Title
Extensions and Applications of Multilevel and Multidimensional Item Response Models

Permalink
https://escholarship.org/uc/item/1mh67889

Author
Choi, In Hee

Publication Date
2013

Peer reviewed|Thesis/dissertation
Extensions and Applications of Multilevel and Multidimensional Item Response Models

By

In Hee Choi

A dissertation submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Education

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Mark Wilson, Chair
Professor Sophia Rabe-Hesketh
Professor Alan Hubbard

Fall 2013
Abstract

Extensions and Applications of Multilevel and Multidimensional Item Response Models

by

In Hee Choi

Doctor of Philosophy in Education

University of California, Berkeley

Professor Mark Wilson, Chair

Multilevel and multidimensional item response models are two commonly used examples as extensions of the conventional item response models. In this dissertation, I investigate extensions and applications of multilevel and multidimensional item response models, with a primary focus on longitudinal item response data that include students’ school switching, classification of examinees into latent classes based on multidimensional aspects, and measurement models for complicated learning progressions. In the first paper, multilevel item response models for longitudinal data are extended to the crossed-classified models (Rasbash & Goldstein, 1994; Raudenbush, 1993) and multiple membership models (Hill & Goldstein, 1998; Rasbash & Browne, 2001) to incorporate students’ school mobility. If students switch school over time in longitudinal studies, the data structure is not strictly hierarchical; therefore, conventional multilevel models are not applicable. In this study, two types of school mobility and corresponding models are specified. Furthermore, this study investigates the impacts of misspecification of school membership in the analysis of longitudinal data. In the second and third paper, mixture models and measurement models based on multidimensional item response models are presented respectively. The second paper investigates possible usefulness of the mixture random weights linear logistic test model (MixRWLLTM) as a means to identify subgroups of examinees as well as to improve interpretations of differences between latent classes. In the proposed MixRWLLTM, examinees are classified with respect to their multidimensional aspects, a general propensity (intercept) and random coefficients of the item properties. In the third paper, a structured constructs model (SCM) for the continuous latent trait is developed to deal with complicated learning progressions, in which relations between levels across multiple constructs are assumed in advance. Based on the multidimensional Rasch model, discontinuity parameters are incorporated to model the hypothesized relations as the advantage or disadvantage for respondents belonging into a certain level in one construct to reach a level in another construct.
Acknowledgements

I would like to take this opportunity to acknowledge with sincere thanks the following people whom I owe deepest gratitude. First of all, I would like to thank my advisor, Mark Wilson, for his support and guidance. His sharp comments and suggestions make me think about practical significance of my dissertation beyond statistical aspects. I would also like to thank my second advisor, Sophia Rabe-Hesketh, for her support and encouragement over past years. My gratitude is also due to Alan Hubbard, for his keen advice on my dissertation.

My gratitude must go out to the Kwanjeong Educational Foundation, which has supported me financially for my first four years of graduate studies. Being selected as a scholarship recipient enabled me not only to come to the U.S., but also to explore every possible opportunity to learn and grow. I am deeply grateful to my former advisor, Sun-Geun Baek and professors in the Department of Education, Seoul National University. I also thank to my colleagues and friends in the Quantitative Methods and Evaluation program at UC Berkeley and my dear friends in Berkeley, Athens and Korea. Without their support and friendship, I would not survive from this long journey.

Lastly but most importantly, I greatly appreciate my family for their endless love and sacrifice. I thank my mother, Hosim Joo, who has believed in and supported me. I know it was not easy for her to let me leave her and study abroad, but she has always encouraged me in my decision and prayed for me. I also acknowledge my parents in law, Bum-Sik Yoon and Im-Sook Im, who have been on my side and prayed for me. Most of all, I would like to thank my husband, Seyoon Yoon. He is the one who let me stand up again, even when I’ve lost faith of myself. His love and support make me special and my life at Berkeley blissed. And finally, I dedicate this dissertation to my late father, Gwangjin Choi, who must be proud of me and most delighted with this achievement.
Table of Contents

Acknowledgements ........................................................................................................... i
Table of Contents ............................................................................................................... ii
List of Tables ..................................................................................................................... v
List of Figures .................................................................................................................... vi
Chapter 1. General Introduction ...................................................................................... 1

Chapter 2. Incorporating Subject Mobility in Growth Modeling for Multilevel and Repeated Item Response Data ................................................................. 4
  2.1. Introduction ................................................................................................................ 4
  2.2. Methods ...................................................................................................................... 6
    2.2.1. Multilevel Item Response Models ...................................................................... 6
    2.2.2. Three-Level Hierarchical Generalized Linear Model for Longitudinal Item Response Data (HGLM-LIRD) ................................................................. 8
    2.2.3. Two Types of School Mobility in Longitudinal Item Response Data 12
    2.2.4. Estimation .......................................................................................................... 19
  2.3. Simulation Study ....................................................................................................... 21
    2.3.1. Type I: Data Generation .................................................................................. 21
    2.3.2. Type I: Analysis ............................................................................................... 23
    2.3.3. Type I: Results ................................................................................................. 23
    2.3.4. Type II: Data Generation ................................................................................ 25
    2.3.5. Type II: Analysis ............................................................................................. 25
    2.3.6. Type II: Results ............................................................................................... 26
  2.4. Empirical Data Study ............................................................................................... 28
    2.4.1. Korean Youth Panel Survey (KYPS) ................................................................. 28
    2.4.2. National Educational Longitudinal Study (NELS: 88) .................................... 31
    2.4.3. Early Childhood Longitudinal Study – Kindergarten Cohort (ECLS-K) .......... 32
  2.5. Conclusion and Discussion ...................................................................................... 34
Chapter 3. Multidimensional Classification of Examinees based on the Mixture Random Weights Linear Logistic Test Model ........................................ 40
  3.1. Introduction .................................................................................. 40
  3.2. Methods ....................................................................................... 43
    3.2.1. Mixture Extensions of the LLTM and RWLLTM .................. 43
    3.2.2. Estimation ............................................................................. 45
  3.3. Empirical Data Study ................................................................... 46
    3.3.1. Data Source .......................................................................... 46
    3.3.2. Analysis .............................................................................. 47
    3.3.3. Results ............................................................................. 48
  3.4. Simulation Study .......................................................................... 52
    3.4.1. Data Generation ................................................................. 52
    3.4.2. Analysis .............................................................................. 53
    3.4.3. Results ............................................................................. 53
  3.5. Conclusion and Discussion .......................................................... 54

Chapter 4. Structured Constructs Model for a Continuous Latent Trait with Discontinuity Parameters ......................................................... 57
  4.1. Introduction .................................................................................. 57
  4.2. Structured Constructs Models ..................................................... 59
    4.2.1. Structured Constructs Model based on Latent Class Analysis .... 59
    4.2.2. Structured Constructs Model for the Latent Continuous Trait ... 61
  4.3. Simulation Study .......................................................................... 64
    4.3.1. Data Generation ................................................................. 64
    4.3.2. Analysis .............................................................................. 64
    4.3.3. Results ............................................................................. 65
  4.4. Empirical Data Study ................................................................... 67
    4.3.1. Data Source .......................................................................... 67
    4.3.2. Analysis .............................................................................. 67
    4.3.3. Results ............................................................................. 68
4.5. Conclusion and Discussion...................................................................................... 72

Chapter 5. Summary and Conclusion............................................................................. 74

References...................................................................................................................... 77

Appendix A. WinBUGS Code for the CCMM-LIRD..................................................... 84

Appendix B. Design Matrix of the Verbal Aggression Data.......................................... 87

Appendix C. WinBUGS code for the MixLLTM and MixRWLLTM.............................. 88

Appendix D. WinBUGS Code for the SCM1 of the ADM Analysis......................... 91
List of Tables

Table 2-1. An Example of Coding Scheme for the Two-Piece Linear Growth Model ......................................................................................................................... 14

Table 2-2. Bias and RMSE of the Type I Simulation Study ........................................ 24
Table 2-3. Bias and RMSE of the Type II Simulation Study ....................................... 27
Table 2-4. Contents of Items regarding Future Occupation Selection of the KYPS .... 29
Table 2-5. Results from the KYPS Vocational Maturity Data Analysis ..................... 30
Table 2-6. Contents of Self-Esteem Items of the NELS:88 ........................................ 32
Table 2-7. Results from the NELS:88 Self-Esteem Data Analysis ............................. 33
Table 2-8. Results from the ECLS-K Mathematics Data Analysis ........................... 35
Table 3-1. Coding Scheme for Item Properties in the Verbal Aggression Data ......... 47
Table 3-2. Estimates for the One-Class and Two-Class LLTM and RWLLTM .......... 49
Table 3-3. Estimates of Coefficients for the Behavior Type ..................................... 50
Table 3-4. Gender Compositions in the Two Latent Classes .................................... 52
Table 3-5. Bias and RMSE of the Simulation Study ................................................ 54
Table 4-1. Bias and RMSE in the Multidimensional Rasch Model and the SCM ..... 66
Table 4-2. Parameter Estimates and Standard Errors for the ADM Data ............... 69
List of Figures

Figure 2-1. Diagram for (a) simple two-level nested model, (b) cross-classified model, and (c) multiple membership model (Browne et al., 2001) .......................... 5

Figure 2-2. Path diagram of the two-level framework of the Rasch model ............................ 8

Figure 2-3. Structure of longitudinal item response data in the three-level approach, where responses on items at level 1 are nested in occasions at level 2, which in turn nested in students at level 3 ................................................ 10

Figure 2-4. Path diagram of the three-level approach to longitudinal item response data (no random slope) ........................................................................... 12

Figure 2-5. An example of data structure in the Type I mobility, in which students attended middle schools at time $t'$ and high schools at time $t''$ .............................. 13

Figure 2-6. Diagram of the CCMM-LIRD with the Type I mobility, in which student $j$ migrated from middle school $m$ to high school $h$ between time $t'$ and time $t''$ .......................................................... 16

Figure 2-7. An example of data structure in the Type II mobility, in which student $j$ moved from school $s'$ and school $s''$ between time $t'$ and time $t''$ ...................... 17

Figure 2-8. Diagram of the CCMM-LIRD with the Type II mobility, in which student $j$ moved from school $s'$ and school $s''$ between time $t'$ and time $t''$ ........ 19

Figure 2-9. An example of combination of middle school and high school membership manipulated in the Type I simulation study ........................................ 22

Figure 3-1. Diagram of estimated $\Theta_{pg}$ by each latent dimension ....................................... 56

Figure 4-1. Theoretical learning progressions of the ADM constructs ................................. 58

Figure 4-2. An hypothesized link between two constructs ...................................................... 62

Figure 4-3. An example of the ability distribution in the requirement and target construct ................................................................................................................. 63

Figure 4-4. Distributions of the estimated latent abilities of CoS and Cha .................... 71
Chapter 1.

General Introduction

Multilevel models (Goldstein, 2003; Rabe-Hesketh & Skrondal, 2012), also known as hierarchical linear models (Raudenbush & Bryk, 2002), have been widely applied for hierarchical structured data, such as, for example, students are nested within classes, in the educational and social researches. Applications of multilevel models can be found in item response theory (IRT) models which can be framed as a two-level model, (Adams, Wilson, & Wu, 1997; Rijmen, Tuerlinckx, De Boeck, & Kuppens, 2003). Conceptualization of the IRT models as multilevel models allows us to reflect the multilevel data structure, and, at the same time, to estimate item parameters and person measures (Fox & Glas, 2001; Kamata, 2001). For example, a multilevel structure is found in longitudinal item response data in which the same set of items is administered to the students over time, thus, responses are nested within a certain measurement occasion and occasions are nested within a student. For analyzing longitudinal item response data, multilevel item response models, where responses, occasions and students correspond to level 1, 2 and 3 respectively, are developed (Pastor & Beretvas, 2006; Segawa, 2005).

Another strand in development of the IRT models is associated with extensions into multidimensional models (Reckase, 1985). In most IRT approaches, it is assumed that items measure one common latent variable, referred to as the unidimensionality assumption. However, current practices in measurement, such as standardized tests, often require more than one ability for students to give a correct answer on test items (Adams, Wilson, & Wang, 1997). Multiple latent ability dimensions, which are commonly assumed to be correlated to each other, are incorporated in statistical analyses in the multidimensional item response model.

This dissertation consists of three papers regarding extensions and applications of multilevel and multidimensional item response models. In the first paper, multilevel item response models for longitudinal data are extended to the crossed-classified models (Rasbash & Goldstein, 1994; Raudenbush, 1993) and multiple membership models (Hill & Goldstein, 1998; Rasbash & Browne, 2001) to incorporate students’ school mobility. In the second and third papers, I propose a mixture model and a measurement model based on the multidimensional item response models. These three papers correspond to Chapter 2, 3, and 4 respectively in this dissertation and below I provide brief introductions of each chapter.
Chapter 2. Incorporating Mobility in Growth Modeling for Multilevel and Longitudinal Item Response Data

One of the assumptions in multilevel modeling is that the data structure is strictly hierarchical, such that students are nested within a school and schools are clustered into a neighborhood. However, in educational research, the data structure is often more complicated than this simple and strict form of hierarchy and a canonical example is the case, in which students move from school to school in longitudinal studies. In this study, the cross-classified and multiple membership models for longitudinal item response data (CCMM-LIRD) are proposed to incorporate students’ school mobility. Furthermore, this study investigates the impacts of misspecification of school membership in the analysis of longitudinal data. Two types of school mobility, which are frequently observed in educational research, are described, and corresponding models are specified. Estimation using Bayesian methods of Markov chain Monte Carlo (MCMC) is presented as well. Simulation studies are conducted to evaluate parameter recovery and the consequences of misspecification of the school-level random effects. Three sets of large-scale longitudinal data are analyzed to illustrate applications of the CCMM-LIRD for two types of school mobility.

Chapter 3. Multidimensional Classification of Examinees based on Mixture Random Weights Linear Logistic Test Model

The purpose of this study is to investigate possible usefulness of the mixture random weights linear logistic test model (MixRWLLTM) as a means to identify subgroups of examinees as well as to improve interpretations of differences between latent classes. In particular, for better understanding of characteristics of latent groups, this study takes advantage of explanatory aspects of the linear logistic test model (LLTM: Fischer, 1973), in which item design properties are used to explain item difficulties. Moreover, in the MixRWLLTM, examinees are classified with respect to their multidimensional aspects, a general propensity (intercept) and random coefficients of the item properties. This study presents the conceptual framework of the mixture extensions of the LLTM and RWLLTM, and estimation for the proposed models based on the MCMC algorithm. Results from an empirical example using verbal aggression data and simulation study are illustrated. Moreover, practical issues in Bayesian estimation for the mixture IRT models including model selection and label switching are discussed in the empirical data and simulation studies.

Chapter 4. Structured Constructs Model for the Continuous Latent Trait with Discontinuity Parameters

A structured constructs model (SCM) for the continuous latent trait is developed to deal
with complicated learning progressions. Particularly, in this study, complexity of learning progressions is defined as multiple constructs (or dimensions) and hypothesized links between multiple constructs. Based on the multidimensional Rasch model, the proposed model assumes that each construct is represented as a latent continuum and levels within a continuous construct can be determined by setting the cut score. Furthermore, discontinuity parameters are incorporated to model a hypothesized link between constructs such that students cannot attain a certain level in one construct without reaching a level in the other construct. Therefore, the developed SCM approach differs from the multidimensional Rasch model due to hypothesized relations between multiple dimensions, and distinguishes from the previous SCM approach based on ordered latent class model which assumes that each construct consists of ordered sets of latent classes. In this chapter, theoretical framework of the SCM for the continuous latent trait is described. The simulation study is performed using Bayesian estimation using MCMC algorithm. As an empirical example, mathematics data of two constructs of the Assessing Data Modeling and Statistical Reasoning (ADM) project is analyzed.
Chapter 2.

Incorporating Mobility in Growth Modeling for Multilevel and Repeated Item Response Data

2.1. Introduction

The purpose of multilevel models or hierarchical linear models (e.g., Goldstein, 2003; Raudenbush & Bryk, 2002) is to handle nesting structures, which are frequently observed in social research settings, thus allowing researchers to investigate the effects of contextual factors, such as teachers’ education and school types on student growth. One of the assumptions in multilevel modeling is that the data structure is strictly hierarchical, that is, students are nested within schools, as is illustrated in Figure 2-1 (a). Therefore, shared but unobserved environmental variables for students within schools induce a positive correlation among the outcomes for the students in the same school and the dependence is represented in the random effects of the schools in multilevel modeling. This means that the standard statistical analysis procedures need to be modified to allow for these dependences. However, in educational research, the data structure is often more complicated than this strict form of hierarchy.

The first type of this complex data structure that will be studied in this chapter is found when lower level units are nested within a combination of two or more higher level units. For instance, if the students within a primary school proceed to the same secondary school and the primary schools are nested within the secondary schools, the data structure follows the three levels of the students (level 1), within the primary schools (level 2), within the secondary schools (level 3). However, the students who attended a particular primary school do not necessarily enter the same secondary school; instead, the students within the same primary school will move to multiple secondary schools and the secondary schools will draw students from multiple primary schools. Consequently, the primary schools are not nested purely within secondary school, but rather each student is nested within a single pair of the primary school and secondary school that he or she attended. The data structure in this example can be represented using Figure 2-1 (b) (Browne, Goldstein, & Rasbash, 2001), in which rectangles represent sets of classification units and arrows going from the lower-level unit to the higher-level units describe membership classifications. In a cross-classified model (Rasbash & Goldstein, 1994; Raudenbush, 1993) which was developed to analyze this type of multilevel data, two classifications at level 2 for primary schools and secondary schools (e.g., two separate rectangles in Figure 2-1 (b)) are assumed and the students have a membership in each classification (e.g., one arrow from the student to the primary school and one arrow from the student to the secondary school in Figure 2-1 (b)).
Another complication of the multilevel data structure is addressed by a multiple membership model (Hill & Goldstein, 1998; Rasbash & Browne, 2001), in which lower-level units are simultaneously members of more than one units within the same higher-level classification. In conventional multilevel modeling in educational researches, students are assumed to belong to one primary school or one teacher, but it is not uncommon that students attend one or more primary schools, or that students are taught by multiple teachers. A typical example of the multiple membership model is an analysis of school effects on student achievement test scores at end-of-year exams, when some students have attended multiple schools during a school year. In Figure 2-1 (c), the school is a single classification unit at level 2, represented by a rectangle of the school, and the double arrows from the student to the school display the student’s multiple school membership.

The outcomes of interest in this study are longitudinal test data, in which responses on the same set of items from the same students are collected over time. In this case, the responses on the items are clustered into a certain time point and repeated occasions are nested within a student. If the students have attended the same schools over the course of data collection, the data structure is extended into a four-level strict hierarchy. However, students often move from school to school for various reasons. In other words, this situation can be described as multiple measures of the same student over time and multiple schools that each student attends over time. Therefore, models that account for the likely positive correlation among multiple measures of the same student and among students in the same school are required and cross-classified and multiple membership models are thus considered to be major tools in this study. The use of cross-classified and multiple membership models has increased in empirical research, however, most of the applications have concentrated on cross-sectional data (Chung & Beretvas, 2011; Fielding, 2002; Jayasinghe, Marsh, & Bond, 2003; Meyers...
or longitudinal data with continuous outcomes (Grady & Beretvas, 2010; Jeon & Rabe-Hesketh, 2012; Luo & Kwok, 2012). This study aims to investigate the application of cross-classified and multiple membership models in modeling growth for multilevel longitudinal item response data as an extension of the three-level hierarchical generalized linear model. Accordingly, the proposed models will take advantage of the item response models as well as the cross-classified and multiple membership models: these provide item-level information and accommodate the complicated data structures frequently encountered in longitudinal studies.

The goals of this study are twofold. One is to demonstrate an application of the three-level item response modeling approach to analyzing longitudinal item response data in which students switch schools between measurement occasions. The other is to investigate the impacts of misspecifications of school membership in the analysis of longitudinal data sets that include mobile students: examples of misspecifications include ignoring school membership and using only the information from the school that the students attended at an initial time point.

To this end, this chapter is organized as follows. First, the multilevel item response models and three-level approaches to the longitudinal item response data are introduced. Second, models are proposed to deal with the two types of school mobility based on the cross-classified and multiple membership models, and a brief explanation is given as to how the Bayesian methods of Markov chain Monte Carlo (MCMC) can be employed to fit the proposed models. Third, two simulation studies are conducted for the two types of school mobility to assess the parameter recovery and the impacts of misspecifications. Fourth, empirical examples of real data sets, which were analyzed previously using cross-classified and multiple membership models with a focus on cross-sectional data and continuous outcomes, are illustrated. Lastly, the chapter ends with concluding remarks and suggestions for further studies.

2.2. Method

2.2.1. Multilevel Item Response Models

In the dichotomous Rasch model, the probability of a correct response is written as

\[ P(y_{ij} = 1 | \theta_j) = \frac{\exp(\theta_j - \delta_i)}{1 + \exp(\theta_j - \delta_i)}, \]

where \( y_{ij} \) represents the response to item \( i = 1, \ldots, I \) from student \( j = 1, \ldots, J \), \( \theta_j \) is the ability of student \( j \) and \( \delta_i \) is the difficulty parameter of item \( i \). In the Rasch model, it is common to consider the ability \( \theta_j \) as a random variable, the latent variable of student \( j \),
and the item difficulties $\delta_i$ as fixed parameters. The Rasch model can be interpreted as a two-level generalized linear model, in which the responses and students are level 1 and level 2 units respectively and responses from the same student are nested within the student (Adams, Wilson, & Wu, 1997; Mislevy & Bock, 1989).

In multilevel IRT modeling, Equation (2.1) corresponds to the level 1 model (measurement model) and the probability is rewritten as

$$
\log \left( \frac{P(y_{ij} = 1)}{1 - P(y_{ij} = 1)} \right) = \pi_{0j} + \sum_{q=1}^{I} X_{qj} \pi_{qj},
$$

where $X_{qi}$ is the $q$th indicator variable, with value of -1 when $q = i$ and 0 when $q \neq i$. In other words, $\pi_{0j}$ corresponds to the random intercept for student $j$ and $\pi_{qj}$ refers to the random coefficient for student $j$ associated with the level 1 predictor variable $X_{qi}$ indicating each item, often called an indicator variable. In the level 2 model, $\pi_{qj}$ is specified as constant across students and then corresponds to the item difficulty of each item $\delta_q$ in Equation (2.1), while $\pi_{0j}$ is assumed to vary across students,

$$
\pi_{0j} = \beta_0 + \zeta_{0j} \\
\pi_{1j} = \delta_1 \\
\vdots \\
\pi_{ij} = \delta_I,
$$

where $\beta_0$ is the fixed intercept across students and $\zeta_{0j}$ is the level 2 residual that follows a normal distribution mean zero and constant variance $\sigma^2$, $\zeta_{0j} \sim N(0, \sigma^2)$. For model identification, a constraint such as either $\delta_1 = \sum_{i=1}^{I-1} \delta_i$, thus, $\sum_{i} \delta_i = 0$ or $\beta_0 = 0$, is imposed. In this study, the item difficulty constraint is used and with this constraint, only the difficulties of the first $(I - 1)$ items, $\delta_1, \ldots, \delta_{I-1}$ are freely estimated and the elements of the design matrix for the $I$th item are equal to one, $X_{qI} = 1$.

Substituting Equation (2.3) into Equation (2.2) yields

$$
\log \left( \frac{P(y_{ij} = 1)}{1 - P(y_{ij} = 1)} \right) = \beta_0 + \zeta_{0j} + \sum_{q=1}^{I} X_{qj} \delta_q.
$$

The latent variable of student $j$ in Equation (2.1) is now expressed as a linear regression model with no covariates,

$$
\theta_j = \pi_{0j} = \beta_0 + \zeta_{0j}.
$$

Note that, unlike the conventional regression models, the outcome variable of Equation
is an unobserved latent variable, referred to as a latent regression (Adams, Wilson, & Wu, 1997). The two-level framework of the Rasch model can be presented in a path diagram using the notations suggested by Rabe-Hesketh, Skrondal, and Pickles (2004).

In Figure 2-2, the observed responses $y_{ij}$ are represented with rectangles, the latent response variable $\theta_j$ with an enclosed circle and $\zeta_{0j}$ is the latent variable serving as the residual. The outer frame labeled “Student” represents the nested structure, in which the responses are clustered in the student. An arrow from $\zeta_{0j}$ to $\theta_j$ represents a linear relationship and arrows from $\theta_j$ to $y_{ij}$ represent a nonlinear relationship between the latent variable and the observed responses inducing dependence among the level 1 units (responses) within the same students. A short arrow pointing to the observed variable represents level 1 variability which follows a Bernoulli distribution in this model. As an extension of the unconditional model, item-level covariates can be included in the linear logistic test model (LLTM: Fischer, 1973), and student-level covariates such as gender or age, can be included (Rijmen et al., 2003). In addition, the two-level model can be extended to the three-level models that include teachers or schools as higher level units (Fox & Glas, 2001; Kamata, 2001).

2.2.2. Three-Level Hierarchical Generalized Linear Model for Longitudinal Item Response Data (HGLM-LIRD)

In the two-level generalized linear model formulation of the Rasch model, each student responds to the same set of items and the responses from the same student are more correlated than the responses from the other students, defined as within-cluster correlation in multilevel modeling. In the Rasch model, expressed as Equation (2.4), the level 2 random effect (residual), $\zeta_{0j}$ is the source of within-student correlation (Rabe-Hesketh & Skrondal, 2012; Raudenbush & Bryk, 2002). In education and
psychology research settings, the measurement of individual growth or change in a construct is a focus of studies in many situations. For the purpose of investigating growth, the same set of items (or with common items at least) is administered to the students repeatedly over time, and longitudinal item response data is collected.

Bacci (2012) compared two major approaches based on multidimensional and multilevel item response models to analyzing longitudinal item response data. In the multidimensional framework, multiple dimensions that are specific to each measurement occasion (Andersen, 1985) or that represent an initial latent variable and additional change or growth between consequent occasions are assumed (Embretson, 1991). Recently, Wilson, Zheng, and McGuire (2012) proposed the latent growth item response model, in which the growth is modeled by assuming two dimensions of an initial latent variable and a constant change between consecutive occasions. In the multilevel item response approach, the two-level Rasch model is extended to the three-level model, in which item responses are the level 1 units, measurement occasions are the level 2 units and students are the level 3 units, and change is modeled using growth models by adopting multilevel models (Pastor & Beretvas, 2006; Segawa, 2005).

In the multidimensional approach, separate latent variables for each measurement occasion are specified, however, there is less flexibility in the data collection, such as fixed occasions for all students, and the complexity of estimation increases as the measurement occasions increases. In contrast to the multidimensional approach, the multilevel framework allows for a different number of occasions for different students, and unequally spaced occasions across students. Given the advantages and drawbacks of the two approaches, the choice depends on the data structure and the research purposes. Since this study investigates the trends of individual growth and the related extensions of the growth models, the multilevel approach, which is discussed in detail below, has been chosen to analyze the longitudinal item response data.

While students are measured repeatedly with items in the two-level item response models, in longitudinal item response data, students are measured repeatedly in two aspects, measurement occasions and items (Littell, Milliken, Stroup, Wolfinger, & Schabenberger, 2006), allowing for the three-level modeling of responses. A set of responses from a student on one occasion are more alike than responses from another occasion, and responses from the same student are more correlated than those from another student. In other words, there are two possible types of within-cluster correlations in longitudinal item response data, within-student and within-occasion correlation as well as within-student and between-occasion correlation. In order to deal with these within-student correlations in longitudinal item response data, a strict three-level approach is specified, in which item responses are nested within an occasion, and occasions are nested within students. Figure 2-3, modified from Figure 2-1 (a), displays a nested structure of units (rectangles) which is represented by a single arrow from
lower units to higher units. In such three-level modeling, the level 1 model is for the item response at a specific measurement occasion; variation in the latent variable across measurement occasions within student is expressed in terms of growth models in the level 2 model. Finally, the level 3 model describes the variation in parameters of growth trajectories between students (Hung & Wang, 2012).

**Level 1 Model** The level 1 model, referred to as the measurement model, specifies the item response functions. Let $y_{ijt}$ denote the response to item $i$ at measurement occasion $t$ for student $j$, for $i = 1, \ldots, I$, $t = 1, \ldots, T$, and $j = 1, \ldots, J$. The probability that student $j$ gives a correct response on item $i$ at occasion $t$ is written as

$$P(y_{ijt} = 1 | \theta_j) = \frac{\exp(\theta_j - \delta_i)}{1 + \exp(\theta_j - \delta_i)},$$

where $\theta_j$ represents the latent variable of student $j$ at occasion $t$, and $\delta_i$ denotes the fixed difficulty parameter of item $i$. Unlike the Rasch model, the latent variable $\theta_j$ is occasion-specific as well as student-specific, indicating it is a time-varying variable. In Equation (2.6), item difficulties are fixed to be invariant across measurement occasions, with the following constraint, $\delta_I = -\sum_{i=1}^{I+1} \delta_i$. The probability of a correct response is rewritten as

$$\log \left( \frac{P(y_{ijt} = 1 | \theta_j)}{1 - P(y_{ijt} = 1 | \theta_j)} \right) = \theta_j + \sum_{q=1}^{T} X_{qj} \delta_q.$$  

**Level 2 Model** At level 2, a latent growth curve model (Duncan, Duncan, & Strycker, 2006; McArdle & Epstein, 1987) is specified to model the latent variable of student $j$ at
time $t$ as a function of the time variable, allowing for estimation of individual growth trajectories. To illustrate, the level 2 (between-occasion and within-student) model for the latent variable of the level 1 model is a linear growth model,

$$\theta_j = \pi_{0j} + \pi_{1j}d_t + \epsilon_j,$$  \hfill (2.8)

where $d_t$ is the time variable taking on values of 0, 1, . . . , $T - 1$ for occasion 1, 2, . . . , $T$, $\pi_{0j}$ and $\pi_{1j}$ are the intercept and slope parameter of student $j$, and $\epsilon_j$ is the level 2 random effect (residual) of student $j$ at time $t$. In Equation (2.8), $\pi_{0j} + \pi_{1j}d_t$ is the linear growth trajectory of student $j$, where $\pi_{0j}$ and $\pi_{1j}$ represent the initial status and linear change of the latent variable, and $\epsilon_j$ is the deviation at time $t$ from the linear growth trajectory of student $j$. In the growth models, $\epsilon_j$ is often assumed to be normally distributed with mean zero and a constant variance, $\epsilon_j \sim N(0, \sigma^2)$, that is, an i.i.d. structure. As an extension of the linear growth model, higher order polynomials of the time variable and time-varying covariates can be included and it is possible to assume an alternative specification of $\epsilon_j$, such as an autoregressive structure (Hung & Wang, 2012; Segawa, 2005).

**Level 3 Model** In the level 3 (between-student) model, the student-specific growth parameters serve as dependent variables,

$$\pi_{0j} = \beta_0 + \zeta_{0j},$$

$$\pi_{1j} = \beta_1,$$  \hfill (2.9)

where $\beta_0$ and $\beta_1$ are the fixed intercept and linear growth rate across students respectively, and $\zeta_{0j}$ is the random effect (intercept or residual) of student $j$. It is assumed that $\zeta_{0j}$ follows a normal distribution, $\zeta_{0j} \sim N(0, \psi^2)$, and $\text{Cov}(\zeta_{0j}, \epsilon_j) = 0$.

Substituting Equation (2.9) into Equation (2.8) yields the latent regression for $\theta_j$,

$$\theta_j = \beta_0 + \beta_1d_t + \zeta_{0j} + \epsilon_j.$$  \hfill (2.10)

Equation (2.10) can be rewritten in a matrix form for student $j$ as follows:

$$\Theta_j = X_1B + Z_1\zeta_{0j} + \epsilon_j,$$  \hfill (2.11)

where $\Theta_j = \begin{bmatrix} \theta_{1j} \\ \vdots \\ \theta_{Tj} \end{bmatrix}$, $X_1 = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & T-1 \end{bmatrix}$, $Z_1 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$, $B = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$, and $\epsilon_j = \begin{bmatrix} \epsilon_{1j} \\ \vdots \\ \epsilon_{Tj} \end{bmatrix}$. Keeping the notations in Figure 2-2, the two-level Rasch model can be extended to the three-level model for analyzing longitudinal item response data, for example, in which student $j$ is measured at a certain time point $t$ with $I$ items, represented as in Figure 2-4.
Figure 2-4. Path diagram of the three-level approach to longitudinal item response data (no random slope)

The latent response variable $\theta_{ij}$, represented with an enclosed circle, is expressed as a function of two residual terms, level 3 residual (intercept) $\zeta_{0j}$, which induces dependence among the responses from a student over time and level 2 residual $\epsilon_{ij}$, which induces dependence among the responses at a time point $t$ after conditioning on $\zeta_{0j}$. The outer frame indicates the nesting structure of the occasions and students.

2.2.3. Two Types of School Mobility in Longitudinal Item Response Data

As discussed earlier, when students switch schools in the course of repeated measurements in longitudinal studies, the consequent complicated data structure requires alternative approaches to multilevel modeling. In this section, two types of mobility often observed in longitudinal data (Luo & Kwok, 2012) are described and cross-classified and multiple membership models for longitudinal item response data (CCMM-LIRD) with two types are specified.

**Type I** The first type of students’ mobility considered in this study is that students move simultaneously at a certain time point due to promotion by the educational system, for example, by graduating from middle schools and entering high schools. For instance, in the Korean Youth Panel Survey (KYPs), the first survey was administered to second-year middle school students and followed students once a year until their high school graduation. The National Educational Longitudinal Study (NELS:88) tracked eighth grade students through four follow-ups once every two years, thus, the students migrated to high schools after the first measurement occasion. In these cases, the strict three-level data structure in Figure 2-3 needs to be extended to the cross-classified model, in which students are nested within a combination of middle
Figure 2-5. An example of data structure in the Type I mobility, in which students attended middle schools at time \( t’ \) and high schools at time \( t'' \)

schools and high schools. To illustrate, in Figure 2-5, which is similar to one suggested by Jeon and Rabe-Hesketh (2012), solid rectangles and arrows represent a clustered structure of items, times and students. In particular, middle schools and high schools are represented as two separate and unconnected rectangles located at the same level and the cross-classified relationship is described by two arrows from the students to either middle schools or high schools. Furthermore, dotted rectangles indicate specific time points within the time level. Suppose students attended middle schools at a time point \( t’ \) and high schools at a time point \( t'' \). Therefore, the responses at time \( t’ \) are nested into the middle schools and the ones at time \( t'' \) are nested into the high schools, represented by dotted arrows respectively. Given this nested relationship, the responses at a certain time point are likely to be correlated due to the unobserved effects of the middle or high school that the student attended. In the cross-classified models, the separate random effects of the middle schools and high schools are specified to explain within-middle school or within-high school correlations.

Another characteristic of this type of mobility is that students switch schools at the same time, separating measurement occasions into two distinct periods (e.g., years of middle school and high school). For the purpose of investigating different growth patterns during middle schools and high schools, a piecewise growth model that allows for breaking of the growth trajectories up into several linear components according to distinct developmental periods is used (Li, Duncan, Duncan, & Hops, 2001; Raudenbush & Bryk, 2002). For instance, in the case of Figure 2-5, two time-related
Table 2-1. An Example of Coding Scheme for the Two-Piece Linear Growth Model

<table>
<thead>
<tr>
<th></th>
<th>( t = 1 )</th>
<th>( t = 2 )</th>
<th>( t = 3 )</th>
<th>( t = 4 )</th>
<th>( t = 5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_{1t} )</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>( d_{2t} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

variables are composed using a scheme given in Table 2-1 for the two-piece linear growth model and the coefficients of \( d_{1t} \) and \( d_{2t} \) are the growth rates during middle school and high school respectively.

Suppose that there are \( M \) middle schools and \( H \) high schools and the middle schools and high schools are indexed by \( m = 1, \ldots, M \) and \( h = 1, \ldots, H \). The response on item \( i \) at occasion \( t \) of student \( j \) who attended middle school \( m \) and high school \( h \) is denoted by \( y_{ijmh} \), and the level 1 measurement model is written as

\[
\log \left( \frac{P(y_{ijmh} = 1 \mid \theta_{ijmh})}{1 - P(y_{ijmh} = 1 \mid \theta_{ijmh})} \right) = \theta_{ijmh} + \sum_{q=1}^{l} X_{q} \delta_{q},
\]

where \( \theta_{ijmh} \) is the latent variable at occasion \( t \) of student \( j \) who attended middle school \( m \) and high school \( h \), and \( \delta_{i} \) indicates the fixed difficulty parameter of item \( i \). As the subscripts imply, only the latent ability is allowed to vary across occasions. In the adoption of the two-piece linear growth model with time-related variables associated with two separate periods, \( \theta_{ijmh} \) is written in the reduced form of the latent variable as,

\[
\theta_{ijmh} = \beta_{0} + \beta_{1} d_{1t} + \beta_{2} d_{2t} + \zeta_{0j} + w_{1t} \gamma_{0m} + w_{2t} \eta_{0h} + \varepsilon_{ijmh},
\]

where \( \beta_{0} \) is the fixed intercept, \( \beta_{1} \) and \( \beta_{2} \) represent the fixed slopes while attending middle school and high school respectively, \( \zeta_{0j} \) denotes the random effect of student \( j \) related to the intercept, and \( \varepsilon_{ijmh} \) is the residual at level 2. In order to explain the deviations from a student-specific growth line due to student \( j \)’s studying in middle school \( m \) and high school \( h \), school-specific random effects, and \( \gamma_{0m} \) and \( \eta_{0h} \) related to the intercept for middle schools and high schools, respectively, are specified and \( w_{1t} \) and \( w_{2t} \) are the coefficients that associate middle school and high school effects with the latent variable at a specific time point \( t \).

The latent variable of student \( j \) who attended middle school \( m \) and high school \( h \) in Equation (2.13) can be rewritten in a matrix form as,

\[
\Theta_{jm} = X_{j} \beta + Z_{j} \zeta_{0j} + Z_{2} \gamma_{0m} + Z_{3} \eta_{0h} + \varepsilon_{jm},
\]
The student, presented using a path diagram to affect students’ responses when 0

\[ \begin{bmatrix} \theta_{j,mh} \\ \vdots \\ \theta_{T,jmh} \end{bmatrix} \]

\[ X_1 = \begin{bmatrix} 1 & d_{11} & d_{21} \\ \vdots & \vdots & \vdots \\ 1 & d_{1T} & d_{2T} \end{bmatrix}, \]

\[ Z_1 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad Z_2 = \begin{bmatrix} w_{11}^T \\ \vdots \\ w_{2T}^T \end{bmatrix}, \quad Z_3 = \begin{bmatrix} \vdots \end{bmatrix}, \]

\[ B = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}, \quad \text{and} \quad \varepsilon_{j,mh} = \begin{bmatrix} \varepsilon_{1,j,mh} \\ \vdots \\ \varepsilon_{T,jmh} \end{bmatrix}. \]

The student-level and school-level random intercepts are assumed to follow a normal distribution with mean zero and a constant variance: \( \zeta_{0j} \sim N(0, \psi^2) \), \( \gamma_{0m} \sim N(0, \tau^2) \) and \( \eta_{0h} \sim N(0, \tau^2) \). Thus, \( \tau^2 \) and \( \tau^2 \) indicate variation of the random effects of middle schools and high schools respectively. As in the three-level model, the level 2 residual is assumed to follow a normal distribution with a constant variance: \( \varepsilon_{ij,mh} \sim N(0, \sigma^2) \). It is further assumed that \( \text{Cov}(\zeta_{0j}, \gamma_{0m}) = \text{Cov}(\zeta_{0j}, \eta_{0h}) = \text{Cov}(\zeta_{0j}, \varepsilon_{ij,mh}) = \text{Cov}(\gamma_{0m}, \varepsilon_{ij,mh}) = \text{Cov}(\eta_{0h}, \varepsilon_{ij,mh}) = 0. \)

The coefficients \( w_{11} \) and \( w_{21} \) can be pre-assigned values or unknown parameters that are freely estimated (McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004). In the KYPS example, in which students moved from middle schools to high schools between the second and third occasion over five time points, if the school effects were constant over time and if the middle schools did not affect students’ responses when they were in the high schools, then \( Z_2 = (1, 1, 0, 0, 0)' \) and \( Z_3 = (0, 0, 1, 1, 1)' \). In addition, the cumulative effects of middle school can be specified using vectors, \( Z_2 = (1, 1, 1, 1, 1)' \) and \( Z_3 = (0, 0, 1, 1, 1)' \). However, the assumption of the constant school effects can be relaxed by allowing estimation of the varied impacts of schools using vectors \( Z_2 = (1, w_{12}, w_{13}, w_{14}, w_{15})' \) and \( Z_3 = (0, 0, 1, w_{24}, w_{25})' \) (Jeon & Rabe-Hesketh, 2012). The coefficients \( w_{11} \) and \( w_{23} \) are set to a value of one for model identification and \( w_{21} \) and \( w_{22} \) are fixed to zero since the students were in the middle schools at those time points. In the case of the varying coefficients of school effects, \( Z_2 = (1, w_{12}, w_{13}, w_{14}, w_{15})' \) and \( Z_3 = (0, 0, 1, w_{24}, w_{25})' \) represent how the middle schools and high schools contribute to the students’ current outcomes, compared to the initial time point (e.g., \( t = 1 \) for middle schools and \( t = 3 \) for high schools). Accordingly, if the estimated \( w_{12} \) is greater than one, for instance, the middle school effects increase from the previous year.

In general, the CCMM-LIRD framework for the Type I mobility described in Figure 2-5 can be presented using a path diagram (see Figure 2-6), similar to one suggested by Jeon and Rabe-Hesketh (2012). Specifically, the solid frames represent the nested structure of time, student, middle school, and high school. Note that student \( j \) is nested within both middle school \( m \) and high school \( h \), and the student-level residual \( \zeta_{0j} \) is placed in the intersection of middle school and high school. The dotted frame represents a specific time point within the time level. For instance, time \( t \) when student \( j \) attended middle school \( m \) is nested within student and middle school and the latent response variable at time \( t' \) is expressed as function of the time-level residual \( \varepsilon_{t,j,mh} \), the
Figure 2-6. Diagram of the CCMM-LIRD with the Type I mobility, in which student \( j \) migrated from middle school \( m \) to high school \( h \) between time \( t' \) and time \( t'' \)

student-level residual \( \zeta_{0j} \), and the middle school-level residual \( \gamma_{0m} \). After student \( j \) moved to high school \( h \), the latent response variable is modeled by the high school-level residual \( \eta_{0h} \) as well as the middle school-level residual \( \gamma_{0m} \). In other words, the dotted arrows from the school-level residuals to the latent response variable correspond to the fixed coefficients, \( w_1 \) and \( w_2 \), in Equation (2.13).

**Type II** Another pattern of student mobility is when sub-samples of students switch their school or classroom membership. For example, students can transfer to other schools for various reasons such as family moving, parents’ job change, and other issues during repeated measurement occasions. In such cases, some of the students can move at any time during the data collection, and it is also possible that they can switch their membership multiple times. In several large-scale longitudinal data sets (e.g., NELS:88 and the Early Childhood Longitudinal Study-Kindergarten Class (ECLS-K)), the student mobility rate, defined as the percentage of students who switched schools, ranged from approximately 8% to 17% (Chung & Beretvas, 2011). A report by the U.S Government Accounting Office (U.S. Government Accounting Office, 1994) showed that the average mobility rate was 17%, but for some populations, the rates were much higher: for example, as high as 40% (Grady & Beretvas, 2010).
Consider an example of student achievement measured annually for three years. If the mobility rate is 20%, most of the students will remain within the same school over time, and the structure of their longitudinal item responses corresponds to a strict four-level hierarchy: item responses (level 1), occasions (level 2), students (level 3) and schools (level 4). One group of mobile students is those who switched schools once either between occasion 1 and occasion 2 or between occasion 2 and occasion 3, and another group is students who changed schools at both occasion 2 and occasion 3. Consequently, in this scenario, students can attend more than one school and might have been under the influence of multiple schools. Suppose that student $j$ attended schools $s'$ at time $t'$ and transferred to school $s''$ at time $t''$; the data structure of this example is presented in Figure 2-7. Unlike the data structure of the Type I mobility, the schools are located in one cluster, represented by a solid rectangle, and the particular schools within the school level are displayed by small dotted rectangles. The students’ membership in multiple schools is expressed using double solid arrows from the students to the schools as in Figure 2-1 (c). In addition, dotted arrows show the nested relationship such as the item responses at time $t$ of student $j$ into school $s'$.

Given that the students’ school membership is not constant over time, the impact of the schools on the item responses at a certain time point cannot be modeled as in conventional multilevel modeling. In order to model students’ multiple school membership, a notation suggested by Browne et al. (2001) is used and the schools that student $j$ has attended across occasions are denoted by $s(j)$. Let $S$ denote the total number of schools with $s(j)$ as a subset of the full set of schools: $s(j) \in \{1, \ldots, S\}$. For example, in the case of Figure 2-7, $s(j) = \{s', s''\}$. Then, a response to item $i$ at
measurement occasion \( t \) of student \( j \) who has attended schools \( s(j) \) is written as \( y_{tjs(j)} \). The probability of a correct response is specified as

\[
\log \left( \frac{P(y_{tjs(j)} = 1 \mid \theta_{tjs(j)})}{1 - P(y_{tjs(j)} = 1 \mid \theta_{tjs(j)})} \right) = \theta_{tjs(j)} + \sum_{q=1}^{I} X_{tjs(j)} q_j, \tag{2.15}
\]

where \( \theta_{tjs(j)} \) is the latent variable at occasion \( t \) of student \( j \) who has attended schools \( s(j) \), and \( \delta_i \) indicates the fixed difficulty parameter of item \( i \). To model the growth of the latent variable for student \( j \) over time, a linear growth model with the time variable taking on the values of 0, 1, . . . , \( T - 1 \) for occasion 1, 2, . . . , \( T \) is used,

\[
\theta_{tjs(j)} = \beta_0 + \beta_1 t + \zeta_{0j} + \sum_{k = at(j)} \lambda_{jk} v_{0k} + \epsilon_{tjs(j)}, \tag{2.16}
\]

where \( \beta_0 \) and \( \beta_1 \) are the fixed intercept and linear slope of the linear growth line respectively, \( \zeta_{0j} \) is the random intercept of student \( j \), \( \lambda_{jk} \) is the pre-assigned coefficient for student \( j \) who attended school \( k \) at time \( t \), \( v_{0k} \) is the random effect of school \( k \), and \( \epsilon_{tjs(j)} \) is the level 2 residual. The random intercepts of the students and the schools are assumed to follow a normal distribution with mean zero and a constant variance, \( \zeta_{0j} \sim N(0, \gamma^2) \) and \( v_{0k} \sim N(0, \tau^2) \), where \( \tau^2 \) represents the between-school variance. A constant variance is specified for \( \epsilon_{tjs(j)} \), which follows a normal distribution, \( \epsilon_{tjs(j)} \sim N(0, \sigma^2) \). The random effects of the students and the schools are independent of each other and the level 2 residual is independent of the random effects of the students and the schools, that is, \( \text{Cov}(\zeta_{0j}, v_{0k}) = \text{Cov}(\zeta_{0j}, \epsilon_{tjs(j)}) = \text{Cov}(v_{0k}, \epsilon_{tjs(j)}) = 0 \).

Equation (2.16) can be rewritten in a matrix for student \( j \) as follows:

\[
\Theta_{\mu(j)} = X_1 B + Z_1 \zeta_{0j} + Z_2 v_{0s(j)} + \epsilon_{\mu(j)}, \tag{2.17}
\]

where \( \Theta_{\mu(j)} = \begin{bmatrix} \theta_{tjs(j)} \\ \theta_{tjs(j)} \end{bmatrix} \), \( X_1 = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & T - 1 \end{bmatrix} \), \( Z_1 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \), \( B = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \), and \( \epsilon_{\mu(j)} = \begin{bmatrix} \epsilon_{1tjs(j)} \\ \vdots \\ \epsilon_{Tjs(j)} \end{bmatrix} \).

Furthermore, specifications of \( Z_2 \) and \( v_{0s(j)} \) depend on student school mobility patterns. To illustrate, for student \( j \) who attended school 1 at occasion 1 and 2 and moved to school 2 at occasion 3, \( Z_2 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 2/3 & 1/3 \end{bmatrix} \) and \( v_{s(j)} = \begin{bmatrix} v_{o1} \\ v_{o2} \end{bmatrix} \), and if he or she switched schools two times, such as school 1 at time 1, school 2 at time 2 and school 3 at time 3,
Figure 2-8. Diagram of the CCMM-LIRD with the Type II mobility, in which student $j$ moved from school $s'$ and school $s''$ between time $t'$ and time $t''$

$$Z_{2j} = \begin{bmatrix} 1 & 0 & 0 \\ 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_{s(j)} = \begin{bmatrix} v_{01} \\ v_{02} \\ v_{03} \end{bmatrix}.$$ Lastly, for students who remained in the same school (e.g., school 1) over three time points, $v_{0s(j)} = v_{01}$ and $Z_{2j} = 1$. In other words, the coefficient $\lambda_{ijk}$ in Equation (2.16) indicates the proportion of time that student $j$ has attended school $k$ up to time $t$, thus, $\sum_{k \in s(j)} \lambda_{ijk} = 1$ for each time point $t$ (for each row of $Z_{2j}$). Given that school membership varies across students as well as across occasions, each student can have a different design matrix for the school effects, $Z_{2j}$, as the subscript $j$ implies. Similar to the diagram in Figure 2-6, the CCMM-LIRD for the Type II mobility described in Figure 2-7 is displayed in Figure 2-8. Note that $v_{0s'}$ and $v_{0s''}$ correspond to the school-level random effects (residuals) specific to school $s'$ and $s''$ respectively.

2.2.4. Estimation
A Bayesian approach is used for parameter estimation in the cross-classified and multiple membership models for the longitudinal item response data. I implemented the proposed models in WinBUGS 1.4.3, the free software that implements Markov chain Monte Carlo (MCMC) (Lunn, Thomas, Best, & Spiegelhalter, 2000). Bayesian methods have been widely adopted in complicated item response models, such as multilevel models (Fox & Glas, 2001), longitudinal item response data analysis (Hung & Wang, 2012; Segawa, 2005), and mixture models (Cho & Cohen, 2010; Cohen & Bolt, 2005). In addition, MCMC estimation has been used for fitting the cross-classified effects and multiple membership models, and it has been shown to be feasible for the analysis of models with complex random effects (Browne et al., 2001; Chung & Beretvas, 2011; Grady & Beretvas, 2010; Lockwood, McCaffrey, Mariano, & Setodji, 2007). Due to the complexity of the model structures, in particular, the discrete responses, longitudinal data and complicated nesting structures of the students and the schools, MCMC was chosen for the estimation method in this study. Implementing MCMC in WinBUGS is relatively easy and straightforward, and its flexibility allowed for the incorporation of various design matrices associated with the fixed and random effects, $X$ and $Z$ in the proposed model formulation.

In order to implement MCMC in WinBUGS, the prior distributions for the unknown parameters need to be specified. In this study, a normal prior for the fixed effects, item difficulties ($\delta$), growth parameters ($\beta$), and coefficients of the school effects ($w$), as well as an inverse gamma prior for the time-level residual variance ($\sigma^2$), the student-level residual variance ($\psi^2$), and the school-level residual variances ($\tau^2_1$ and $\tau^2_2$), were assumed. Specifically, the prior distributions for the CCMM-LIRD for the Type I mobility were specified as follows:

$$\delta_i \sim N(0,1), \ i = 0,..., I - 1,$$

$$\beta_k \sim N(0,1), \ k = 0,1,2,$$

$$w_{it} \sim N(0,10^3), \ t = 2,3,4,5,$$

$$w_{2t} \sim N(0,10^3), \ t = 4,5,$$

$$\sigma^2 \sim \text{Inverse-Gamma}(10^{-3},10^{-3}),$$

$$\psi^2 \sim \text{Inverse-Gamma}(10^{-3},10^{-3}),$$

$$\tau^2_1 \sim \text{Inverse-Gamma}(10^{-3},10^{-3}),$$

$$\tau^2_2 \sim \text{Inverse-Gamma}(10^{-3},10^{-3}).$$

Note that non-informative priors were specified for the coefficients of the school effects and the variances of the random effects. For the item difficulties ($\delta$) and growth parameters ($\beta$), a mildly informative prior, normal distribution with mean zero and variance 1, was set to make the fitting procedures more stable by providing rough bounds on the model parameters (Bolt, Cohen, & Wollack, 2002; Cho & Cohen, 2010).
Similarly, normal priors for the regression parameters and inverse gamma priors for the variance components were specified for the CCMM-LIRD for the Type II mobility. In particular, the model for the Type II mobility, students’ school membership at each time point and school switching patterns were specified (for details, see the WinBUGS code in Appendix A). For all of the models considered in this study, three chains with dispersed starting values were run with 5,000 iterations after a burn-in of 5,000 iterations. Convergence of the three chains was examined using the \( \hat{R} \) index proposed by Gelman and Rubin (1992) with a critical value of 1.01. In addition, the deviance information criterion (DIC; Spiegelhalter, Best, Carlin, & Van Der Linde, 2002) which is a fit index used in Bayesian model selection was used to compare model fit.

### 2.3. Simulation Study

#### 2.3.1. Type I: Data Generation

In order to simulate data with a Type I mobility, students were assumed to have moved from middle school to high school between occasion 2 and occasion 3 over five occasions as in the case of KYPS. The data were generated using the CCMM-LIRD for the Type I with the two-piece linear growth model, Equations (2.12) and (2.13). The number of items \( I \) and measurement occasions \( T \) were set as 10 and 5, respectively, and the two time-related variables, \( d_{1t} \) and \( d_{2t} \), took on the values in Table 2-1. The level 2 residual \( \epsilon_{ijmh} \) was generated from a normal distribution with mean zero and variance 0.4 \( (\sigma^2 = 0.4) \). The student-specific random effect \( \zeta_{0j} \) was generated from a normal distribution with mean zero and variance \( \psi^2 = 0.2 \). The random effects of middle schools, \( \gamma_{0m} \), was generated to be normally distributed with mean zero and variance \( \tau_1^2 = 0.2 \). Likewise, the random effects of high schools, \( \eta_{0h} \), were generated from a normal distribution with mean zero and variance \( \tau_2^2 = 0.2 \), independently of \( \zeta_{0j} \) and \( \gamma_{0m} \). In addition, the varied coefficients for the school effects, to be specific, the decreasing effects of middle schools \( Z_2 = (1, 0.8, 0.6, 0.4, 0.2)' \) and the increasing effects of high schools, \( Z_3 = (0, 0, 1, 1.2, 1.4)' \), were specified. The fixed intercept and slopes for the students while attending middle school and high school were \( \beta_0 = 0.1, \beta_1 = 0.1 \) and \( \beta_2 = 0.2 \), respectively. The item difficulty parameters were generated from a normal distribution of mean zero and variance 1, \( \delta_1 \sim N(0, 1) \) \( (i = 1, \ldots, 9) \) and \( \delta_{10} = -\sum_{i=1}^{9} \delta_i \).

In a Type I mobility, since student school membership changes simultaneously from middle school to high school, combinations of middle school and high school membership for each student need to be generated. In large-scale surveys, it is common to employ a multistage sampling method, in which clusters are sampled first and then units in the cluster are sampled. For example, in educational surveys, school districts are sampled first, schools from each selected district are sampled next, and then
students in every selected school are sampled. In the case of a Type I mobility, the students were sampled when they attended middle school, thus, the number of middle schools and the number of students per middle school at occasion 1 were controlled by the survey design. However, most longitudinal surveys are observational studies that follow students who graduated from middle school and entered high school during study periods; as a result, the students within a middle school do not necessarily enter the same high school. Because of the sampling design, there are usually more high schools than middle schools and the number of students per school varies across high schools more than across middle schools.

In this study, it was assumed that there were 10 school districts and 10 middle schools per school district were selected. For each middle school, 30 students were sampled at the first occasion. Thus, the total number of students and middle schools were \( J = 3,000 \) and \( M = 100 \), respectively. Furthermore, the students were assumed to enter high schools located in the same school district and the number of high schools
was eight times greater than the number of middle schools \( (H = 800) \), mimicking the empirical examples (e.g., KYPS). To illustrate, in Figure 2-9, 10 middle schools (MS 1 \( \sim \) MS 10) were selected and 30 students were sampled from each middle school within District 1. Thus, there were 300 students (Student 1 \( \sim \) Student 300) in District 1. Between occasion 2 and occasion 3, 300 students in District 1 entered one of the 80 high schools located in the same district (HS 1 \( \sim \) HS 80). Because the students were assumed to choose high schools randomly, the actual number of chosen high schools varied across districts and the number of students per school also differed across high schools. The R software (R Core Team, 2013) was used to generate data.

2.3.2. Type I: Analysis

Once the data sets were generated, the three-level HGLM-LIRD (M1) and the CCMM-LIRD for the Type I mobility were fitted for each data set. In M1, student school membership was not considered and the data structure followed the strict hierarchy as in Figure 2-3. For the CCMM-LIRD analysis, two different models, one with constant coefficients of the school effects (M2), \( Z_2 = (1, 1, 1, 1)' \) and \( Z_3 = (0, 0, 1, 1, 1)' \), and the other assuming varying coefficients (M3), \( Z_2 = (1, w_{12}, w_{13}, w_{14}, w_{15})' \) and \( Z_3 = (0, 0, 1, w_{24}, w_{25})' \), were employed. Therefore, M3 was the data-generating model and the coefficients \( w_{12}, w_{13}, w_{14}, w_{15}, w_{24}, \) and \( w_{25} \) were also estimated in this model. In other words, M1 was fitted in order to investigate the consequences of ignoring the school-level random effects by assuming \( Z_2 = Z_3 = 0 \) and M2 was considered to investigate the influences of misspecifying the school effects coefficients. Three models were fitted and a total of 30 replicates were made. Bias and root mean square error (RMSE) were used to assess the parameter recovery of each model.

2.3.3. Type I: Results

Across 30 replicates, the estimated DIC values were consistently lower for the generating model, M3, than the other two models, which suggested a 100% correct model identification. Specifically, the DIC values of M3 were smaller than ones of M1 and M2 by more than 5 units which was the minimum cut-off representing a substantial drop to support the better fit (Li, Bolt, & Fu, 2006). The average of the DIC values of M3 across the replicates was 171250.3, and ones of M1 and M2 were 171975.4 and 171389.6, respectively. In general, the CCMM-LIRD fit better than the three-level HGLM-LIRD when students switched schools simultaneously over repeated observations.

In M3, there were 22 parameter estimates, nine item difficulty estimates \( (\delta_i, i = 1, \ldots, 9) \), three growth parameter estimates including the fixed intercept \( (\hat{\beta}_0) \) and slopes during middle school years \( (\hat{\beta}_1) \) and high school years \( (\hat{\beta}_2) \), six coefficients of the
Table 2-2. Bias and RMSE of the Type I Simulation Study

<table>
<thead>
<tr>
<th></th>
<th>M1</th>
<th></th>
<th>M2</th>
<th></th>
<th>M3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
<td>Bias</td>
</tr>
<tr>
<td>δ₁</td>
<td>0.003</td>
<td>0.022</td>
<td>0.003</td>
<td>0.022</td>
<td>0.003</td>
</tr>
<tr>
<td>δ₂</td>
<td>0.002</td>
<td>0.015</td>
<td>0.001</td>
<td>0.015</td>
<td>0.001</td>
</tr>
<tr>
<td>δ₃</td>
<td>-0.001</td>
<td>0.022</td>
<td>-0.001</td>
<td>0.022</td>
<td>-0.001</td>
</tr>
<tr>
<td>δ₄</td>
<td>0.001</td>
<td>0.024</td>
<td>0.002</td>
<td>0.024</td>
<td>0.002</td>
</tr>
<tr>
<td>δ₅</td>
<td>0.004</td>
<td>0.015</td>
<td>0.003</td>
<td>0.015</td>
<td>0.003</td>
</tr>
<tr>
<td>δ₆</td>
<td>-0.003</td>
<td>0.015</td>
<td>-0.003</td>
<td>0.015</td>
<td>-0.003</td>
</tr>
<tr>
<td>δ₇</td>
<td>0.002</td>
<td>0.016</td>
<td>0.002</td>
<td>0.016</td>
<td>0.002</td>
</tr>
<tr>
<td>δ₈</td>
<td>-0.001</td>
<td>0.020</td>
<td>-0.002</td>
<td>0.020</td>
<td>-0.002</td>
</tr>
<tr>
<td>δ₉</td>
<td>-0.002</td>
<td>0.016</td>
<td>-0.003</td>
<td>0.016</td>
<td>-0.002</td>
</tr>
<tr>
<td>β₀</td>
<td>0.009</td>
<td>0.048</td>
<td>0.006</td>
<td>0.047</td>
<td>0.005</td>
</tr>
<tr>
<td>β₁</td>
<td>-0.002</td>
<td>0.017</td>
<td>-0.001</td>
<td>0.016</td>
<td>-0.002</td>
</tr>
<tr>
<td>β₂</td>
<td>-0.004</td>
<td>0.015</td>
<td>-0.001</td>
<td>0.014</td>
<td>0.002</td>
</tr>
<tr>
<td>w₁₂</td>
<td></td>
<td>0.006</td>
<td></td>
<td>0.064</td>
<td></td>
</tr>
<tr>
<td>w₁₃</td>
<td></td>
<td>-0.002</td>
<td></td>
<td>0.060</td>
<td></td>
</tr>
<tr>
<td>w₁₄</td>
<td></td>
<td></td>
<td>-0.014</td>
<td></td>
<td>0.061</td>
</tr>
<tr>
<td>w₁₅</td>
<td></td>
<td></td>
<td>-0.018</td>
<td></td>
<td>0.046</td>
</tr>
<tr>
<td>w₂₄</td>
<td></td>
<td></td>
<td>0.014</td>
<td></td>
<td>0.069</td>
</tr>
<tr>
<td>w₂₅</td>
<td></td>
<td>0.010</td>
<td></td>
<td>0.073</td>
<td></td>
</tr>
<tr>
<td>σ²</td>
<td>0.112</td>
<td>0.113</td>
<td>0.025</td>
<td>0.027</td>
<td>0.003</td>
</tr>
<tr>
<td>ψ²</td>
<td>0.151</td>
<td>0.152</td>
<td>-0.003</td>
<td>0.012</td>
<td>0.002</td>
</tr>
<tr>
<td>τ₁</td>
<td>-0.115</td>
<td>0.116</td>
<td>0.002</td>
<td>0.040</td>
<td></td>
</tr>
<tr>
<td>τ₂</td>
<td>0.094</td>
<td>0.096</td>
<td>-0.001</td>
<td>0.020</td>
<td></td>
</tr>
</tbody>
</table>

school effects (\( \hat{w}_{12}, \hat{w}_{13}, \hat{w}_{14}, \hat{w}_{24}, \hat{w}_{25} \)) and four variance component estimates of the random effects (\( \hat{\sigma}^2, \hat{\psi}^2, \hat{\tau}_1^2, \hat{\tau}_2^2 \)). As shown in Table 2-2, the bias of M3 ranged in magnitude from -0.018 to 0.014 and the RMSE ranged from 0.011 to 0.073. Even though the bias values of the \( \hat{w}_{1t} \) and \( \hat{w}_{2t} \) were slightly greater than those of the other estimates, none of the bias estimates were significantly different from zero at the 5% level, according to the one-sample t-tests. These results suggest that the estimates of the generating model were unbiased.

Comparisons with the results other two models reveal that the estimates of the fixed effects remained unbiased under M1 and M2 as well. The bias of the fixed effect parameters in M1 and M2 was negligible, indicating that the estimates of the fixed effects were unaffected when the school-level random effect was ignored (M1) and the incorrect coefficients associated with the school-level random effect were assumed (M2). The variance estimates of the random effects, however, were affected by the
misspecification of the school-level random effects. Under M1, assuming $Z_2 = Z_3 = 0$ resulted in overestimation of the time-level residual variance ($\hat{\sigma}^2$) and the variance of the student-level random effect ($\hat{\psi}^2$). On the other hand, under M2, the time-level residual variance was slightly overestimated compared to that in M3, but the estimated values of the bias were not as great as those in M1, and the variance of the student-level random effect was not biased. A difference between M2 and M3 was also found in the variance estimates of the school-level random effects. M2 yielded underestimation of the variance of the middle school random effect ($\hat{\eta}^{2}_{1}$) and overestimation of the variance of the high school random effect ($\hat{\eta}^{2}_{2}$), when compared to M3.

2.3.4. Type II: Data Generation

The data with a Type II mobility were generated using the CCMM-LIRD with the linear growth model, Equations (2.15) and (2.16). The number of items and measurement occasions were set as 10 ($I = 10$) and 3 ($T = 3$) respectively. The time variable, $d_t$, took on the values of 0, 1, and 2 corresponding to occasions 1, 2, and 3. The level 2 residual $\epsilon_{ij(t)}$ was generated from a normal distribution with mean zero and variance 0.4 ($\zeta^2 = 0.4$). The student-specific random effect $\delta_{0j}$ and the school-specific random effect $\nu_{0k}$ were generated to be normally distributed with mean zero and variance 0.2 ($\psi^2 = 0.2$ and $\eta^2 = 0.2$). The fixed intercept and slope of the growth trajectories were assumed as $\beta_0 = 0.4$ and $\beta_1 = 0.2$. The item difficulty parameters generated for the Type I simulation were used.

A few conditions related to the cross-classified and multiple membership models have been considered in previous simulation studies, for example, the number of schools, the number of students per school, the magnitude of the variance of random effects, the intra-class correlations and the mobility rate. Among these conditions, the mobility rate was the most significant factor influencing the observed bias (Chung & Beretvas, 2011; Grady & Beretvas, 2010; Luo & Kwok, 2012). For this reason, mobility rates were specified as 10% and 20% in this simulation study. It was assumed that there were 100 schools ($S = 100$) at the first occasion and 30 students were assigned to each school. Therefore, there were 3,000 students ($J = 3,000$) and a randomly chosen 10% or 20% out of 3,000 students moved to another school between occasion 1 and occasion 2 as well as between occasion 2 and occasion 3. Consequently, there were three types of students with school membership: those who remained in the same school across all occasions, those who moved to a different school once either at occasion 2 or at occasion 3, and those who switched schools two times at occasion 2 and occasion 3.

2.3.5. Type II: Analysis
After the data were generated, each data set was analyzed using three models, the three-level HGLM-LIRD that ignored student school membership (M1), the four-level HGLM-LIRD that assumed that the students did not switch schools (M2), and the CCMM-LIRD for the Type II mobility, the model used to generate the data (M3). In M1, the strict three-level data structure was assumed, in which responses were nested within the measurement occasion and in which the occasions are nested within the student. On the other hand, the students were assumed to remain within the same school assigned at occasion 1 over repeated occasions in M2; therefore, some students had the wrong school membership at occasion 2 and occasion 3. In M3, the student’s correct membership in schools, which varied across measurement occasions, was considered, and the effects of the schools on student growth were investigated. For the students who switched schools over time, in M3, an equal interval between consecutive occasions was assumed for the coefficients associated with the school-level random effect \( \lambda_{ijk} \), as previously illustrated. As in the simulation study for the Type I, three models were fitted and a total of 30 replicates were made.

2.3.6. Type II: Results

Similar to the results of the first simulation study, the generating model, M3, had the smallest DIC values across the 30 replicates under the two mobility conditions and the differences of the DIC values from M1 and M2 were significant. Under the 10% mobility rate condition, the average DIC value of M3 was 101555.2, those of M1 and M2 were 101829.7 and 101589, and the average DIC values were 101765.9 (M1), 101540.4 (M2) and 101477.5 (M3) under the 20% condition. Hence, the CCMM-LIRD was the better-fitting model than the three-level and four-level HGLM-LIRDS, in which the student time-varying school membership was not modeled adequately.

The bias and RMSE of the fixed and random effect parameters of the three models are listed in Table 2-3. The results had similarities with those in the Type I simulation study. To be specific, under the two mobility conditions, the estimated bias values of the fixed effect parameters, including the item difficulties and the fixed intercept and slope of the growth line, were small and acceptable in M3 as well as in the school-level random effect. However, the impacts on the variance estimates of the random effects were different from those of the first simulation. The time-level residual variance estimate \( (\hat{\sigma}^2) \) was also unbiased in M1 and M2, and the variance estimate of the student-level random effect \( (\hat{\psi}^2) \) was overestimated only in M1. In M2, the variance of the school-level random effect \( (\hat{\tau}^2) \) was underestimated and the estimated bias increased alongside the mobility rate, which was augmented from 10% to 20%. In M3, none of the bias estimates were significantly different from zero at the 5% level when one-sample t-tests were used. In sum, the results indicate that the appropriate modeling of multiple school membership through the CCMM-LIRD for the Type II mobility did not yield biased estimates in the 10% and 20% mobility rate conditions.
Table 2.3. Bias and RMSE of the Type II Simulation Study

<table>
<thead>
<tr>
<th>M1</th>
<th>M2</th>
<th>M3</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
</tr>
<tr>
<td>10% mobility rate</td>
<td>10% mobility rate</td>
<td>10% mobility rate</td>
</tr>
<tr>
<td>( \delta_1 )</td>
<td>-0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>( \delta_2 )</td>
<td>-0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>( \delta_3 )</td>
<td>-0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>( \delta_4 )</td>
<td>-0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>( \delta_5 )</td>
<td>-0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>( \delta_6 )</td>
<td>-0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>( \delta_7 )</td>
<td>-0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>( \delta_8 )</td>
<td>-0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>( \delta_9 )</td>
<td>-0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>( \beta_0 )</td>
<td>-0.18</td>
<td>0.18</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>-0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>-0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>( \gamma_2' )</td>
<td>-0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>( \gamma_2'' )</td>
<td>-0.007</td>
<td>0.007</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>M1</th>
<th>M2</th>
<th>M3</th>
</tr>
</thead>
<tbody>
<tr>
<td>20% mobility rate</td>
<td>20% mobility rate</td>
<td>20% mobility rate</td>
</tr>
<tr>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
</tr>
<tr>
<td>( \delta_1 )</td>
<td>-0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>( \delta_2 )</td>
<td>-0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>( \delta_3 )</td>
<td>-0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>( \delta_4 )</td>
<td>-0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>( \delta_5 )</td>
<td>-0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>( \delta_6 )</td>
<td>-0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>( \delta_7 )</td>
<td>-0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>( \delta_8 )</td>
<td>-0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>( \delta_9 )</td>
<td>-0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>( \beta_0 )</td>
<td>-0.18</td>
<td>0.18</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>-0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>-0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>( \gamma_2' )</td>
<td>-0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>( \gamma_2'' )</td>
<td>-0.007</td>
<td>0.007</td>
</tr>
</tbody>
</table>

27
2.4. Empirical Data Study

2.4.1. Korean Youth Panel Survey (KYPS)

Data Source The Korean Youth Panel Survey (KYPS) data, which was collected by the National Youth Policy Institute in South Korea, was used as the first example of the Type I mobility. The students in the survey were second-year middle school students in Korea as of April 1, 2003, and stratified multi-stage cluster sampling was conducted to compose a representative sample. The first survey was administered in 2003 and the students were followed every year from 2004 to 2007. Because the students graduated from their middle schools and moved to high schools between the second and third measurement occasion, the first two occasions were nested within the middle schools, the last three were nested within the high schools, and the data structure of a Type I mobility was found in the KYPS data.

The dependent variables of interest were the responses on 14 items: 7 items intended to measure student maturity regarding specific occupation selection and 7 items regarding decision related to the students’ future career path in general. For example, the contents of the items for occupation selection are given in Table 2-4 and the same items were used to measure maturity in deciding upon a career path (Item 8 ~ Item 14) (NYPI, 2009). The items were negatively stated, the 5-point Likert-type responses were dichotomized; “strongly disagree” and “disagree” were recorded as 1 and “strongly agree,” “agree,” and “neutral” as 0. To examine the effects of schools on the growth of student maturity in deciding upon a job and future plans, the sample of 2,582 students with full information on school identification at each measurement occasion and complete data on the dependent variables were selected for analysis. Specifically, the number of middle schools at the first occasion was 104 and the average number of students in a middle school was 24.83. Two years after the first survey, the students moved simultaneously to 819 high schools and the average number of students in a high school was 3.15.

Results The KYPS data were analyzed using three models, the three-level HGLM-LIRD (M1), the CCMM-LIRD for the Type I mobility assuming the constant effects of schools over time (M2), and the CCMM-LIRD for the Type I mobility with the varying school effects (M3). The parameters and standard error estimates as well as the deviance and DIC values estimated using the three models are given in Table 2-5. The DIC value of M3 was the lowest among the three models and the difference was significant to support a better model fit. In other words, the CCMM-LIRD, which incorporates students’ Type I mobility and allows for the varying effects of schools over time, explains the growth of students’ vocational maturity better than the three-level model in which the students’ school membership was not considered and the CCMM-LIRD that assumed the constant school effects.
Table 2-4. Contents of Items regarding Future Occupation Selection of the KYPS

<table>
<thead>
<tr>
<th>Item No.</th>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 1</td>
<td>I don’t know well my talents.</td>
</tr>
<tr>
<td>Item 2</td>
<td>I don’t know well the types and characteristics of occupations because I don’t have enough information.</td>
</tr>
<tr>
<td>Item 3</td>
<td>I have difficulties in occupation selection because there are lots of things that I want to do.</td>
</tr>
<tr>
<td>Item 4</td>
<td>My plans for future occupation are frequently changing.</td>
</tr>
<tr>
<td>Item 5</td>
<td>I cannot decide my future occupation because I have frequently conflicts with my parents.</td>
</tr>
<tr>
<td>Item 6</td>
<td>It is meaningless to decide future occupation beforehand because the future is uncertain.</td>
</tr>
<tr>
<td>Item 7</td>
<td>I usually follow my parents’ opinion in deciding future occupation.</td>
</tr>
</tbody>
</table>

In M3, there were three types of fixed effect estimates: the item difficulty parameters ($\delta_i$, $i = 1, \ldots, 14$), the fixed intercept and slope parameters in the growth trajectories of the latent variable ($\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$), and the fixed coefficients of the school-level random effects ($\hat{\omega}_{12}$, $\hat{\omega}_{13}$, $\hat{\omega}_{14}$, $\hat{\omega}_{15}$, $\hat{\omega}_{24}$, $\hat{\omega}_{25}$). For the item difficulty estimates, the difficulties of the first 13 items were freely estimated, and the difficulty of the last item was constrained to be equal to the negative sum of the difficulty estimates of the previous items for model identification. As shown in Table 2-5, the item difficulty estimates and associated standard errors were almost identical across the three models. In both the occupation selection ($\delta_1 \sim \delta_7$) and career path items ($\delta_8 \sim \delta_{14}$), the difficulties of the first four items were estimated to be positive and the last three items to be negative. Moreover, the patterns of the estimated item difficulties were very similar in two respects: for example, the students seemed to experience the most difficulty in having access to enough information about occupations and careers ($\delta_2 = 1.090$ and $\delta_9 = 1.066$) and it was not relatively difficult for them to resolve conflicts with parents ($\delta_5 = -1.013$ and $\delta_{12} = -1.076$). These results can provide suggestions for teachers and parents to help the youth prepare for their future.

The regression coefficients of the two-piece linear growth model suggest that on average student awareness and preparation for future plan increased more quickly while attending high school ($\hat{\beta}_2 = 0.247$) than while in middle school ($\hat{\beta}_1 = 0.130$). In addition to the item difficulties and growth parameters, the coefficients of the school effects were estimated in M3, which implies that the school effects on the students’ vocational maturity change over time. Given that the coefficient of the middle school

---

1 The items were originally administered in Korean and they were translated in English and provided by the National Youth Policy Institute (NYPI, 2009).
Table 2-5. Results from the KYPS Vocational Maturity Data Analysis

<table>
<thead>
<tr>
<th></th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>SE</td>
<td>Est.</td>
</tr>
<tr>
<td>(\delta_1)</td>
<td>0.778</td>
<td>0.02</td>
<td>0.778</td>
</tr>
<tr>
<td>(\delta_2)</td>
<td>1.089</td>
<td>0.02</td>
<td>1.090</td>
</tr>
<tr>
<td>(\delta_3)</td>
<td>0.399</td>
<td>0.02</td>
<td>0.400</td>
</tr>
<tr>
<td>(\delta_4)</td>
<td>0.081</td>
<td>0.02</td>
<td>0.082</td>
</tr>
<tr>
<td>(\delta_5)</td>
<td>-1.012</td>
<td>0.02</td>
<td>-1.013</td>
</tr>
<tr>
<td>(\delta_6)</td>
<td>-0.977</td>
<td>0.02</td>
<td>-0.978</td>
</tr>
<tr>
<td>(\delta_7)</td>
<td>-0.483</td>
<td>0.02</td>
<td>-0.484</td>
</tr>
<tr>
<td>(\delta_8)</td>
<td>0.881</td>
<td>0.02</td>
<td>0.881</td>
</tr>
<tr>
<td>(\delta_9)</td>
<td>1.066</td>
<td>0.02</td>
<td>1.066</td>
</tr>
<tr>
<td>(\delta_{10})</td>
<td>0.402</td>
<td>0.02</td>
<td>0.403</td>
</tr>
<tr>
<td>(\delta_{11})</td>
<td>0.095</td>
<td>0.02</td>
<td>0.096</td>
</tr>
<tr>
<td>(\delta_{12})</td>
<td>-1.075</td>
<td>0.02</td>
<td>-1.076</td>
</tr>
<tr>
<td>(\delta_{13})</td>
<td>-0.787</td>
<td>0.02</td>
<td>-0.788</td>
</tr>
<tr>
<td>(\delta_{14})</td>
<td>-0.456</td>
<td></td>
<td>-0.456</td>
</tr>
<tr>
<td>(\beta_0)</td>
<td>0.345</td>
<td>0.03</td>
<td>0.349</td>
</tr>
<tr>
<td>(\beta_1)</td>
<td>0.093</td>
<td>0.02</td>
<td>0.083</td>
</tr>
<tr>
<td>(\beta_2)</td>
<td>0.191</td>
<td>0.02</td>
<td>0.191</td>
</tr>
<tr>
<td>(w_{12})</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(w_{13})</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(w_{14})</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(w_{15})</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(w_{24})</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(w_{25})</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\sigma^2)</td>
<td>1.157</td>
<td>0.03</td>
<td>1.113</td>
</tr>
<tr>
<td>(\psi^2)</td>
<td>1.081</td>
<td>0.04</td>
<td>0.970</td>
</tr>
<tr>
<td>(\tau^2_1)</td>
<td></td>
<td></td>
<td>0.080</td>
</tr>
<tr>
<td>(\tau^2_2)</td>
<td></td>
<td></td>
<td>0.207</td>
</tr>
</tbody>
</table>

Deviance 177715.9  177697.5  177683
DIC 187477  187427  187369

effects was set to one at time 1 \((w_{11} = 0)\), the estimated coefficients represent how middle schools contribute to the latent variable at the current time point compared to the initial observation. The estimated coefficients of the middle school effects at times 2 and 3 suggested a decline of school effects; however, they did not differ significantly from zero at the 5% level. However, at times 4 and 5, the coefficients were negative, which means that the middle schools contributed inversely after the students had moved to the high schools and these estimates were significant at the 5% level. More
specifically, a positive middle school effect contributes negatively and a negative middle school effect contributes positively to the latent variable at occasion 4 and 5 (McCaffrey et al., 2004). In contrast to the middle school effects, the coefficients of the high school effects were greater than 1 at times 4 and 5, therefore, the high school effects increased over time.

Modeling student promotion from the middle schools to the high schools in M3 enabled us to include the random effects of the middle schools and high schools related to the intercept. In M2 and M3, the variability of the random effects between the middle schools and between the high schools was estimated in addition to the within-student variance and the between-student variance. In M3, the estimated within-student variance ($\hat{\sigma}^2 = 1.081$) and between-student variance ($\hat{\psi}^2 = 1.071$) were smaller than the estimates in M1 and, compared to the estimates in M2, incorporating the varying coefficients of the school effects resulted in smaller estimates of the between-middle school variance ($\hat{\tau}_1^2 = 0.067$) and the between-high school variance ($\hat{\tau}_2^2 = 0.111$). In general, the variance estimates of the random effects in M3 suggest that the between-student variance and the within-student variance were greater than the between-middle school and between-high school variance, and there was more variability of the random effects between the high schools than between the middle schools, indicating that high schools have more influence on students’ growth of career-related preparation than middle schools.

2.4.2. National Educational Longitudinal Study (NELS: 88)

Data Source Another example of the cross-classified data structure caused by the Type I mobility is found in the National Educational Longitudinal Study (NELS:88) data, in which eighth grade students moved to high schools after the base year survey. Because of the data structure, the NELS:88 data have been analyzed using the cross-classified model in previous studies (e.g., Meyers & Beretvas, 2006; Palardy, 2010; Shi, Leite, & Algina, 2010) with a focus on continuous outcomes such as test scores during 10th grade. In this study, the responses on 13 self-esteem items listed in Table 2-6 (McLaughlin, Cohen, & Lee, 1997) over the three in-school waves of the data collection (eighth graders in the spring of 1988, sophomores in the spring of 1990, and seniors in the spring of 1992) were analyzed using the proposed CCMM-LIRD. For the positively stated items, “strongly agree” and “agree” were recorded as 1 and “strongly disagree,” and “disagree” as 0 and the responses were recorded in reverse for the negatively stated items. The final sample size was 4,799 with 269 middle schools and 401 high schools. The average number of students per school was 17.84 in middle school and 12.59 in high school.
<table>
<thead>
<tr>
<th>Item No.</th>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 1</td>
<td>I feel good about myself.</td>
</tr>
<tr>
<td>Item 2</td>
<td>I don’t have enough control over the direction my life is taking.</td>
</tr>
<tr>
<td>Item 3</td>
<td>In my life, good luck is more important than hard work for success.</td>
</tr>
<tr>
<td>Item 4</td>
<td>I feel I am a person of worth, the equal of other people.</td>
</tr>
<tr>
<td>Item 5</td>
<td>I am able to do things as well as most other people.</td>
</tr>
<tr>
<td>Item 6</td>
<td>Every time I try to get ahead, something or somebody stops me.</td>
</tr>
<tr>
<td>Item 7</td>
<td>My plans hardly ever work out, so planning only makes me unhappy.</td>
</tr>
<tr>
<td>Item 8</td>
<td>On the whole, I am satisfied with myself.</td>
</tr>
<tr>
<td>Item 9</td>
<td>I certainly feel useless at times.</td>
</tr>
<tr>
<td>Item 10</td>
<td>At times I think I am no good at all.</td>
</tr>
<tr>
<td>Item 11</td>
<td>When I make plans, I am almost certain I can make them work.</td>
</tr>
<tr>
<td>Item 12</td>
<td>I feel I do not have much to be proud of.</td>
</tr>
<tr>
<td>Item 13</td>
<td>Chance and luck are very important for what happens in my life.</td>
</tr>
</tbody>
</table>

**Results** The results of fitting the models to the NELS:88 self-esteem data are summarized in Table 2-7. Similar to the KYPS data analysis, the DIC index suggests that the CCMM-LIRD with the varying school effects (M3) fit better than the three-level HGLM-LIRD (M1) and the CCMM-LIRD with the constant effects of schools over time (M2). Again, the item difficulty estimates were not very dissimilar across the three models. Specifically, two items related to judging and evaluating oneself in comparison to others were the easiest among the thirteen items ($\hat{\delta}_4 = -1.288$ and $\hat{\delta}_5 = -1.366$) and two items measuring self-confidence related to ability appeared to be relatively more difficult than the other items ($\hat{\delta}_9 = 2.213$ and $\hat{\delta}_{10} = 1.503$). With respect to the average linear growth trajectory, the mean self-esteem at occasion 1 (eighth grade in middle school) was estimated as $\hat{\beta}_0 = 2.312$ and the estimated fixed growth rate between occasion 2 and occasion 3 ($\hat{\beta}_2 = 0.192$) was much higher than the linear change between occasion 1 and occasion 2 ($\hat{\beta}_1 = 0.014$). The coefficient estimates of the school effects suggest that the middle school effects decreased over time, while at occasion 3 the high schools contributed as they did at occasion 2. In contrast to the KYPS data, there was more variability between the middle schools than between the high schools ($\hat{\tau}^2_1 = 0.149$ and $\hat{\tau}^2_2 = 0.076$).

**2.4.3. Early Childhood Longitudinal Study-Kindergarten Class (ECLS-K)**

**Data source** A goal of the Early Childhood Longitudinal Study-Kindergarten Class (ECLS-K) was to promote the extensive understanding of children’s development from kindergarten to middle school, including academic performance and social-emotional...
For the purpose of anchoring different test forms across time points and
examinees, 14 common items were presented at least in two occasions for the same student from kindergarten to third grade. In this analysis, a subset of the responses on these common items was selected as dependent variables of interest. Among the 14 items, the actual number of items that the students responded to ranged from 5 to 14 on each occasion. A correct response to an item was scored as 1, and an incorrect response was scored as 0. The sample consisted of 4,261 students with their full information about their school identifications at each occasion. At occasion 1, there were 379 schools and the average number of students per school was 11.24. At occasion 2 and occasion 3, there were 381 and 380 schools with 11.18 and 11.21 students on average, respectively. Of the 4,261 students, 3,913 (91.83%) attended the same school throughout the first three measurement occasions, 333 (7.82%) attended two schools, and 15 (0.35%) students attended three schools.

Results Table 2-8 gives a summary of the analysis of the ECLS-K data via the three models: the three-level HGLM-LIRD (M1), the four-level HGLM-LIRD (M2), and the CCMM-LIRD for the Type II mobility (M3). For the three models, the linear growth model with the time variable \( (d_t) \) taking the values of 0, 1, and 2 for the three time points was employed. In M3, students’ multiple school membership was modeled by assigning equal weights to each school attended. Specifically, if a student attended three different schools over the three time points, \( \lambda_{ijk} \) at occasion 3 took the values of 1/3, 1/3 and 1/3 for each school attended at occasion 1, 2, and 3.

As found in the simulation study, M3 was the best-fitting model according to the estimated DIC values, suggesting that the CCMM-LIRD was a more appropriate model when the Type II mobility was encountered in the data than the strict HGLM-LIRD which ignored student school membership (M1) or assumed that students stayed in the same schools over time (M2). Regardless of the differences in the model specification related to the school-level random effect, fitting the three models resulted in similar estimates for the fixed effect parameters, which include the item difficulty estimates \( (\delta_i, i = 1, \ldots, 14) \) and growth parameters \( (\beta_0, \beta_1) \), and the residual variance estimate \( (\sigma^2) \). However, the variance estimate of the student-specific random effect associated with the intercept \( (\psi^2) \) was larger in M1 than the estimates using M2 and M3. In M2, \( \delta^2 \) was almost identical to the estimate in M3, but \( \psi^2 \) was smaller. As a whole, these results were consistent with the findings of the simulation study for the Type II mobility.

2.5. Conclusion and Discussion

Multilevel models assume that the units in the lower levels are nested purely within one and only one unit in the higher levels, for example, that students are clustered within a school and schools are nested within a neighborhood. However, in reality, a number of data structures may not be in accordance with this strict hierarchy and one typical example in longitudinal studies is the case when students move from
### Table 2-8. Results from the ECLS-K Mathematics Data Analysis

<table>
<thead>
<tr>
<th></th>
<th>M1</th>
<th></th>
<th>M2</th>
<th></th>
<th>M3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>SE</td>
<td>Est.</td>
<td>SE</td>
<td>Est.</td>
<td>SE</td>
</tr>
<tr>
<td>δ₁</td>
<td>-1.013</td>
<td>0.03</td>
<td>-1.013</td>
<td>0.03</td>
<td>-1.014</td>
<td>0.03</td>
</tr>
<tr>
<td>δ₂</td>
<td>-2.437</td>
<td>0.03</td>
<td>-2.436</td>
<td>0.03</td>
<td>-2.436</td>
<td>0.03</td>
</tr>
<tr>
<td>δ₃</td>
<td>-1.933</td>
<td>0.03</td>
<td>-1.933</td>
<td>0.03</td>
<td>-1.933</td>
<td>0.03</td>
</tr>
<tr>
<td>δ₄</td>
<td>-1.281</td>
<td>0.03</td>
<td>-1.280</td>
<td>0.03</td>
<td>-1.280</td>
<td>0.03</td>
</tr>
<tr>
<td>δ₅</td>
<td>0.287</td>
<td>0.02</td>
<td>0.287</td>
<td>0.02</td>
<td>0.287</td>
<td>0.02</td>
</tr>
<tr>
<td>δ₆</td>
<td>0.372</td>
<td>0.02</td>
<td>0.372</td>
<td>0.02</td>
<td>0.372</td>
<td>0.02</td>
</tr>
<tr>
<td>δ₇</td>
<td>1.081</td>
<td>0.02</td>
<td>1.080</td>
<td>0.02</td>
<td>1.081</td>
<td>0.02</td>
</tr>
<tr>
<td>δ₈</td>
<td>0.581</td>
<td>0.02</td>
<td>0.581</td>
<td>0.02</td>
<td>0.581</td>
<td>0.02</td>
</tr>
<tr>
<td>δ₉</td>
<td>-1.664</td>
<td>0.03</td>
<td>-1.664</td>
<td>0.03</td>
<td>-1.663</td>
<td>0.03</td>
</tr>
<tr>
<td>δ₁₀</td>
<td>1.152</td>
<td>0.03</td>
<td>1.151</td>
<td>0.03</td>
<td>1.151</td>
<td>0.03</td>
</tr>
<tr>
<td>δ₁₁</td>
<td>1.224</td>
<td>0.03</td>
<td>1.224</td>
<td>0.03</td>
<td>1.224</td>
<td>0.03</td>
</tr>
<tr>
<td>δ₁₂</td>
<td>1.449</td>
<td>0.03</td>
<td>1.448</td>
<td>0.03</td>
<td>1.449</td>
<td>0.03</td>
</tr>
<tr>
<td>δ₁₃</td>
<td>1.687</td>
<td>0.03</td>
<td>1.687</td>
<td>0.03</td>
<td>1.687</td>
<td>0.03</td>
</tr>
<tr>
<td>δ₁₄</td>
<td>0.495</td>
<td>0.02</td>
<td>0.495</td>
<td>0.02</td>
<td>0.495</td>
<td>0.02</td>
</tr>
<tr>
<td>β₀</td>
<td>-1.732</td>
<td>0.02</td>
<td>-1.741</td>
<td>0.03</td>
<td>-1.737</td>
<td>0.03</td>
</tr>
<tr>
<td>β₁</td>
<td>1.456</td>
<td>0.01</td>
<td>1.455</td>
<td>0.01</td>
<td>1.454</td>
<td>0.01</td>
</tr>
<tr>
<td>σ²</td>
<td>0.376</td>
<td>0.02</td>
<td>0.375</td>
<td>0.02</td>
<td>0.373</td>
<td>0.02</td>
</tr>
<tr>
<td>ψ²</td>
<td>0.981</td>
<td>0.03</td>
<td>0.869</td>
<td>0.03</td>
<td>0.869</td>
<td>0.03</td>
</tr>
<tr>
<td>τ²</td>
<td>0.103</td>
<td>0.02</td>
<td>0.119</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Deviance:  115905.8  115917.4  115915.6  
DIC:       122106  122079  122072

In this study, the cross-classified and multiple membership models were developed to incorporate students’ school mobility in the longitudinal item response data. First, two types of school switching observed in longitudinal studies were described, specifically, all of the students switch schools simultaneously at some time point, such as from middle school to high school (Type I), and some of the students change schools at any time during the data collection (Type II). For the Type I mobility, separate school effects of the middle schools and high schools were assumed, and the degree to which schools influence responses was allowed to vary over time. In the Type II, the effects of the multiple schools that the students have attended up to a certain time point were modeled using the appropriate design matrix.

In general, the results of the simulation studies suggest that the appropriate modeling of Type I and Type II mobility for school membership through the use of the CCMM-LIRD yields fairly good recovery of the fixed and random effect parameters. Another primary goal of this study in addition to developing models that handle school switching in longitudinal item response data was to investigate the impacts of
misspecifying school membership. When mobile students were found in the longitudinal data, one option for researchers who rely solely on the traditional multilevel models is to ignore school membership and use the three-level model. In this case, unobserved school effects, shared by the students who attended the same schools, are not modeled properly. Another possible option is to assume that the students stay within the same school by using only the information for the first school they attended; therefore, the possible effects of multiple schools on students who have attended multiple schools cannot be modeled in this approach.

For the Type I simulation study, the three-level model in which school membership was ignored and the CCMM-LIRD that assumed the constant effects of schools over time were compared to the CCMM-LIRD that allowed the varying school effects. For the Type II simulation study, the CCMM-LIRD that incorporated multiple school membership was compared to the three-level model that did not model school membership, and the four-level model, in which the students were assumed to remain within the first school. In both of the two types of mobility, the fixed effect parameters including the item difficulty parameters and growth trajectory parameters were not affected by misspecification of the school effects as shown in previous studies (Chung & Beretvas, 2011; Grady & Beretvas, 2010; Luo & Kwok, 2012; Meyers & Beretvas, 2006). However, the consequences of ignoring or misspecifying the school effects on the variance component estimates of the random effects were dissimilar according to the mobility patterns and model specifications.

Specifically, in the Type I mobility, given that the time-level residuals ($\epsilon_{ijmh}$), student-level random effects ($\zeta_{ij}$), and school-level random effects ($\gamma_{0m}$ and $\eta_{0h}$) are assumed to be normally distributed with a constant variance and independent of each other, the variance of the latent variable $\theta_{ijmh}$ of the data-generating model, Equation (2.13), is written as

$$\text{var}(\theta_{ijmh}) = \psi^2 + w_{1t}^2 \tau_1^2 + w_{2t}^2 \tau_2^2 + \sigma^2.$$  
(2.18)

If school membership is not considered as in M1 of the Type I simulation study by assuming $w_{1t} = w_{2t} = 0$, the variance of $\theta_{ijmh}$ is estimated as $\hat{\psi}^2 + \hat{\sigma}^2$ and both the between-student and within-student variances were overestimated, compared to the true values. In M2 of the Type I simulation study, $w_{1t}$ and $w_{2t}$ take value of 1 or 0 according to their status at a certain time. For example, at occasion 1 and 2, the variance is estimated as $\hat{\psi}^2 + \hat{\tau}_1^2 + \hat{\sigma}^2$ and $\hat{\psi}^2 + \hat{\tau}_1^2 + \hat{\tau}_2^2 + \hat{\sigma}^2$ after occasion 3. Note that $w_{1t}$ is assumed to decrease over time and $w_{2t}$ to increase over time in M3. Thus, in M3, $w_{1t}$ is always less than or equal to and $w_{2t}$ is greater than or equal to the corresponding coefficients in M2. As a consequence of misspecifying the design matrix of the school-level random effects, in M2, the variance component of the middle school random effect was underestimated (associated with larger coefficients than the true values) and the variance of the high school random effect was overestimated.
(associated with smaller coefficients than the true values), while the variance estimates of the time-level residuals and student-level random effects were relatively not biased.

Similarly, in the simulation of the Type II mobility, the variance of the latent variable \( \theta_{t(i,j)} \) of the true model, Equation (2.16), is expressed as,

\[
\text{var}(\theta_{t(i,j)}) = \psi^2 + \tau^2 \sum_{k \text{est}(j)} \lambda_{ijk}^2 + \sigma^2. \tag{2.19}
\]

In M1, \( \lambda_{ijk} \) is assumed to be zero and the variance is estimated as \( \hat{\psi}^2 + \hat{\tau}^2 + \hat{\sigma}^2 \), and only the between-student variance is overestimated than the true value. Under M2, the students were assumed to stay within the first school they attended, thus, \( \lambda_{ijk} \) took a value of 1 associated with a school \( k \) that student \( j \) attended at occasion 1 and the estimated variance is \( \hat{\psi}^2 + \hat{\tau}^2 + \hat{\sigma}^2 \) at any time point \( t \). On the other hand, when student mobility is modeled, \( \lambda_{ijk} \) indicates the relative contributions of school \( k \) on student \( j \) at time \( t \) and \( \lambda_{ijk} \) is less than or equal to 1. Hence, with the existence of mobile students attending multiple schools, \( \sum_{k \text{est}(j)} \lambda_{ijk}^2 \) in the true model is always less than or equal to 1, yielding underestimation of the between-school variance in M2. For the same reason, when there are more mobile students, that is, the mobility rate increases, the degree of underestimation of the between-school variance in M2 increases as shown in the simulation results and previous studies (e.g., Chung & Beretvas, 2011).

The results suggest that when the school-level random effects were not included in the models, as in the three-level models, the between-school variance was redistributed to the lower-levels. More specifically, in the Type I mobility, ignoring the random effects of the middle schools and high schools yielded overestimation of the student-level and time-level residual variances, and the between-school variance was redistributed to the between-student variance in the Type II mobility. Additionally, the use of incorrect design matrices associated with the school-level random effects produced overestimated or underestimated between-school variances. In particular, this result indicates that with more students who switch schools in the Type II mobility, the four-level modeling yielded a smaller estimate of the between-school variance. In sum, ignoring or misspecifying the school-level random effects in analyzing complicated longitudinal item response data could lead researchers to conclude that more or less variability exists than really does. Given that the identification of a substantial variance of the school-level random effects often directs researchers to investigate school characteristics that may explain the variability across schools (Meyers & Beretvas, 2006), it is important to model the cross-classified and multiple school membership appropriately.

Finally, this study was a preliminary investigation of the extensions of the cross-classified and multiple membership models to longitudinal item response data. A major limitation is that the CCMM-LIRD models employed in this study assumed only
random intercepts of the growth trajectories, and that unconditional growth models without any covariates except for the time variable were used. The growth models can be extended by incorporating student-specific random slopes and student-level and school-level explanatory variables.

In addition, real data in educational research may be more complex than the data considered in this study. For instance, in the Type II simulation study, students who switched schools between time points were randomly selected, however, action of switching schools could be associated with student background and school characteristics. One possible factor is student achievement, and previous studies have shown a negative relation between school change and academic achievement (e.g., Heinlein & Shinn, 2000; Rumberger, 2003; Rumberger & Larson, 1998; Temple & Reynolds, 2000).

In order to investigate the impacts of non-random school mobility, a set of simulation studies with a small number of replications were additionally performed, in which the probabilities of switching schools were negatively associated with student’s latent ability, \( \zeta_{0j} \). Hence, students whose latent ability was lower were more likely to change schools, and about 20% of students switch schools between time points across three measurement occasions in total. The results suggest similar patterns to those obtained from previous simulation studies with random school mobility under the 20% mobility condition. While fixed effects parameters including the item difficulties and growth trajectory parameters remained unbiased under the three models (M1, M2, and M3), the variance of student-level random effects (\( \psi \)) was overestimated in M1 and the variance of school-level random effects (\( \tau \)) was underestimated in M2. However, when mobile students were deleted from the sample and the data were analyzed using the strict four-level model, the fixed intercept of the growth trajectory (\( \beta_0 \)) was overestimated and the variance of student-level random effects (\( \psi \)) was underestimated, compared to the estimates in M3. Given that the mobile students were not randomly selected, thus, deleting those students yielded different samples from the population, these results can be explained. However, in the simulation study with random school mobility, fitting the strict four-level model after deleting the mobile students produced negligible bias and RMSE values of all parameters. This is interesting finding and it needs further investigation to draw complete conclusions.

Another complication can be found if students might move from middle schools to high schools and switch schools during middle school years or high school year as well, for which a combined model of the Type I and Type II mobility would be required. Lastly, this study assumed equal intervals between time points for the students who attended multiple schools in the Type II mobility (e.g., for a student who attended three different schools across the three time points, \( \lambda_{ijk} = (1/3, 1/3, 1/3) \)). However, if we have information regarding the duration of studying in a particular school, \( \lambda_{ijk} \) could be specified accordingly. By investigating these additional factors, further research
could reflect the complexity of real data and improve the generalizability of the findings from the current study.
Chapter 3.
Multidimensional Classification of Examinees based on the Mixture Random Weight Linear Logistic Test Model

3.1. Introduction

Mixture item response theory (IRT) models have been developed to represent the possibility that students may not be a homogeneous population as assumed in the conventional IRT models, but a mixture of multiple latent subpopulations or classes. The distinguishing features of the mixture IRT models are that students from distinct populations are qualitatively different (De Boeck, Wilson, & Acton, 2005), and each person’s population membership is unknown; instead, it is a latent variable. Thus, in the mixture IRT models, finding discrete characteristics that define each latent class of examinees is important. Applications of the mixture IRT models in educational and psychological contexts have attempted to enhance our understanding of the differences between examinees in different classes. For example, latent classes differ in their use of strategies for test items (e.g., Bolt, Cohen, & Wollack, 2001; Rost, 1990), developmental stages in task solution (e.g., Draney, Wilson, Gluck, & Spiel, 2008; Wilson, 1989) and individual differences in the presence of the test speededness (e.g., Bolt et al., 2002; De Boeck, Cho, & Wilson, 2011; Meyer, 2010).

Mislevy and Verhelst (1990) incorporated the linear logistic test model (LLTM; Fischer, 1973) into the mixture IRT models by relating characteristics of each class to known features of items through psychological and cognitive theory. The key characteristics that differentiate the LLTM from the Rasch model is that item properties or item design factors are used to explain the differences in difficulty between items. In this study, verbal aggression data (Vansteelandt, 2000) is taken as an example of the LLTM and its extended models for the purpose of identifying latent classes of examinees. In this example data, the items are built based on four factors that describe a person’s propensity of verbal aggression.

The first design factor is related to the Behavior Mode that differentiates between two levels of behavior, wanting to do (termed as Want) and actual doing (termed as Do). This differentiation is meaningful considering that we do not always actually do whatever we want to do. The second design factor is the Situation Type contrasting situations in which someone else is to blame (termed as Other-to-blame) such as missing a bus or train because a bus fails to stop or a clerk gave me wrong information, and situations in which oneself is to blame (termed as Self-to-blame) such as the grocery store closing because I am late or the telephone operator disconnecting
because I do not have enough coins. The second factor reflects the expected tendency that people would display more verbal aggression in other-to-blame situations. The last two design factors, related to the Behavior Type, include three levels, *Curse, Scold*, and *Shout*. The third and fourth factors are the *Blaming* and *Expressing*, which deal with the extent to which respondents ascribe blame and express aggression respectively. Among the three behavior types, cursing and scolding are regarded as blaming and cursing and shouting as expressive.

The LLTM can be employed to explain how these item features influence responses on test with a prior item structure like the verbal aggression data. Suppose that there are *K* item properties. The difficulty of item *i* is expressed in the LLTM as:

\[ \beta^*_i = \sum_{k=0}^{K} X_{ik} \beta_k, \]  

(3.1)

where \( X_{ik} \) is the value of the \( I \times (K + 1) \) design matrix of item *i* on property *k*, and \( \beta_k \) is the coefficient of *k* \((1, \ldots, K)\). For \( k = 0 \), \( \beta_0 \) is the item intercept with \( X_{i0} = 1 \) for all items *i*, and from 1 to *K*, \( X_{ik} \) reflects the pre-specified structure of item properties composing difficulty \( \beta^*_i \), and \( \beta_k \) represents the difficulty of property *k*. In the LLTM, item difficulties are defined as a linear function of the difficulties of item properties and the LLTM is referred to as explanatory item response models with respect to items (De Boeck & Wilson, 2004).

Therefore, under the LLTM, the probability that person *p* gives the correct response on item *i* is written as:

\[ P(Y_{pi} = 1 | \theta_p) = \frac{\exp(\theta_p - \beta^*_i)}{1 + \exp(\theta_p - \beta^*_i)} = \frac{\exp(\theta_p - \sum_{k=0}^{K} X_{ik} \beta_k)}{1 + \exp(\theta_p - \sum_{k=0}^{K} X_{ik} \beta_k)}, \]  

(3.2)

where \( \theta_p \) is the latent ability of person *p* that corresponds to the random intercept following an underlying population distribution (e.g., a normal distribution with mean zero and a constant variance) and \( \beta_k \) is the fixed coefficient of item property *k*. In other words, \( \beta_k \) indicates the contribution of item design feature *k* to item difficulty. Applications of the LLTM can provide a means of evaluating cognitive theories empirically and enable researchers to predict item difficulties such as in rule-based item generation (Embretson, 1998; Freund, Hofer, & Holling, 2008; Geerlings, Glas, & van der Linden, 2011; Hornke & Habon, 1986). If the LLTM is extended to the mixture model, the assumption of one homogeneous population with respect to the latent ability is relaxed and each class is defined using class-specific ability distribution and class-specific item property coefficients, which will be discussed below.
Another extension of the LLTM is the random weights LLTM (RWLLTM; Rijmen & De Boeck, 2002) that relaxes the assumption of invariant effects of item properties by incorporating person-specific random coefficients. In addition to explanatory aspect, the LLTM has the advantages in its parsimony: item difficulties are explained in terms of item features and there are usually fewer item features than items. However, the assumption that item properties explain the item difficulty perfectly and that the effects of the item features are constant for all persons might be too unrealistic and strict in some situations. To overcome the limitations of the LLTM, person-specific random coefficients $\Theta_{pk'}$ are assumed for a subset of $K$ item properties, $K'$ of which coefficients are assumed to vary over persons. Therefore, $X_{is} (s \in K')$ is the element of the sub-matrix of the full design matrix associated with random coefficients (or random slopes) $\theta_{ps} (s \in K')$. For instance, if the random coefficients are assumed for the first and second item properties among four item properties, $K'$ corresponds to {1, 2}, $X_{is}$ is the element of the matrix consisting of the second and third columns of the full design matrix $X$, and $\Theta_{pk'} = (\theta_{p0}, \theta_{p2})'$. In the RWLLTM, $\sum_{s \in K'} X_{is} \theta_{ps}$ is added to the difficulty of item $i$ for person $p$,

$$
\beta_{pi}^{\Theta} = \sum_{k=0}^{K} X_{ik} \beta_k + \sum_{s \in K'} X_{is} \theta_{ps},
$$

(3.3)

Alternatively, given that $\Theta_{pk'}$ are the person-specific random effects, in the RWLLTM, the person ability is a multidimensional parameter, $\Theta_p = (\theta_{p0}, \Theta_{pk'})'$, a set of the random intercept $\theta_{p0}$ and random coefficients $\theta_{ps} (s \in K')$. Thus, the required ability for person $p$ to response item $i$ is formulated as:

$$
\theta_{pi}^{\Theta} = \sum_{s=0}^{S} Z_{is} \theta_{ps},
$$

(3.4)

where $Z_{is}$ is the value of the $I \times (S + 1)$ matrix that appends a constant vector of 1 with the length of $I$ for the random intercept $\theta_{p0}$ and sub-matrix of the design matrix $X$ for the random coefficients $\theta_{ps} (s \in K')$. Particularly, for $s = 0$, $Z_{i0} = 1$ for all items. For $s$ from 1 to $S$, $Z_{is}$ is the same as $X_{is} (s \in K')$ and $S$ is equal to the number of random coefficients $K'$ (e.g., $S = 2$ in the above example). In the RWLLTM, the probability that person $p$ gives the correct response on item $i$ is written as:

---

2 The model title, random weights LLTM, is used as it was proposed in Rijmen and De Boeck (2002), however, in this study, the term, “coefficient”, is also used interchangeably with the identical meaning with “weight” in Rijmen and De Boeck (2002).
$$P(Y_{pi} = 1 | \Theta_p) = \frac{\exp(\theta_{pi}^* - \beta_i^*)}{1 + \exp(\theta_{pi}^* - \beta_i^*)} = \frac{\exp(\sum_{s=1}^{S}Z_{is}\theta_{ps} - \sum_{k=0}^{K}X_{ik}\beta_k)}{1 + \exp(\sum_{s=0}^{S}Z_{is}\theta_{ps} - \sum_{k=0}^{K}X_{ik}\beta_k)}. \quad (3.5)$$

In fact, the model framework (3.5) is a special case of an earlier model, the multidimensional random coefficients multinomial logit model (MRCMLM; Adams, Wilson, & Wang, 1997), in which $Z$ and $X$ correspond to the scoring matrix and design matrix of the MRCMLM respectively. The random effects $\Theta_p$ are assumed to follow a multivariate normal distribution, therefore, the RWLLTM can be considered as a multidimensional extension of the LLTM that includes additional dimensions corresponding to person-specific random effects associated with item properties.

The primary objective of this study is to investigate the use of the mixture RWLLTM (MixRWLLTM) to distinguish a subpopulation of examinees. A mixture extension of the RWLLTM provides a useful tool to identify latent classes that differ in multidimensional aspects, specific latent dimensions defined by item design features as well as a general latent trait. Each class is defined with class-specific ability structure, separate mean and variance-covariance structure, and class-specific item property coefficients. This chapter is organized as follows. First, mixture extensions of the LLTM and the RWLLTM are briefly described with respect to model specifications. Following that, I introduce the MCMC procedure using WinBUGS 1.4.3 (Lunn et al., 2000) for parameter estimation of the proposed models. Then the results of the verbal aggression data analysis are presented to show how the MixRWLLTM can be applied to an empirical example. Finally, a simulation conducted to assess parameter recovery and correct identification of class membership of the MixRWLLTM is discussed.

### 3.2. Methods

#### 3.2.1. Mixture Extensions of the LLTM and RWLLTM

The mixture LLTM (MixLLTM) was developed by Mislevy and Verhelst (1990). In their study, each student was assumed to belong to one of a number of exhaustive and mutually-exclusive classes that differ in item-solving strategies. That is, the distinctive characteristics determining each class of students were latent features, for example, item-solving strategies. For each item, the difficulty for each class of students could be explained by known item features through psychological and substantive theory. This is, of course, the prime characteristic of the LLTM.

Combining these two assumptions about the students and items is the rationale of formulating the MixLLTM. In the MixLLTM, similar to the mixture Rasch model (Rost, 1990) in which the Rasch model with class-specific person ability and class-
specific item difficulty parameters is assumed for each latent class, the conditional probability that a person \( p \) endorses item \( i \) under the condition that this person belongs to latent class \( g \) is

\[
P(Y_{pi} = 1 | \theta_{pg}, g) = \frac{\exp(\theta_{pg} - \beta_{ig}^*)}{1 + \exp(\theta_{pg} - \beta_{ig}^*)} = \frac{\exp(\theta_{pg} - \sum_{k=0}^{K} X_{ik} \beta_{kg})}{1 + \exp(\theta_{pg} - \sum_{k=0}^{K} X_{ik} \beta_{kg})}.
\]  

(3.6)

As shown in Equation (3.6), in the MixLLTM, the conditional probability is the same as in the LLTM, but with class-specific ability \( \theta_{pg} \) and class-specific item property coefficient \( \beta_{kg} \), as subscript \( g \) indicates. Due to the class-specific coefficients, the item difficulties become class-specific as well. It is common to assume that the ability follows a normal distribution with class-specific mean and variance, \( \theta_{pg} \sim N(\mu_g, \sigma_{\theta_{pg}}^2) \). Class membership \( g \) is regarded as a latent variable with the class size parameters or the mixing proportions \( \pi_g \), having constructs, \( 0 \leq \pi_g \leq 1 \) and \( \sum_g \pi_g = 1 \). Therefore, each person belongs to one of the classes with the probability \( \pi_g \). The marginal probability of a correct response in the MixLLTM is specified as:

\[
P(Y_{pi} = 1 | \theta_{pg}, g, \pi_g) = \sum_{g=1}^{G} \pi_g P(Y_{pig} = 1) = \sum_{g=1}^{G} \pi_g \frac{\exp(\theta_{pg} - \sum_{k=0}^{K} X_{ik} \beta_{kg})}{1 + \exp(\theta_{pg} - \sum_{k=0}^{K} X_{ik} \beta_{kg})}.
\]  

(3.7)

The MixLLTM is useful to identify distinct classes that differ in a general level of propensity, where each class is defined by class-specific ability distributions and item property parameters. However, it is also possible to assume that classes are distinguished by the degree to which item properties influence the item difficulty as well as the general propensity, and this goal can be achieved by extending the random weights LLTM into a mixture model (Fieuws, Spiessens, & Draney, 2004). Considering the model framework of the RWLLTM and MixLLTM in Equation (3.5) and Equation (3.7), the marginal probability that a person \( p \) endorses item \( i \) in the mixture random weights LLTM (MixRWLLTM) can be represented as:

\[
P(Y_{pi} = 1 | \Theta_{pg}, g, \pi_g) = \sum_{g=1}^{G} \pi_g \frac{\exp(\sum_{x=0}^{S} Z_{ixg} \theta_{pg} - \sum_{k=0}^{K} X_{ik} \beta_{kg})}{1 + \exp(\sum_{x=0}^{S} Z_{ixg} \theta_{pg} - \sum_{k=0}^{K} X_{ik} \beta_{kg})}.
\]  

(3.8)

In each class, as presented in Equation (3.8), there are multiple random effects; class-specific intercept \( \theta_{pg0} \) and class-specific random coefficients of item property \( \theta_{pgk} \), and the classes are defined by random effects \( \Theta_{pg} = (\theta_{pg0}, \theta_{pg1}, \ldots, \theta_{pgS})' \) and fixed
coefficients of item property $\beta_{kg}$. For example, in the case of incorporating just one random coefficient $\theta_{pg}$ in addition to the random intercept $\theta_{pg0}$, the $Z$ matrix is composed of first two columns of $X$ matrix and the random effects $\Theta_{pg} = (\theta_{pg0}, \theta_{pg1})$ are assumed to follow a bivariate normal distribution with constant variance-covariance matrix for each class as:

$$\Theta_{pg} = \begin{bmatrix} \theta_{p0g} \\ \theta_{pg1} \end{bmatrix} \sim \text{MVN}_2\left( \begin{bmatrix} \mu_{0g} \\ \mu_{1g} \end{bmatrix}, \begin{bmatrix} \sigma_{0g}^2 & \sigma_{01g} \\ \sigma_{01g} & \sigma_{1g}^2 \end{bmatrix} \right).$$

(3.9)

### 3.2.2. Estimation

A Markov chain Monte Carlo (MCMC) estimation (which is known to be useful in estimating mixture distributions (Diebolt & Robert, 1994)) was selected to estimate the parameters of the MixLLTM and MixRWLLTM. WinBUGS 1.4.3. (Lunn et al., 2000) software can be used for this purpose. In order to implement the MCMC algorithm using WinBUGS, distributions must be specified for all parameters, which include person-specific ability with class-specific mean and variance, class-specific item property coefficients, group membership and mixture probabilities. Although each parameter has possibly a number of different prior distributions, this study limits its scope to the simple and straightforward commonly-used ones such as the conjugate priors that make the posterior distribution belong to the same family. More specifically, assuming a normal distribution is standard practice for the ability and item parameters, and the conjugate prior for the variance of the normal distribution is the inverse-gamma distribution. It is reasonable to assume that, given mixture probabilities, each individual’s group membership follows a multinomial distribution, and one of the conjugate priors for the mixture probabilities is the Dirichlet distribution (Cho, Cohen, & Kim, 2013; Cohen & Bolt, 2005; Ntzoufras, 2009).

Thus, the following prior distributions were used to estimate the MixLLTM in this study,

$$\beta_{kg} \sim N(0,1), k = 0,\ldots,K, g = 1,\ldots,G,$$

$$\theta_{pg} | \sigma_{0g} \sim N(0, \sigma_{0g}), g = 1,\ldots,G,$$

$$\sigma_{0g}^2 \sim \text{Inverse-Gamma}(1,1), g = 1,\ldots,G,$$

$$g \sim \text{Multinomial}(1, (\pi_1, \pi_2,\ldots, \pi_G)),$$

$$\pi = (\pi_1, \pi_2,\ldots, \pi_G) \sim \text{Dirichlet}(\alpha_1, \alpha_2,\ldots, \alpha_G).$$

For identification, the means of the ability distributions were set to zero for each class. Mildly informative prior distributions for item property coefficients $\beta_{kg}$ and variance of the ability $\sigma_{0g}^2$ were used, and, for mixture probabilities, a non-informative Dirichlet
prior with $\alpha_g = 0.5$ was set as well (Bolt et al., 2001; Cho et al., 2013; Cohen & Bolt, 2005). Therefore, based on the probability and priors, the posterior distribution can be written as

$$P\left(\theta_{pg}, \sigma_{0g}^2, \beta_g, g, \pi_g \mid Y\right) \propto$$

$$P\left(Y \mid \theta_{pg}, \sigma_{0g}^2, \beta_k, g, \pi_g\right) P\left(\theta_{pg} \mid \sigma_{0g}^2\right) P\left(\sigma_{0g}^2 \mid \pi_g\right) P\left(\beta_k \mid \pi_g\right) P\left(\pi_g\right).$$

The only difference between the MixLLTM and MixRWLLTM lies in the ability parameter $\Theta_{pg}$. For the ability parameter $\Theta_{pg}$, which includes $\theta_{p0g}$ and $\theta_{psg}$, a multivariate normal distribution with mean zero and constant variance-covariance matrix $\Sigma_g$ was assumed for each class as in the RWLLTM, and an inverse-Wishart distribution, a conjugate prior of the variance and covariance of the multivariate normal distribution, was assumed for $\Sigma_g$ (Gelman, Carlin, Stern, & Rubin, 2004). The prior and hyper-prior distributions of ability in the MixRWLLTM were as follows:

$$\Theta_{pg} \mid \Sigma_g \sim MVN(0, \Sigma_g), g = 1, \ldots, G,$$

$$\Sigma_g \sim \text{Inverse-Wishart}(R_\theta, r), g = 1, \ldots, G,$$

where $R_\theta$ and $r$ represent the scale matrix and degree of freedom of the inverse-Wishart distribution. The dimensions of the multivariate normal distribution and the parameters of the inverse-Wishart distribution are determined by model specification. For parameters other than the ability, the same prior distributions as the MixLLTM can be assumed. The posterior distribution of the MixRWLLTM is written as

$$P\left(\Theta_{pg}, \Sigma_g, \beta_k, g, \pi_g \mid Y\right) \propto$$

$$P\left(Y \mid \Theta_{pg}, \Sigma_g, \beta_k, g, \pi_g\right) P\left(\Theta_{pg} \mid \Sigma_g\right) P\left(\Sigma_g \mid \pi_g\right) P\left(\beta_k \mid \pi_g\right) P\left(\pi_g\right).$$

### 3.3. Empirical Data Study

#### 3.3.1. Data Source

Verbal aggression data (Vansteelandt, 2000) previously analyzed by De Boeck (2008) as well as Ip, Smits, and De Boeck (2009), was selected to illustrate how the proposed model can be applied to real data (The data can be downloaded from http://bearcenter.berkeley.edu/EIRM/). A total of 316 persons, 243 females and 73 males, responded to 24 items and each item described verbally aggressive reactions in a frustrating situation, as described above. Responses were dichotomized as 0 for “no”, and 1 for “perhaps” or “yes”.

As illustrated previously, the items were built based on the four design factors:
Table 3-1. Coding Scheme for Item Properties in the Verbal Aggression Data

<table>
<thead>
<tr>
<th>Design factor</th>
<th>Coding Scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavior Mode ((k=1))</td>
<td>Do = 1</td>
</tr>
<tr>
<td></td>
<td>Want = 0</td>
</tr>
<tr>
<td>Situation Type ((k=2))</td>
<td>Other-to-blame = 1</td>
</tr>
<tr>
<td></td>
<td>Self-to-blame = 0</td>
</tr>
<tr>
<td>Behavior Type: Blaming ((k=3))</td>
<td>Curse, Scold = 1/2</td>
</tr>
<tr>
<td></td>
<td>Shout = -1</td>
</tr>
<tr>
<td>Behavior Type: Expressing ((k=4))</td>
<td>Curse, Shout = 1/2</td>
</tr>
<tr>
<td></td>
<td>Scold = -1</td>
</tr>
</tbody>
</table>

Behavior Mode (Want vs. Do), Situation Type (Self-to-blame vs. Other-to-blame), Behavior Type: Blaming (Curse, Scold vs. Shout), and Behavior Type: Expressing (Curse, Shout vs. Scold). For example, an item, “A bus fails to stop for me. I would want to curse” describes factors of Want (Behavior Type), Other-to-blame (Situation Type), and Curse (Blaming and Expressing). The four design factors are referred to as the item properties and these item designs enable application of the LLTM and its extended models. The coding scheme for the item properties which designated the values of the design matrix is presented in Table 3-1. Dummy coding was used for the behavior mode and the situation type, in which the Want behavior mode and the Self-to-blame situation type were the reference categories; and contrast coding was used for the behavior type where the overall mean was the reference category. The item design matrix with the constant item predictor \((k=0)\) is given in Appendix B.

3.3.2. Analysis

In this study, a Markov chain Monte Carlo (MCMC) as implemented in WinBUGS is used to extend the LLTM and RWLLTM into mixture models using the verbal aggression data. WinBUGS was run using three chains with different lengths of iterations depending on the model specification: for example, for the LLTM and RWLLTM, three chains were run with 3,000 iterations with a burn-in of 3,000 iterations, and for more complicated models such as the MixLLTM and MixRWMLLTM, 10,000 iterations were made after 10,000 iterations of a burn-in. In order to check convergence, time-series plots are monitored and three chains with differed initial values are specified. Convergence of the three chains is examined using the \(\hat{R}\) indexed proposed by Gelman and Rubin (1992) with a critical value of 1.01.

Furthermore, for ease of interpretation, one random coefficient for the behavior mode \((k = 1)\), \(\theta_{p1}\), is assumed in addition to the random intercept \(\theta_{p0}\) for the random weights models, thus, \(S = 1\) and \(Z\) corresponds to the first two columns of the design matrix \(X\). In the mixture models, two latent classes \(G = 2\) are assumed. Therefore, in the RWLLTM and MixRWLLTM, the ability parameters, \(Theta\) and \(Phi_{pg}\), follow a bivariate normal distribution, and in the MixLLTM and MixRWMLLTM, group membership \(g\) follows a Bernoulli distribution.
Given that the four models considered above are not nested, a likelihood ratio (LR) test is not appropriate to compare the relative fit of the models. Li et al. (2009) examined the performance of model selection indices for mixture dichotomous IRT models in the context of Bayesian estimation. They compared two information-based criteria, Akaike’s (1974) information criterion (AIC) and Schwarz’s (1978) Bayesian information criterion (BIC), and three Bayesian methods including the deviance information criterion (DIC; Spiegelhalter et al., 2002), and found that the BIC selects the true data-generating model better than the other methods based on the simulation results. Hence, this study, for the sake of investigating the goodness of fit of the four models, the AIC and BIC indices are reported, and the BIC is used to determine the better fitting model. Specifically, this study follows Li et al. (2009) to define the AIC and BIC for MCMC estimation as:

\[
\text{AIC} = D(\bar{\xi}) + 2m, \\
\text{BIC} = D(\bar{\xi}) + m(\log N),
\]

where \(D(\bar{\xi})\) is the posterior mean of the deviance, \(\bar{\xi}\) represents all parameters under the model, \(m\) refers to the number of estimated parameters, and \(N\) indicates the sample size.

Another critical issue in mixture IRT modeling is label switching problem (Cho et al., 2013; Li et al., 2009). The first type of label switching occurs across iterations within a single MCMC chain and the second type arises when the latent classes switch over replications or for different initial values. An occurrence of the first type of label switching results in multiple modes of the density for the parameters, thus, the estimated marginal posterior densities are examined in empirical data analysis in order to detect label switching. The second type of label switching is often observed in simulation studies, as detailed below.

### 3.3.3. Results

Table 3-2 summarizes the parameter estimates and corresponding standard errors obtained from applications of the one-class and two-class LLTM and RWLLTM to the verbal aggression data. First of all, under the LLTM, the fixed effect estimates represent the intercept \((\beta_0)\) and the item property coefficients of the four design factors \((\beta_1 \sim \beta_4)\). The estimate of the first design factor was \(\hat{\beta}_1 = 0.670\), suggesting that the probability of being verbally aggressive decreased when actually doing compared to wanting to do. In contrast, the negative estimate of \(\hat{\beta}_2 = -1.023\) indicates that examinees became more verbally aggressive in other-to-blame situations than in self-to-blame situations, as we could expect.

The estimates of the behavior type (e.g., Blaming and Expressing) were -1.358 and -0.701 respectively, indicating that the blaming aspect of a behavior has greater
Table 3-2. Estimates for the One-Class and Two-Class LLTM and RWLLTM

<table>
<thead>
<tr>
<th>Parameter</th>
<th>LLTM</th>
<th>MixLLTM</th>
<th>RWLLTM</th>
<th>MixRWLLTM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$ (Intercept)</td>
<td>0.311 (0.09)</td>
<td>0.317 (0.10)</td>
<td>0.295 (0.21)</td>
<td>0.408 (0.16)</td>
</tr>
<tr>
<td>$\beta_{01}$</td>
<td>0.104 (0.26)</td>
<td>0.451 (0.18)</td>
<td>0.408 (0.16)</td>
<td>0.451 (0.18)</td>
</tr>
<tr>
<td>$\beta_{02}$</td>
<td>0.500 (0.18)</td>
<td>0.451 (0.18)</td>
<td>0.408 (0.16)</td>
<td>0.451 (0.18)</td>
</tr>
<tr>
<td>$\beta_1$ (Do)</td>
<td>0.670 (0.06)</td>
<td>0.723 (0.08)</td>
<td>0.670 (0.06)</td>
<td>0.723 (0.08)</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>1.083 (0.25)</td>
<td>0.802 (0.19)</td>
<td>0.802 (0.19)</td>
<td>0.802 (0.19)</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>0.451 (0.18)</td>
<td>0.736 (0.13)</td>
<td>0.736 (0.13)</td>
<td>0.736 (0.13)</td>
</tr>
<tr>
<td>$\beta_2$ (Other-to-blame)</td>
<td>-1.023 (0.06)</td>
<td>-1.071 (0.06)</td>
<td>-1.023 (0.06)</td>
<td>-1.071 (0.06)</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>-1.011 (0.16)</td>
<td>-0.912 (0.15)</td>
<td>-0.912 (0.15)</td>
<td>-0.912 (0.15)</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>-1.117 (0.11)</td>
<td>-1.129 (0.12)</td>
<td>-1.129 (0.12)</td>
<td>-1.129 (0.12)</td>
</tr>
<tr>
<td>$\beta_3$ (Blaming)</td>
<td>-1.358 (0.05)</td>
<td>-1.421 (0.52)</td>
<td>-1.358 (0.05)</td>
<td>-1.421 (0.52)</td>
</tr>
<tr>
<td>$\beta_{31}$</td>
<td>-2.575 (0.22)</td>
<td>-2.625 (0.22)</td>
<td>-2.625 (0.22)</td>
<td>-2.625 (0.22)</td>
</tr>
<tr>
<td>$\beta_{32}$</td>
<td>-0.603 (0.12)</td>
<td>-0.608 (0.14)</td>
<td>-0.608 (0.14)</td>
<td>-0.608 (0.14)</td>
</tr>
<tr>
<td>$\beta_4$ (Expressing)</td>
<td>-0.701 (0.05)</td>
<td>-0.734 (0.05)</td>
<td>-0.701 (0.05)</td>
<td>-0.734 (0.05)</td>
</tr>
<tr>
<td>$\beta_{41}$</td>
<td>-1.078 (0.15)</td>
<td>-1.039 (0.13)</td>
<td>-1.039 (0.13)</td>
<td>-1.039 (0.13)</td>
</tr>
<tr>
<td>$\beta_{42}$</td>
<td>-0.487 (0.09)</td>
<td>-0.542 (0.10)</td>
<td>-0.542 (0.10)</td>
<td>-0.542 (0.10)</td>
</tr>
<tr>
<td>$\sigma_0^2$</td>
<td>1.820 (0.18)</td>
<td>2.206 (0.25)</td>
<td>2.206 (0.25)</td>
<td>2.206 (0.25)</td>
</tr>
<tr>
<td>$\sigma_{01}^2$</td>
<td>2.919 (0.91)</td>
<td>3.559 (0.94)</td>
<td>3.559 (0.94)</td>
<td>3.559 (0.94)</td>
</tr>
<tr>
<td>$\sigma_{02}^2$</td>
<td>1.588 (0.43)</td>
<td>1.989 (0.49)</td>
<td>1.989 (0.49)</td>
<td>1.989 (0.49)</td>
</tr>
<tr>
<td>$\sigma_1^2$</td>
<td>1.005 (0.18)</td>
<td>2.044 (0.61)</td>
<td>2.044 (0.61)</td>
<td>2.044 (0.61)</td>
</tr>
<tr>
<td>$\sigma_{11}$</td>
<td>0.794 (0.29)</td>
<td>0.794 (0.29)</td>
<td>0.794 (0.29)</td>
<td>0.794 (0.29)</td>
</tr>
<tr>
<td>$\sigma_{12}$</td>
<td>0.794 (0.29)</td>
<td>0.794 (0.29)</td>
<td>0.794 (0.29)</td>
<td>0.794 (0.29)</td>
</tr>
<tr>
<td>$\sigma_{01}$</td>
<td>-0.424 (0.18)</td>
<td>-1.509 (0.59)</td>
<td>-1.509 (0.59)</td>
<td>-1.509 (0.59)</td>
</tr>
<tr>
<td>$\sigma_{011}$</td>
<td>0.025 (0.29)</td>
<td>0.025 (0.29)</td>
<td>0.025 (0.29)</td>
<td>0.025 (0.29)</td>
</tr>
<tr>
<td>$\sigma_{012}$</td>
<td>0.025 (0.29)</td>
<td>0.025 (0.29)</td>
<td>0.025 (0.29)</td>
<td>0.025 (0.29)</td>
</tr>
<tr>
<td>$\pi_1$</td>
<td>0.477 (0.07)</td>
<td>0.482 (0.07)</td>
<td>0.482 (0.07)</td>
<td>0.482 (0.07)</td>
</tr>
<tr>
<td>AIC</td>
<td>7593.6</td>
<td>7196.5</td>
<td>7297.5</td>
<td>6872.0</td>
</tr>
<tr>
<td>BIC</td>
<td>7616.1</td>
<td>7245.4</td>
<td>7327.6</td>
<td>6935.8</td>
</tr>
</tbody>
</table>
Table 3-3. Estimates of Coefficients for the Behavior Type

<table>
<thead>
<tr>
<th></th>
<th>LLTM</th>
<th>MixLLTM</th>
<th>RWLLTM</th>
<th>MixRWLLTM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{\text{Curse}}$</td>
<td>-1.030</td>
<td></td>
<td>-1.078</td>
<td></td>
</tr>
<tr>
<td>$\beta_{\text{Curse}(1)}$</td>
<td>-1.827</td>
<td>-1.832</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{\text{Curse}(2)}$</td>
<td>-0.545</td>
<td>-0.575</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{\text{Scold}}$</td>
<td>0.022</td>
<td>0.024</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{\text{Scold}(1)}$</td>
<td>-0.210</td>
<td>-0.274</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{\text{Scold}(2)}$</td>
<td>0.186</td>
<td>0.238</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{\text{Shout}}$</td>
<td>1.008</td>
<td>1.054</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{\text{Shout}(1)}$</td>
<td>2.036</td>
<td>2.106</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{\text{Shout}(2)}$</td>
<td>0.360</td>
<td>0.337</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

effects on verbal aggression than the expression aspect (De Boeck & Wilson, 2004). In order to examine the effects of three behaviors, coefficients of Curse, Scold and Shout were calculated based on the coding scheme and the estimates of the third and fourth item properties, as shown in Table 3-3. Among three levels of the behavior type, cursing, the combination of blaming and expressing, was a more likely response and shouting was a least likely response. In the random part, $\hat{\sigma}_0^2$ is the variance estimates of the random intercept ($\theta_{\beta0}$), estimated as 1.820.

The third column of Table 3-2 displays the results of extending the LLTM into the two-class model. The model selection indices, the AIC and BIC, indicate that the two-class LLTM fit better than the one-class model. The difference of model fit is also observed in terms of effect sizes. Under the MixLLTM, two classes of examinees differed in the levels of general propensity of aggressiveness, therefore, the original normal distribution with respect to the random intercept was replaced by a mixture of two normal distributions for each class.

More specifically, the two-class LLTM produced class proportions of approximately 48% in class 1 and 52% in class 2. The variance estimate of the random intercept in class 1 ($\hat{\sigma}_{\beta0}^2$) was greater than in class 2 ($\hat{\sigma}_{\beta0}^2$), which suggests that there was more variability in the general propensity in class 1 than in class 2. Furthermore, two classes differed in the fixed effects of the item properties. In general, the patterns of the estimated difficulties of the item properties in each class were similar to those in the LLTM. The probability of being verbally aggressive decreased when going from wanting to doing in two classes, however, in class 1, the probability decreased more in doing. In addition, examinees in the two classes were more likely to be aggressive in other-to-blame situations than in self-to-blame situations. In class 2, there was a small difference between blaming and expressing, while, in class 1, the effect of blaming was
much greater than expressing, and this resulted in a larger coefficient of shouting in class 1 (see Table 3-3).

In the one-class and two-class RWLLTM, a random coefficient of the behavior mode ($\theta_{p1}$) was incorporated in addition to the random intercept ($\theta_{p0}$). In other words, there were individual differences in the degree of being verbally aggressive for actually doing as well as for the general propensity. Compared to the one-class and two-class LLTM, the one-class RWLLTM had a better fit in terms of the AIC and BIC than the one-class LLTM, which implies that allowing individual differences in the effects of the behavior mode yielded an improved fit, however, the two-class LLTM fit still better. Hence, a mixture of two normal distributions of one random effect was a better solution to a bivariate normal distribution of two random effects. Again, considering the effect-size difference, it can be noted that, in the one-class RWLLTM, the estimated variance of the random coefficient ($\hat{\sigma}_{11}^2$) was smaller than the random intercept ($\hat{\sigma}_{00}^2$). The correlation indicates that the random intercept and coefficient were negatively correlated. The fixed coefficients of item properties were not much different from ones of the one-class LLTM.

Now compare the one-class RWLLTM to the two-class RWLLTM. In terms of the AIC and BIC, the two-class RWLLTM yielded the better model fit than the one-class RWLLTM and two-class LLTM: the two-class RWLLTM was the best-fitting model among the four models, considered in this study. The two-class RWLLTM produced class proportions of approximately 48.2% in class 1 and 51.8% in class 2. The classes were not only defined in terms of the intercept, but also by the coefficient of the behavior mode. Thus, the latent trait was assumed to follow a mixture of two bivariate normal distributions. Even though the estimates for the fixed coefficients of item properties in the two-class RWLLTM were not differentiated much from the two-class LLTM, the two classes did differ in a meaningful way with respect to the random effects.

In particular, in class 1, the estimated variance of the intercept ($\hat{\sigma}_{00}^2$) was greater than the varinace of the random coefficient of the behavior mode ($\hat{\sigma}_{11}^2$), and there was a negative association between the two random effects. The estimated correlation was -0.567 which was significantly different from zero at the 5% level. This negative correlation means that, in class 1, people who have higher propensity toward verbal aggression tend to have to smaller random coefficient for the behavior mode. Accordingly, they were relatively less verbally aggressive in actually doing than in wanting to do. Similar to class 1, in class 2, the variance estimate of the intercept ($\hat{\sigma}_{02}^2$) was greater than the variance estimates of the random coefficient of the behavior mode ($\hat{\sigma}_{12}^2$), even though the estimates in class 2 were smaller than those in class 1. Unlike in class 1, the estimated covariance of the two random effects was a small positive value. The estimated correlation was 0.02, which was not significant. Therefore, in class 2, the
Table 3-4. Gender Compositions in the Two Latent Classes

<table>
<thead>
<tr>
<th>Latent Class</th>
<th>Gender</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
<td>Male</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td>Class 1</td>
<td>110 (71.0%)</td>
<td>45 (29.0%)</td>
<td>155 (49.1%)</td>
<td></td>
</tr>
<tr>
<td>Class 2</td>
<td>133 (82.6%)</td>
<td>28 (17.4%)</td>
<td>161 (50.9%)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>243 (76.9%)</td>
<td>73 (23.1%)</td>
<td>316</td>
<td></td>
</tr>
</tbody>
</table>

general propensity of verbal aggression and the random coefficient of the behavior mode were virtually independent of each other.

In addition to estimates of the item parameters, the variances of the latent ability distributions and the mixing proportions, examinees in mixture IRT modeling are characterized by a parameter that indicates each examinee’s latent group membership as well. The estimated mixing proportions classified 155 (110 females and 45 males) examinees into class 1 and 161 (133 females and 28 males) examinees into class 2 (see Table 3-4). The proportions of females and males were 76.9% and 23.1% in the total sample; however, the proportions in class 1 and class 2 were 71% and 29% as well as 82.6% and 17.4% respectively. This is, there were more males in class 1 and more females in class 2 than in the total sample. The chi-square test of independence indicated that gender was associated with class membership ($p < 0.05$), although the correlation was weak ($\rho = 0.138$).

3.4. Simulation Study

3.4.1. Data Generation

The simulation design mimicked the empirical example of the verbal aggression data described previously. The data were generated from the two-class RWLLTM, in which 1,000 examinees responded to test items designed based on four item properties, as in the empirical application. The simulation design included two test lengths, 24 items and 48 items. For the 24-item condition, the design matrix used for the verbal aggression was assumed. In the case of 48-item, the elements of the design matrix for the first 24 items were repeated for last 24 items.

Keeping the structure of the verbal aggression data, the estimates of the two-class RWLLTM, presented in the fifth column of Table 3-2, were assumed as the true values in the data generation. More specifically, two latent classes with the class size parameters, $\pi_g = (0.482, 0.518)$, were assumed and only one coefficient of the first item design factor was treated as random. In other words, the data generating model was a two-class and two-dimensional model containing one random intercept and one random
coefficient. Thus, the latent traits follow a bivariate normal distribution with class-specific means and variance and covariance matrix. For model identification, the means of the random effects were constrained to be zero in each class. The variance-covariance matrix of the random effects for each class were specified as:

\[
\Theta_{p1} \sim MVN(0, \begin{bmatrix} 3.559 & -1.509 \\ -1.509 & 1.989 \end{bmatrix}), \Theta_{p2} \sim MVN(0, \begin{bmatrix} 2.044 & 0.025 \\ 0.025 & 0.794 \end{bmatrix}).
\]

In addition, the two classes differed in the fixed coefficients of the item properties. The R software (R Core Team, 2013) was used to generate the data and 30 replications were made for each condition of two test lengths.

### 3.4.2. Analysis

Once the data was generated, the two-class RWLLTM, was applied using the MCMC algorithm. As implemented in the empirical data application, WinBUGS was run using three chains with 10,000 iterations after discarding 10,000 burn-in periods. Convergence of the three chains was determined by the Gelman and Rubin (1992) method.

The second type of label switching in mixture IRT modeling, which refers to class switching over replications, was observed in the simulation study described here. For example, if label switching has occurred, class 1 in one replication corresponds to class 2 in the true model, thus, labels of the parameter estimates and group membership need to be switched, such as from class 1 to class 2. Given that we know the true values of the parameters in the simulation study, the detection of label switching is possible by simply comparing the item parameter estimates and estimated group membership with the generating values (Cho et al., 2013; Li et al., 2009). Specifically, in this simulation study, the covariance of the random effects, of which true value in class 1 was negative and larger in the absolute value than one in class 2, was used to detect label switching.

### 3.4.3. Results

In order to investigate the extent to which the generating parameters are recovered from the simulated dataset, the recovery of the simulated fixed and random effect parameters including the fixed intercept and coefficients of the item properties, the variances and covariance of the random effects, and the class mixing proportions as well as the recovery of simulated latent group membership were examined.

After adjusting for label switching, the bias and root mean squared error (RMSE) of the parameters in each class were assessed and reported in Table 3-5. In general, the estimated biases were not substantial under the two conditions of test lengths. According to the one-sample t-test, none of these bias estimates were significantly
Table 3-5. Bias and RMSE of the Simulation Study

<table>
<thead>
<tr>
<th>Class 1</th>
<th></th>
<th>P = 1000, I = 24</th>
<th></th>
<th>P = 1000, I = 48</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True</td>
<td>Bias</td>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
</tr>
<tr>
<td>β01</td>
<td>0.295</td>
<td>0.023</td>
<td>0.076</td>
<td>-0.035</td>
<td>0.079</td>
</tr>
<tr>
<td>β11</td>
<td>0.802</td>
<td>-0.024</td>
<td>0.088</td>
<td>0.008</td>
<td>0.076</td>
</tr>
<tr>
<td>β21</td>
<td>-0.912</td>
<td>-0.004</td>
<td>0.066</td>
<td>-0.012</td>
<td>0.037</td>
</tr>
<tr>
<td>β31</td>
<td>-2.625</td>
<td>-0.001</td>
<td>0.080</td>
<td>-0.010</td>
<td>0.031</td>
</tr>
<tr>
<td>β41</td>
<td>-1.039</td>
<td>0.015</td>
<td>0.044</td>
<td>-0.001</td>
<td>0.029</td>
</tr>
<tr>
<td>σ01^2</td>
<td>3.559</td>
<td>0.006</td>
<td>0.083</td>
<td>-0.007</td>
<td>0.079</td>
</tr>
<tr>
<td>σ11^2</td>
<td>2.044</td>
<td>0.003</td>
<td>0.096</td>
<td>-0.013</td>
<td>0.087</td>
</tr>
<tr>
<td>σ011</td>
<td>-1.509</td>
<td>-0.019</td>
<td>0.090</td>
<td>-0.002</td>
<td>0.079</td>
</tr>
<tr>
<td>π1</td>
<td>0.482</td>
<td>0.015</td>
<td>0.029</td>
<td>0.021</td>
<td>0.027</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Class 2</th>
<th></th>
<th>P = 1000, I = 24</th>
<th></th>
<th>P = 1000, I = 48</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True</td>
<td>Bias</td>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
</tr>
<tr>
<td>β02</td>
<td>0.408</td>
<td>-0.027</td>
<td>0.086</td>
<td>0.015</td>
<td>0.076</td>
</tr>
<tr>
<td>β12</td>
<td>0.736</td>
<td>0.006</td>
<td>0.068</td>
<td>-0.004</td>
<td>0.052</td>
</tr>
<tr>
<td>β22</td>
<td>-1.129</td>
<td>0.017</td>
<td>0.044</td>
<td>0.008</td>
<td>0.034</td>
</tr>
<tr>
<td>β32</td>
<td>-0.608</td>
<td>-0.007</td>
<td>0.043</td>
<td>-0.005</td>
<td>0.033</td>
</tr>
<tr>
<td>β42</td>
<td>-0.542</td>
<td>-0.013</td>
<td>0.034</td>
<td>-0.005</td>
<td>0.025</td>
</tr>
<tr>
<td>σ02^2</td>
<td>1.989</td>
<td>-0.001</td>
<td>0.097</td>
<td>0.021</td>
<td>0.074</td>
</tr>
<tr>
<td>σ12^2</td>
<td>0.794</td>
<td>0.016</td>
<td>0.091</td>
<td>0.001</td>
<td>0.080</td>
</tr>
<tr>
<td>σ012</td>
<td>0.025</td>
<td>0.005</td>
<td>0.083</td>
<td>-0.015</td>
<td>0.083</td>
</tr>
</tbody>
</table>

different from zero at the 5% level. These results suggest that the estimates of the generating model were approximately unbiased. In the 48-item condition, the RMSEs were slightly smaller than those in the 24-item condition.

In addition, the recovery of group membership was assessed by comparing the estimated latent group membership with the generating one, and the percentage of correct identification was evaluated in each replication. The averages of the percentage of correct identification across replications were 86.69% and 93.8% for the 24- and 48-item conditions respectively, which indicates that the recovery of group membership increases as the test length increases.

### 3.5. Conclusion and Discussion

Mixture item response theory models have been proposed as a useful approach
to explore differences on the latent variables among two or more groups in the population, in which the groups might respond to an instrument in significantly different ways. In this sense, it is rather crucial to investigate how and why the differences exist between latent classes in mixture IRT applications. This study examines possible usefulness of the mixture RWLLTM as a means to find subgroups of examinees as well as to improve interpretations of differences between latent classes. In particular, for better understanding of characteristics of latent groups, this study takes advantage of explanatory aspects of the LLTM, in which item design properties are used to explain item difficulties. This study described the conceptual framework of the MixLLTM and MixRWLLTM, and estimation for the proposed models based on the MCMC algorithm. Moreover, practical issues in Bayesian estimation for the mixture IRT models including model selection and label switching were discussed in the empirical data and simulation studies.

The results from the illustrative example using the verbal aggression data indicate that the two-class RWLLTM, which allows two latent classes to have individual differences in a general propensity and the coefficient of the behavior mode, performs best among the considered models. The estimates of the latent variables related to the general propensity (intercept) and the coefficient of the behavior mode for each class in the MixRWLLTM can be graphically displayed as in Figure 3-1. While the general propensity of verbal aggression and the random coefficient of the behavior mode seems to be unrelated to each other in class 2, in class 1 those who have higher general propensity of verbal aggression are less likely to act some verbally aggressive behavior in frustrating situations. In other words, if we solely rely on the LLTM or the RWLLTM, the existence of two classes and differences between classes in their multidimensional aspects would not be detected.

The simulations show that the Bayesian estimation using WinBUGS appears to recover the parameters and group membership of the MixRWLLTM fairly well and that increasing the number of test items seems to yield better recovery of group membership. In sum, the results from the empirical example and simulation study suggest that the MixRWLLTM could be employed for the purpose of multidimensional classification of examinees.

Finally, the chapter ends by addressing limitations of the current study and suggestions for future studies. This study used conjugate and mildly informative prior distributions for the Bayesian estimation in order to make the fitting procedures more stable (Bolt et al., 2002; Cho & Cohen, 2010). In fact, the use of improper priors and diffuse priors yielded a number of traps in WinBUGS. However, given that the specification of prior distributions could have substantial impacts on estimation (Gelman, 2006), it is worth investigating more deeply the use of different prior distributions. Furthermore, the Bayesian approach adopted in this study required substantial computing time for convergence, which is not uncommon in MCMC. To improve the practical use of the proposed model, other software which handles
multidimensional mixture models for discrete data (e.g., LatentGold (Vermunt & Magidson, 2005)) can be considered for future studies. Even though this study was restricted to two latent classes and one random coefficient additional to the intercept, further extensions of the MixRWLLTM such as allowing more than two latent classes and more random coefficients are also possible.
Chapter 4.
Structured Constructs Model for the Continuous Latent Trait with the Discontinuity Parameters

4.1. Introduction

Corcoran, Mosher, and Rogat (2009, p. 37) defined learning progressions as the descriptions of the successively more sophisticated ways of thinking about an important domain of knowledge and practice. Students follow these ways of thinking, as they learn about and investigate a topic over a broad span of time. This comprehensive definition of learning progressions explicates the fundamentals of the learning process and development, that is, the presence of levels or stages and progress from lower levels to higher levels. A construct map approach is a common in which core ideas of learning progressions are embodied with respect to curriculum development and assessment (Wilson, 2005). A construct defines an underlying theoretical object of interest, assumed to be continuous, ranging from one extreme to another, and a construct map is an ordering of qualitatively different levels of performance or competence defined on the continuum of the construct. In other words, the construct is what is to be learned and what is to be measured, and the construct map is a visual representation of learning progressions denoted by ordered levels of development.

An approach based on a single construct map focuses on one fully ordered set of ways of thinking; however, in the context of instruction and learning, situations can often be more complex than this. For instance, Wilson (2012) pointed out the for multiple constructs and hypothesized links between multiple constructs as an important challenge in measurement practice related to learning progressions. For example, suppose that a set of educational test items requires more than one ability for students to give a correct response on items, and hence, that multiple latent ability dimensions are incorporated in statistical analyses, such as in a multidimensional item response model (Reckase, 1985). Multiple latent ability dimensions, which are often assumed to be correlated to each other, can represent multiple constructs in learning progressions. In addition, educational theories can suggest a complicated interplay between levels across constructs that entails measurement models other than the conventional multidimensional models. To illustrate, in the case of the Assessing Data Modeling and Statistical Reasoning (ADM) Project as presented in Figure 4-1 (Diakow, Irribarra, & Wilson, 2011; Lehrer & Wilson, 2011), each of seven columns represent a construct and small blocks within a construct indicate levels of the construct. Arrows connect a

---

3 For more detail, see section 4.4.1 below.
specific level in one construct to a specific level in another construct, which imply hypothesized links between levels across constructs. For example, an arrow connecting level 3 on the Concept of Statistics construct (CoS3) to level 3 on the Chance construct (Cha3) represents a hypothesis such that students cannot reach level 3 on the Chance construct if they have not reached level 3 on the Concept of Statistics construct. In other words, level 3 of the Concept of Statistics construct is a “prerequisite” to attaining level 3 on the Chance construct.

Wilson (2009) proposed a family of structured constructs model (SCM) as a new class of measurement models, which handles complex structures of learning progressions, particularly focusing on hypothesized connections between levels across multiple dimensions. Diakow et al. (2011) explored the SCM under the framework of ordered latent class models. In this approach, levels can be detected via model-based approaches by use of latent classes, where examinees within the same level are assumed to be homogeneous. This study explores an alternative approach to the SCM by placing cut scores along a latent continuum to identify levels of the construct. Given that the proposed model deals with multiple constructs, it still belongs to the class of multidimensional IRT models. Moreover, some constraints, such that respondents in
one level of the first construct are more likely to belong to a particular level of the second construct, are incorporated to model hypothesized links between levels of different constructs as assumed in the ADM project. This study discusses these constraints, which are modeled using discontinuity parameters, in more detail below.

In sum, the main goal of this study is to describe and propose one possible way to formulate a measurement model for complicated learning progressions through the SCM approach based on continuous latent variables, which is an alternative to the latent class model approach. This chapter briefly describes previous frameworks for the SCM based on the latent class model first, and presents an SCM for the latent continuous trait. Subsequently, the results of applying the method to the simulated data and the ADM data are discussed. Lastly, the chapter concludes with discussions and suggestions for further research.

4.2. Structured Constructs Models

4.2.1. Structured Constructs Model based on the Latent Class Analysis

As discussed above, the SCM provides a theoretical framework of measurement models for complicated learning progressions, in which multiple constructs are involved and relations among levels of the constructs are hypothesized in advance. Therefore, two important steps in SCM modeling are to identify levels of learning progressions and classify examinees into the levels, and to incorporate the hypothesized links across the constructs into the measurement models.

Latent class analysis is one explicit way to define the levels within one construct (Lazarsfeld & Henry, 1968). Specifically, latent class analysis primarily aims at finding subgroups of examinees by relating observed variables to a set of discrete latent variables. These categorical latent variables indicate class membership of examinees, which are mutually exclusive and exhaustive to each other. Therefore, applications of latent class analysis to learning progressions suggest that the latent classes can be interpreted as levels in learning progressions, such as proficient and non-proficient levels. After detecting latent groups, each examinee is classified into one of the levels (classes) according to his or her latent class membership, indexed by $g$ (e.g., Junker & Sijtsma, 2001; Maris, 1999). Then, the probability of person $p$ having response vector $y_p$ depends on his or her class membership as followings:

$$\Pr(y_p) = \sum_{g=1}^{G} \pi_g \prod_{i=1}^{I} \Pr(y_{pig} \mid g),$$

(4.1)

where $\pi_g$ represents the probability of person $p$ belonging to class $g$, and $\Pr(y_{pig} \mid g)$ is the conditional probability of giving response on item $i$ of person $p$ in class $g$. In other
words, $\pi_g$ is the parameter indicating the class size or the mixing proportion of the classes, having constraints such as $0 \leq \pi_g \leq 1$ and $\sum_g \pi_g = 1$. Furthermore, given that respondents within the same class are homogeneous in latent class analysis, the conditional probability, $\Pr(y_{ig} | g)$, is the same across respondents who belong to class $g$; therefore, the conditional probability can be written as $\Pr(y_{ig} | g)$.

In addition, ordered latent class models (Croon, 1990) add an ordinal structure of latent classes to latent class analysis, in which classes are ordered from low to high along the (latent) continuum by using inequality constraints on item responses. Consequently, as examinees progress from lower classes to higher classes, the probability of giving a correct answer on items increases. Suppose that there are two latent classes, $g_1$ and $g_2$, if $g_2$ corresponds to the higher proficiency class than $g_1$, then the conditional probability of answering item $i$ in class $g_2$ is always greater than or equal to the conditional probability in class $g_1$, that is:

$$\forall \ g_2 > g_1: \Pr(y_{ig} \mid g_2) \geq \Pr(y_{ig} \mid g_1). \quad (4.2)$$

Considering the increasing aspect in learning progressions from lower levels to higher levels, ordered latent class models are a suitable way to describe learning progressions (Wilson et al., 2012). Diakow et al. (2011) elaborated the SCM under the framework of ordered latent class models and extended into multiple latent variables. To illustrate, for two constructs of the ADM project, Cha and CoS, 24 crossed levels of the two constructs (four levels of CoS $\times$ six levels of Cha) were considered as classes. In this case, each examinee belongs to one of the 24 classes, and $\pi_{r,t}$ denotes the joint probability of belonging to level $r$ on the Cha construct and level $t$ on the CoS construct. Similar to latent class analysis, the probability of having response vector $y_p$ is expressed as:

$$\Pr(y_p) = \sum_{r,t} \pi_{r,t} \prod_{i=1}^{J} \Pr(y_{ig} \mid r,t) = \sum_{r,t} (\pi_r \times \pi_{t|r}) \prod_{i=1}^{J} \Pr(y_{ig} \mid r,t), \quad (4.3)$$

where $\Pr(y_{ig} \mid r, t)$ is the conditional probability of response on item $i$ of persons in level $r$ on CoS and level $t$ in Cha; $\pi_r$ is the marginal probability of being in level $r$ on CoS; and $\pi_{t|r}$ indicates the conditional probability of belonging in level $t$ on Cha given belonging in level $r$ on CoS. Equation (4.3) also shows that the joint probability can be rewritten as a product of the marginal probability and the conditional probability.

Moreover, the links between constructs were expressed using these joint probabilities of class membership. If there is no association between constructs, class membership on Cha does not depend on the status on CoS. As a result, the conditional probability $\pi_{t|r}$ is equal to the marginal probability $\pi_t$. However, hypothesized relations between levels on the two constructs would imply that status on Cha is influenced by status on CoS, as described in Figure 4-1. For example, the hypothesis that CoS3 is required to reach Cha3 can be paraphrased such that it is very unlikely or impossible
for examinees in level 2 on CoS to belong to level 3 or higher levels on Cha. These constraints can be modeled by assuming \( \pi_{3|2} = \pi_{4|2} = \pi_{5|2} = \pi_{6|2} = 0 \). The same principles can be applied to other links between levels across the constructs.

### 4.2.2. Structured Constructs Model for the Latent Continuous Trait

In the SCM based on latent class analysis, levels are identified using latent classes, and examinees within the same level are assumed to be identical with respect to their probabilities of responding items. This section considers an SCM approach based on a latent continuum, rather than assuming each construct as an ordered set of latent classes, which allows differences between persons within the same level.

Let a simple form of the SCM involve a single connection between two constructs. Each construct is assumed to be continuous and to be composed of two levels (e.g., master versus non-master, or proficiency versus non-proficiency). Following the notations of Diakow et al. (2011), the construct from which the link initiates is referred to as the “requirement” and the construct at which the link terminates is referred to as the “target”, and \( \theta_{pR} \) and \( \theta_{pT} \) denote the continuous latent variable of person \( p \)’s proficiency in the requirement and target constructs respectively. Note that in latent class analysis the latent variables were categorical variables indicating class membership of examinees. For simplicity, each item is assumed to relate to one construct as in the between-item multidimensional item response model (Adams, Wilson, & Wang, 1997). Specifically, \( \beta_{iR} \) (\( i = 1, \ldots, I_R \)) indicates the difficulty of the \( i \)th item in the requirement construct and \( \beta_{iT} \) (\( i = 1, \ldots, I_T \)) is the difficulty of the \( i \)th item in the target construct.

Furthermore, if person \( p \) is considered as being proficient in the requirement construct, he or she is assumed to be more likely to be classified into the proficient level on the target construct. This hypothesized link between the two constructs is expressed as an arrow from the proficient level in the requirement construct to the proficient level in the target construct in Figure 4-2. In current study, this link between levels of the two constructs is modeled by the introduction of a discontinuity parameter, which is similar to the saltus parameter (Draney et al., 2008; Wilson, 1989). In the saltus model, as individuals progress from lower levels to higher levels, a sudden spurt or change occurs; consequently discontinuities are inherent in the cognitive developmental levels. Specifically, in the saltus model, classes of persons, to be estimated, represent different development stages or levels, and groups of items are specified to allow persons at or above the developmental stage to have the advantage in answering items in that stage. The saltus parameter, \( \tau_{ck} \), quantifies these discontinuities as additive effects on the item parameters of all items in item group \( k \) when people in group \( c \) respond to those items.
In the SCM framework, hypothesized links between the two constructs are assumed to induce discontinuities. Particularly, in the requirement construct, the probability that person $p$ gives a correct response on item $i$ is written, according to the Rasch model, as:

$$\Pr( y_{piR} = 1 | \theta_{pR} ) = \frac{\exp(\theta_{pR} - \beta_{iR})}{1 + \exp(\theta_{pR} - \beta_{iR})}. \quad (4.4)$$

However, the probability of success on the items in the target construct depends on the ability in the target construct, $\theta_{pT}$, as well as the ability in the requirement construct, $\theta_{pR}$. More specifically, if $\theta_{pR} \geq C_1$, where $C_1$ is the cut score in the requirement construct (to be determined), the probability is augmented by $\delta_1$, and if $\theta_{pR} < C_1$, the probability is augmented by $\delta_2$. In other words, two discontinuity parameters, $\delta_1$ and $\delta_2$, can be considered as advantage (or disadvantage) parameters in the target construct according to the level in the requirement construct. Then, the probability of person $p$’s correct answer on item $i$ in the target construct is expressed as:

$$\Pr( y_{piT} = 1 | \theta_{pT}, \theta_{pR} ) = \frac{\exp(\theta_{pT} + \delta_1 f(\theta_{pR}) + \delta_2 (1 - f(\theta_{pR})) - \beta_{iT})}{1 + \exp(\theta_{pT} + \delta_1 f(\theta_{pR}) + \delta_2 (1 - f(\theta_{pR})) - \beta_{iT})}. \quad (4.5)$$

where $f(\theta_{pR}) = 1$ when $\theta_{pR} \geq C_1$ and $f(\theta_{pR}) = 0$, otherwise, and the cut score $C_1$ may be calculated by using an approximation based on the difficulties of the items, or it may be estimated directly. As in the multidimensional Rasch models, $\theta_{pR}$ and $\theta_{pT}$ are assumed to follow a bivariate normal distribution,

$$\begin{bmatrix} \theta_{pR} \\ \theta_{pT} \end{bmatrix} \sim \text{MVN} \left( \begin{bmatrix} \mu_R \\ \mu_T \end{bmatrix}, \begin{bmatrix} \sigma^2_R & \sigma_{RT} \\ \sigma_{RT} & \sigma^2_T \end{bmatrix} \right).$$
Figure 4-3 plots an illustrative example of the latent ability distributions in the requirement and target constructs. Each solid curve in the two constructs represents the distribution of the latent variables, $\theta_{pR}$ and $\theta_{pT}$, without discontinuity parameters, which consequently follow a bivariate normal distribution with $\mu_R = \mu_T = 0$ and a constant variance-covariance matrix. If the cut score in the requirement construct is specified as zero, $C_1 = 0$, persons whose latent ability in the requirement construct is greater than or equal to zero are classified into the proficient level. A link between the two constructs is assumed to induce discontinuities such that persons who are in the proficient level in the requirement construct are more likely to reach the proficient level in the target construct, as expressed in Figure 4-2. Accordingly, for persons in the proficient level on the requirement construct, a positive discontinuity parameter $\delta_1$ represents the constraint and the (upper) dotted curve in Figure 4-3 displays the distribution of $\theta_{pT}$ which is boosted by $\delta_1$, $\theta_{pT} + \delta_1$. Similarly, the other (lower) dotted curve displays the distribution of $\theta_{pT} + \delta_2$ for persons in the non-proficient level on the requirement construct, which is augmented by a negative discontinuity parameter $\delta_2$. As shown in Figure 4-3, incorporating the discontinuity parameters distinguishes clearly two groups of examinees in the target construct, and the magnitude and sign of the discontinuity parameters are associated with separation of the two distributions.
4.3. Simulation Study

4.3.1. Data Generation

A simulation study was designed to assess the recovery of the parameters of the proposed model. The data were generated using Equation (4.4) and (4.5), and the R software (R Core Team, 2013) was used to generate data. As illustrated in the model framework, two constructs, referred to as the requirement and target constructs, were specified, and examinees were assumed to be classified into one of two levels based on the cut score. The number of examinees was set as 1,000 and the latent abilities of examinees in the two constructs were generated from a bivariate normal distribution,

\[ \Theta_p = \begin{bmatrix} \theta_{pR} \\ \theta_{pT} \end{bmatrix} \sim MVN \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \].

The number of items was 30 for each construct \( (I_R = I_T = 30) \) and item difficulties in each construct were generated from a uniform distribution between -3 and 3. The mean of the item difficulties was constrained to be zero within each construct. In addition, two discontinuity parameters were specified as \( \delta_1 = 1 \) and \( \delta_2 = -1 \). Thus, \( \delta_1 \) can be considered as an advantage for examinees who were in the proficient level on the requirement construct, to reaching the proficient level on the target construct, while \( \delta_2 \) corresponds to a disadvantage for examinees who did not reach the non-proficient level on the requirement construct.

More importantly, five values of the cut score in the requirement construct \( (C_1 = -1, -0.5, 0, 0.5, 1) \) were selected to investigate the impacts of the cut score on the model parameter estimation. Given that the latent ability in the requirement construct was assumed to follow a normal distribution with mean zero and variance 1, about 84\% of examinees were classified into the proficient level in the requirement construct when \( C_1 = -1 \), while about 16\% of examinees were in the proficient level in the case of \( C_1 = 1 \). In other words, different values of the cut score are associated with the percentage of examinees in each level in the requirement construct.

4.3.2. Analysis

After the data were generated, each dataset was analyzed using two models, the conventional multidimensional Rasch model and the proposed SCM for the continuous latent trait with discontinuity parameters. In particular, the multidimensional Rasch model corresponds to a constrained version of the SCM for the continuous latent trait, in which two discontinuity parameters are equal to zero \( (\delta_1 = \delta_2 = 0) \), thus, allowing us to investigate the consequences of ignoring discontinuity parameters. For parameter estimation of the two models, a Bayesian approach using Markov chain Monte Carlo (MCMC) algorithm was implemented in WinBUGS 1.4.3 (Lunn et al., 2000). Specifically, prior distributions were specified for the SCM as follows:
\[
\begin{align*}
\beta_{ir} & \sim N(0, 10^3), \\
\beta_{it} & \sim N(0, 10^3), \\
\delta_1 & \sim N(0, 10^3), \\
\delta_2 & \sim N(0, 10^3), \\
\Theta_p | \Sigma & \sim MVN(0, \Sigma), \\
\Sigma & = \begin{bmatrix}
\sigma_R^2 & \sigma_{RT} \\
\sigma_{RT} & \sigma_T^2
\end{bmatrix} \sim Inverse-Wishart(\Psi, \nu).
\end{align*}
\]

Following conventions in Bayesian item response modeling, a normal distribution was used for the fixed effects parameters, item difficulties and discontinuity parameters, and an inverse-Wishart distribution was specified for the variance and covariance matrix of the latent ability variables in the two dimensions (for detail, see the WinBUGS code in Appendix D). The same prior distributions were specified for the multidimensional Rasch model except for the discontinuity parameters. For all models, three chains with dispersed starting values were run and the convergence of the chains was determined by use of the \(\hat{R}\) index (Gelman & Rubin, 1992) with a critical value of 1.01. For the SCM analysis, 10,000 post-burn-in iterations were used to provide sampled parameter values for posterior distributions, after a conservative burn-in of 10,000 iterations. In the multidimensional Rasch model, 5,000 iterations of post-burn-in were used after 5,000 iterations of burn-in. A total of 30 replicates were made, and bias and root mean square error (RMSE) were reported.

4.3.3. Results

Table 4-1 provides the recovery results in the multidimensional Rasch model and the SCM under the five conditions of the cut score (\(C_1\)). For the item difficulties, the table presents the averages and standard deviations of the estimated bias and RMSE values of 60 items. First of all, comparing the results from the two models, the bias and RMSE values of the item difficulties and the variance of the requirement construct did not differ significantly in two models. However, the SCM with the two discontinuity parameters, were much smaller with respect to other parameters, suggesting that the proposed model appeared to be recovered well.

Additionally, in the SCM model, as the cut score increases from -1 to 1, the RMSE of the first discontinuity parameter (\(\delta_1\)) increases, while the RMSE of the second discontinuity parameter (\(\delta_2\)) decreases with an increasing cut score. As delineated above, given that the different values of the cut score represent the percentage of examinees in the each level of the requirement construct, the recovery of the discontinuity parameters could be influenced by the sample size in each level.
Table 4-1. Bias and RMSE in the Multidimensional Rasch Model and the SCM

<table>
<thead>
<tr>
<th></th>
<th>$C_1 = -1$</th>
<th>$C_1 = -0.5$</th>
<th>$C_1 = 0$</th>
<th>$C_1 = 0.5$</th>
<th>$C_1 = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
<td>Bias</td>
</tr>
<tr>
<td>Multi Rasch</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta^*$</td>
<td>0</td>
<td>0.093</td>
<td>0</td>
<td>0.091</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.020)</td>
<td>(0.024)</td>
<td>(0.020)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>$\sigma^2_R$</td>
<td>0.012</td>
<td>0.050</td>
<td>0.019</td>
<td>0.058</td>
<td>0.008</td>
</tr>
<tr>
<td>$\sigma^2_T$</td>
<td>1.087</td>
<td>1.095</td>
<td>1.675</td>
<td>1.681</td>
<td>1.975</td>
</tr>
<tr>
<td>$\sigma_{RT}$</td>
<td>0.518</td>
<td>0.521</td>
<td>0.767</td>
<td>0.769</td>
<td>0.860</td>
</tr>
<tr>
<td>$\rho_{RT}$</td>
<td>0.201</td>
<td>0.201</td>
<td>0.268</td>
<td>0.268</td>
<td>0.288</td>
</tr>
<tr>
<td>SCM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta^*$</td>
<td>0</td>
<td>0.091</td>
<td>0</td>
<td>0.089</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.019)</td>
<td>(0.022)</td>
<td>(0.018)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>0.004</td>
<td>0.046</td>
<td>-0.017</td>
<td>0.071</td>
<td>0.003</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>-0.006</td>
<td>0.107</td>
<td>0.039</td>
<td>0.093</td>
<td>0.014</td>
</tr>
<tr>
<td>$\sigma^2_R$</td>
<td>-0.013</td>
<td>0.045</td>
<td>-0.002</td>
<td>0.054</td>
<td>-0.009</td>
</tr>
<tr>
<td>$\sigma^2_T$</td>
<td>0.009</td>
<td>0.088</td>
<td>0.024</td>
<td>0.101</td>
<td>0.008</td>
</tr>
<tr>
<td>$\sigma_{RT}$</td>
<td>0.019</td>
<td>0.052</td>
<td>0.043</td>
<td>0.077</td>
<td>0.020</td>
</tr>
<tr>
<td>$\rho_{RT}$</td>
<td>0.020</td>
<td>0.039</td>
<td>0.037</td>
<td>0.059</td>
<td>0.020</td>
</tr>
</tbody>
</table>

* The averages of the bias and RMSE values of estimated difficulties of 60 items are presented in the top row and the numbers in parenthesis correspond to the standard deviations of the bias and RMSE.
Specifically, when $C_1 = -1$, most of examinees were classified into the proficient level (about 84%) in the requirement construct, and the discontinuity parameter associated with examinees in the proficient level ($\delta_1$) was estimated more accurately than one in the case of $C_1 = -1$, in which less examinees (about 16%) were in the proficient level.

Interestingly, the impacts of excluding the discontinuity parameters were most obvious on the variance of the latent ability in the target construct ($\zeta^2_T$) and the covariance of the latent variables in the two constructs ($\zeta_{RT}$). As shown in Table 4-1, the magnitude of the bias and RMSE of $\sigma^2_T$ were fairly large, suggesting that the variance of the latent ability in the target construct was considerably overestimated compared to the true value across all values of the cut score. Similarly, $\sigma_{RT}$ was estimated to be much greater than the true value. As a result, the correlation between the two constructs ($\rho_{RT}$) was overestimated. These results indicate that ignoring the discontinuity parameters yields a wrong conclusion that there is both more variability in the latent variable in the target construct, and a higher association between the latent variables in the two dimensions, than there actually are.

4.4. Empirical Data Analysis

4.4.1. Data Source

A subset of the ADM project data, which consisted of responses to the items of the two selected constructs, the Concept of Statistics (CoS) construct and the Chance (Cha) construct, was analyzed for empirical illustration of the SCM using real data. The CoS construct describes how students develop their concepts of the meaning and uses of statistics, from describing a distribution informally using shape to understanding statistics as measures of summarizing a sampling distribution. The Cha construct represents students’ progression in understanding probability as a measure of uncertainty. Students are expected to understand that chance yields a distribution of outcomes, as they progress to more sophisticated levels in the Cha construct. In this study, being proficient in the CoS construct was regarded as the requirement to reach the proficient level in the Cha construct.

In total, 16 items for the CoS construct and 18 items for the Cha construct were considered. Due to test form design, the number of items administered to each student was not the same. Responses from 489 middle school students who answered at least two items in each construct were analyzed.

4.4.2. Analysis

In order to demonstrate the use of the proposed SCM models, the selected ADM data was analyzed using the multidimensional Rasch model and two models
based on the SCM. As discussed earlier in the simulation study, the multidimensional Rasch model can be considered as the SCM, in which the discontinuity parameters were constrained to be zero.

The first SCM analysis incorporated two discontinuity parameters applied to examinees in the proficient and non-proficient level on the CoS (requirement) construct respectively, as described in Equation (4.4) and (4.5). The items were scored dichotomously and the mean of the item difficulties was constrained to be zero. In addition, the cut score in the CoS construct was specified using the mean of the item difficulties, $C_1 = 0$. Therefore, about 50% of examinees were classified in the proficient level in the CoS construct. In the second SCM analysis, only one discontinuity parameter for the examinees in the proficient level in the CoS construct was modeled by assuming $\delta_2 = 0$. Hence, the model in the Cha construct was reduced to:

$$\text{Pr}(Y_{pT} = 1|\theta_{pT}, \theta_{pR}) = \frac{\exp(\theta_{pT} + \delta_1 f(\theta_{pR}) - \beta_{IT})}{1 + \exp(\theta_{pT} + \delta_1 f(\theta_{pR}) - \beta_{IT})} \quad (4.6)$$

Similar to the simulation study, all models were estimated using WinBUGS with MCMC estimation.

### 4.4.3. Results

The parameters estimates in CoS and Cha of the multidimensional Rasch model and the two SCM analyses are listed in Table 4-2. In the multidimensional Rasch model, for CoS, the estimated item difficulties were between -2.115 and 2.150, and difficulties in Cha items ranged from -2.175 and 2.125. As shown in Table 4-2, in the multidimensional Rasch model, the variances of the latent ability in the two constructs were estimated as 1.872 and 1.805 respectively, and the covariance was estimated as 1.251.

Table 4-2 lists parameter estimates of the two SCM analyses as well. The item difficulties from the multidimensional model and SCM analyses were very much alike, as shown in the simulation results. In particular, the SCM analyses produced more similar estimates of the item difficulties to each other. In the first SCM analysis (SCM 1), in which two discontinuity parameters were incorporated, $\delta_1$ and $\delta_2$ were estimated as 1.178 and 0.687 respectively. These estimated discontinuity parameters suggest that both examinees in the proficient level and non-proficient level on CoS have an advantage in attaining the proficient level on Cha (e.g., positive values of $\delta_1$ and $\delta_2$), however, examinees who were classified into the proficient level in CoS have a greater advantage in Cha, compared to examinees in the non-proficient level in CoS, in terms of magnitude of the discontinuity parameters. Furthermore, in SCM 1, the variance of the latent ability in Cha ($\sigma_{cha}^2$) and the correlation between the two constructs ($\rho_{CoS,Cha}$) were estimated to be smaller than those in the multidimensional
Table 4-2. Parameter Estimates and Standard Errors for the ADM Data

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Multidimensional</th>
<th>SCM 1</th>
<th>SCM 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>SE</td>
<td>Est.</td>
</tr>
<tr>
<td>(\beta_{1CoS})</td>
<td>-0.820</td>
<td>0.16</td>
<td>-0.789</td>
</tr>
<tr>
<td>(\beta_{2CoS})</td>
<td>0.101</td>
<td>0.20</td>
<td>0.090</td>
</tr>
<tr>
<td>(\beta_{3CoS})</td>
<td>-0.173</td>
<td>0.21</td>
<td>-0.166</td>
</tr>
<tr>
<td>(\beta_{4CoS})</td>
<td>-2.115</td>
<td>0.21</td>
<td>-1.998</td>
</tr>
<tr>
<td>(\beta_{5CoS})</td>
<td>-1.925</td>
<td>0.20</td>
<td>-1.824</td>
</tr>
<tr>
<td>(\beta_{6CoS})</td>
<td>-0.894</td>
<td>0.16</td>
<td>-0.859</td>
</tr>
<tr>
<td>(\beta_{7CoS})</td>
<td>1.554</td>
<td>0.17</td>
<td>1.491</td>
</tr>
<tr>
<td>(\beta_{8CoS})</td>
<td>1.238</td>
<td>0.16</td>
<td>1.184</td>
</tr>
<tr>
<td>(\beta_{9CoS})</td>
<td>1.441</td>
<td>0.26</td>
<td>1.346</td>
</tr>
<tr>
<td>(\beta_{10CoS})</td>
<td>-1.446</td>
<td>0.17</td>
<td>-1.382</td>
</tr>
<tr>
<td>(\beta_{11CoS})</td>
<td>0.845</td>
<td>0.20</td>
<td>0.801</td>
</tr>
<tr>
<td>(\beta_{12CoS})</td>
<td>0.343</td>
<td>0.20</td>
<td>0.322</td>
</tr>
<tr>
<td>(\beta_{13CoS})</td>
<td>2.150</td>
<td>0.24</td>
<td>2.014</td>
</tr>
<tr>
<td>(\beta_{14CoS})</td>
<td>-1.772</td>
<td>0.21</td>
<td>-1.667</td>
</tr>
<tr>
<td>(\beta_{15CoS})</td>
<td>0.649</td>
<td>0.17</td>
<td>0.634</td>
</tr>
<tr>
<td>(\beta_{16CoS})</td>
<td>0.824</td>
<td>0.18</td>
<td>0.802</td>
</tr>
<tr>
<td>(\beta_{17CoS})</td>
<td>0.850</td>
<td>0.12</td>
<td>0.833</td>
</tr>
<tr>
<td>(\beta_{18CoS})</td>
<td>-2.175</td>
<td>0.29</td>
<td>-1.984</td>
</tr>
<tr>
<td>(\beta_{19CoS})</td>
<td>-0.685</td>
<td>0.19</td>
<td>-0.652</td>
</tr>
<tr>
<td>(\beta_{20CoS})</td>
<td>-0.796</td>
<td>0.19</td>
<td>-0.752</td>
</tr>
<tr>
<td>(\beta_{11Cha})</td>
<td>1.063</td>
<td>0.16</td>
<td>1.043</td>
</tr>
<tr>
<td>(\beta_{12Cha})</td>
<td>-0.126</td>
<td>0.18</td>
<td>-0.128</td>
</tr>
<tr>
<td>(\beta_{13Cha})</td>
<td>0.198</td>
<td>0.17</td>
<td>0.204</td>
</tr>
<tr>
<td>(\beta_{14Cha})</td>
<td>-0.207</td>
<td>0.18</td>
<td>-0.192</td>
</tr>
<tr>
<td>(\beta_{15Cha})</td>
<td>0.710</td>
<td>0.13</td>
<td>0.696</td>
</tr>
<tr>
<td>(\beta_{16Cha})</td>
<td>0.038</td>
<td>0.16</td>
<td>0.020</td>
</tr>
<tr>
<td>(\beta_{17Cha})</td>
<td>-1.497</td>
<td>0.18</td>
<td>-1.434</td>
</tr>
<tr>
<td>(\beta_{18Cha})</td>
<td>-1.474</td>
<td>0.18</td>
<td>-1.413</td>
</tr>
<tr>
<td>(\beta_{19Cha})</td>
<td>-1.564</td>
<td>0.19</td>
<td>-1.494</td>
</tr>
<tr>
<td>(\beta_{20Cha})</td>
<td>0.091</td>
<td>0.24</td>
<td>0.069</td>
</tr>
<tr>
<td>(\beta_{21Cha})</td>
<td>2.145</td>
<td>0.26</td>
<td>1.979</td>
</tr>
<tr>
<td>(\beta_{22Cha})</td>
<td>1.840</td>
<td>0.25</td>
<td>1.705</td>
</tr>
<tr>
<td>(\beta_{23Cha})</td>
<td>-0.023</td>
<td>0.22</td>
<td>-0.034</td>
</tr>
<tr>
<td>(\beta_{24Cha})</td>
<td>1.613</td>
<td>0.21</td>
<td>1.534</td>
</tr>
</tbody>
</table>

(continued)
Table 4-2. (continued)

<table>
<thead>
<tr>
<th></th>
<th>Multidimensional</th>
<th>SCM 1</th>
<th>SCM 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>SE</td>
<td>Est.</td>
</tr>
<tr>
<td>( \delta_1 )</td>
<td>1.178</td>
<td>0.37</td>
<td>1.212</td>
</tr>
<tr>
<td>( \delta_2 )</td>
<td>0.687</td>
<td>0.38</td>
<td></td>
</tr>
<tr>
<td>( \sigma^2_{CoS} )</td>
<td>1.872</td>
<td>0.25</td>
<td>1.742</td>
</tr>
<tr>
<td>( \sigma^2_{Cha} )</td>
<td>1.805</td>
<td>0.23</td>
<td>1.425</td>
</tr>
<tr>
<td>( \sigma_{CoS,Cha} )</td>
<td>1.251</td>
<td>0.16</td>
<td>0.912</td>
</tr>
<tr>
<td>( \rho_{CoS,Cha} )</td>
<td>0.680</td>
<td>0.579</td>
<td>0.379</td>
</tr>
<tr>
<td>DIC</td>
<td>7227.16</td>
<td></td>
<td>7221.33</td>
</tr>
</tbody>
</table>

Rasch model. These results were consistent with the findings in the simulation study, where the variance of the target construct and the covariance were overestimated in the multidimensional Rasch model. Lastly, the estimated DIC indicates that SCM 1 fit better than the multidimensional Rasch model.

Given that, comparing the estimate with its standard error, \( \delta_2 \) was not significantly different from zero at the 5% level in SCM 1, only \( \delta_1 \) for examinees in the proficient level on CoS was included in the second SCM analysis (SCM 2). Based on the estimated DIC value, SCM 2 fit better than the multidimensional Rasch model as well as SCM 1. To illustrate, in SCM 2, the discontinuity parameter (\( \delta_1 \)) was greater, and variance of the latent ability in Cha (\( \sigma^2_{Cha} \)) and covariance between the two constructs (\( \sigma_{CoS,Cha} \)) were estimated to be smaller than ones in SCM 1. Additionally, compared to the estimates of the multidimensional Rasch model, \( \sigma^2_{Cha} \) and \( \sigma_{CoS,Cha} \) were estimated to be much smaller in SCM 2. In other words, decreased correlation between the two constructs (\( \rho_{CoS,Cha} = 0.373 \)), which was substantially smaller than those obtained under the multidimensional Rasch model and SCM 1, suggests that incorporating a discontinuity parameter for the examinees in the proficient level on CoS into the multidimensional Rasch model explains much of the raw correlation of the latent ability variables in CoS and Cha in the multidimensional model analysis.

Finally, Figure 4-4 presents the latent ability distribution in CoS and Cha as estimated under the multidimensional Rasch model and the two SCM analyses. The solid lines represent the distribution of \( \hat{\theta}_{pCoS} \) and \( \hat{\theta}_{pCha} \) in the multidimensional Rasch model. The dotted lines in Figure 4-4 (a) display \( \hat{\theta}_{pCoS} \) (upper panel), and \( \hat{\theta}_{pCha} \) augmented by \( \hat{\delta}_1 \) and \( \hat{\delta}_2 \) (lower panel) in SCM 1. More specifically, the dotted curve in the lower panel combines the distribution of \( \hat{\theta}_{pCha} + \hat{\delta}_1 \) for examinees in the proficient level on CoS and \( \hat{\theta}_{pCha} + \hat{\delta}_2 \) for those in the non-proficient level on CoS. Likewise, the dotted line in Figure 4-4 (b) plots \( \hat{\theta}_{pCha} \) of examinees in the non-proficient level on CoS and \( \hat{\theta}_{pCha} + \hat{\delta}_1 \) of examinees in the proficient level on CoS as estimated by SCM 2.
Figure 4-4. Distributions of the estimated latent abilities of CoS and Cha
In general, compared to the distribution of the multidimensional Rasch model, the SCM models, especially SCM 2, differentiate examinees on Cha more clearly into two groups (i.e., show a bi-modal distribution).

4.5. Conclusion and Discussion

This study provides a preliminary investigation of ideas for measurement models in complicated learning progressions, in which relations between levels across multiple constructs are hypothesized. For this purpose, this study discusses the previous approach, the SCM based on ordered latent class analysis, and presents an alternative to latent class models. The SCM framework, proposed in this chapter, assumes that each construct is a continuous latent variable indicating examinees’ proficiency and examinees are assigned into a certain level based on the cut score. Therefore, each examinee’s proficiency in this study is represented in terms of the level on the construct as well as the latent continuum.

In the SCM model, the hypothesized relations between levels across multiple constructs are modeled by incorporating the discontinuity parameters into the multidimensional Rasch model. Specifically, these hypothesized links represent assumptions such that reaching a particular level on the first construct is required for respondents to attain a certain level on the second construct. In this sense, the two constructs are referred to as the requirement and the target construct respectively, and the discontinuity parameters describe the advantage or disadvantage for respondents in a level on the requirement construct to reach a level on the target construct.

Results from the simulation study indicate that the proposed SCM model appears to recover the parameters well, and that the estimation accuracy of the discontinuity parameters depends on the sample size in the relevant level on the requirement construct. Moreover, ignoring the discontinuity parameters and applying the multidimensional Rasch model yields overestimation of the variance of the latent variable on the target construct and the covariance between the latent variables on the two constructs. In the empirical example of the ADM data using the CoS and Cha constructs, this study finds that the SCM with one discontinuity parameter, which implies the assumption that students in the proficient level on CoS have an advantage to reach the proficient level on Cha, fits better than both the multidimensional Rasch model and the SCM with the two discontinuity parameters. Both the simulation study and empirical data analysis suggest that the discontinuity parameters are closely related to the association between the two constructs, which leads overestimation of the covariance of the latent variables in the multidimensional models.

The simple SCM model with two levels in two constructs with a single connection can be extended to more complicated models such as those with more levels within one construct, more than two constructs, and more connections between levels.
across constructs. However, given that the SCM modeling is substantially based on the assumption that there are relations between constructs, it is crucial to have validated theory and data to support the hypotheses. An extension, including other response types such as polytomous responses, is also an important development of the model.

Another limitation of this current study is that the cut score, which plays an important role in classifying examinees into the levels, is assumed to be determined using the mean of item difficulties. However, in many circumstances, the cut score needs to be estimated as well. For example, Jiao, Lissitz, Macready, Wang, and Liang (2011) proposed one possible way to estimate the cut score. They deployed a mixture Rasch model to find subgroups of examinees and allowed inter-individual differences within a subgroup. They specified the cut score as the intersecting point of two adjoining distributions of the latent ability in two adjacent latent classes. These extensions may enhance the application of the SCM modeling as a measurement model for complex learning progressions.
Chapter 5.
Summary and Conclusion

In this dissertation, I investigated extensions and applications of multilevel and multidimensional item response models, focusing on longitudinal item response data that include students’ school switching, classification of examinees into latent classes based on multidimensional aspects, and measurement models for complicated learning progressions. This dissertation consists of three papers, Chapters 2, 3, and 4. I present brief summary and conclusion of each chapter below.

In Chapter 2, the cross-classified multiple membership models for longitudinal item response data (CCMM-LIRD) were proposed to incorporate students’ school mobility, which is often observed in longitudinal studies. The Type I mobility pattern describes students’ simultaneous school switching, such as graduating from middle school and entering high schools, and the crossed-classified models were incorporated into the three-level hierarchical generalized linear model for longitudinal item response data (HGLM-LIRD). More specifically, the random effects of the middle schools and high schools were included and time-varying coefficients associated school effects were assumed. In the Type II mobility, some of students transfer from school to school at any time of measurement occasions. In order to deal with this type of school mobility, students’ membership of more than one school were modeled through the use of the multiple membership models, in which the effects of the schools were specified according to proportions of time that the students have attended the school up to a certain time point.

The results of the simulation studies indicate that the proposed approaches yield fairly good recovery of the parameters in both types of school mobility considered in this chapter. Furthermore, in both types, the fixed effect parameters such as the item difficulties and the growth trajectory parameters were not influenced by misspecifying the school-level random effects, including ignoring school effects or assuming that students stay within the first school. However, ignoring school-level random effects resulted in redistribution of the between-school variance into the lower-levels, yielding overestimation of the variances of the time-level or (and) the student-level random effects. Moreover, incorrect specification of the school-level random effects produced overestimation or underestimation of the between-school variances. These results emphasize the importance of proper modeling of school mobility when the data sets include mobile students.

In addition to the simulation studies, three sets of large-scale longitudinal data in education, vocational maturity data of the KYPS and self-esteem data of the NELS: 88 for Type I, and mathematics data of the ECLS-K for Type II, were analyzed to
illustrate applications of the proposed models. The CCMM-LIRD models allowed us to disclose distinct contributions of the middle schools and high schools as well as to investigate differential contributions of the schools over time, when the data sets contained the Type I mobility. In the ECLS-K data, the CCMM-LIRD which included students who attended multiple schools explained the growth of students better than the three-level and four-level models. The consequences of misspecifying the school-level random effects in the empirical data studies were consistent with findings from the simulation studies.

In Chapter 3, the mixture random weights linear logistic test model (MixRWLLTM) was presented for classifying examinees into subgroups which are qualitatively distinguished and for defining characteristics of latent classes. According to the proposed model, latent classes are defined based on multiple aspects, a general propensity (intercept) and random coefficients of the item properties. In other words, the item properties, which are used to explain the item difficulties in the LLTM, can be employed to describe latent classes as well.

As an empirical data study, verbal aggression data in which items were designed based on the four design factors, was analyzed using the one- and two-class LLTM and RWLLTM. The results reveal that the two-class RWLLTM fitted better than the other models. Under the MixRWLLTM, in one class, examinees whose general propensity of verbal aggression was higher tended to do verbally aggressive reaction, while there was no association between the general propensity and the degree to which they actually do verbally aggressive behaviors in the other class. The simulation study suggests that the applications of Bayesian estimation appeared to recover the parameters in the MixRWLLTM well and, as the test length increased, the accuracy of correct classification rate increased.

Lastly, in Chapter 4, the structured constructs model (SCM) for the continuous latent trait was developed as a suitable measurement model in complex learning progressions, in which relations between levels across multiple constructs are assumed in advance. In the proposed SCM approach, each examinee’s ability is defined as continuous latent variables, and levels in each construct are determined based on the cut scores. Moreover, the discontinuity parameters model the hypothesized relations as the advantage of disadvantage for respondents belonging into a certain level in one construct to reach a level in another construct.

In the simulation study, two constructs, referred to as the requirement and target constructs, and two levels in each construct were assumed. Two discontinuity parameters were specified for examinees in each level on the requirement construct. The results of fitting the SCM and the multidimensional Rasch model to the simulated data sets indicate that parameters were recovered pretty well in the proposed model and ignoring discontinuity parameter resulted in substantial overestimation of the variance of the latent variable on the target construct and the covariance of the two latent
variables. Applications the SCM model to the mathematics data on the CoS and Cha constructs of the ADM project suggest that the SCM with one discontinuity parameter for examinees in the proficient level on CoS fitted better than the multidimensional Rasch model without discontinuity parameters. This result confirms the hypothesis that students who reach the proficient level on CoS are more likely to attain the proficient level on Cha. In sum, this study shows that the proposed SCM approach could be useful to analyze data from the learning progressions with complex outcome progression structures.
References


schools frequently, harming their education (GAO/HEHS publication no. 94-45).


Statistical Innovations Inc.

*Psychological Bulletin, 105*(2), 276-289.

Lawrence Earlbaum Associates.


Wilson, M. (2012). Responding to a challenge that learning progressions pose to measurement
practice. In A. C. Alonzo & A. W. Gotwals (Eds.), *Learning progressions in science* (pp.

Appendix A. WinBUGS Code for the CCMM-LIRD

1. WinBUGS code for Type I (KYPS data analysis)

model {
  for (j in 1:J) {
    for (t in 1:T) {
      theta[j,t] <- b0 + b1*d1[t] + b2*d2[t] + zeta[j] +
      w1[t]*gamma[msid[j]] + w2[t]*eta[hsid[j]] + epsilon[j,t]
      for (i in 1:I) {
        logit(prob[j,(t-1)*I+i]) <- theta[j,t] - delta[i]
        resp[j,(t-1)*I+i] ~ dbern(prob[j,(t-1)*I+i])
      }
    }
  }
  for (m in 1:M) {
    gamma[m] ~ dnorm(mu3, iv31)
  }
  mu3 <- 0
  iv31 ~ dgamma(0.001, 0.001)
  tau1 <- 1/iv31
  for (h in 1:H) {
    eta[h] ~ dnorm(mu4, iv41)
  }
  mu4 <- 0
  iv41 ~ dgamma(0.001, 0.001)
  tau2 <- 1/iv41
  for (j in 1:J) {
    zeta[j] ~ dnorm(mu2, iv21)
  }
  mu2 <- 0
  iv21 ~ dgamma(0.001, 0.001)
  psi <- 1/iv21
  for (j in 1:J) {
    for (t in 1:T) {
      epsilon[j,t] ~ dnorm(mu1, iv11)
    }
  }
}
mu1 <- 0
iv11 ~ dgamma(0.001, 0.001)
sigma <- 1/iv11
for (i in 1:(I-1)) {
    delta[i] ~ dnorm(0,1)
}
delta[I] <- -sum(delta[1:(I-1)])
b0 ~ dnorm(0, 1)
b1 ~ dnorm(0, 1)
b2 ~ dnorm(0, 1)
w1[1] <- 1
for(t in 2:T) {
    w1[t] ~ dnorm(0, 0.001)
}
w2[1] <- 0
w2[2] <- 0
w2[3] <- 1
for (t in 4:T) {
    w2[t] ~ dnorm(0, 0.001)
}

2. WinBUGS code for Type II (ECLS-K data analysis)
model {
    for (j in 1:J) {
            +step(3-sch[j])*nu[smem1[j,1]]
            +step(sch[j]-2)*(1/2*nu[smem1[j,1]]+1/2*nu[smem1[j,2]])
            +equals(sch[j],1)*nu[smem1[j,1]]
            +equals(sch[j],2)*(2/3*nu[smem1[j,1]]+1/3*nu[smem1[j,3]])
            +equals(sch[j],3)*(1/3*nu[smem1[j,1]]+2/3*nu[smem1[j,2]])
            +equals(sch[j],4)*(2/3*nu[smem1[j,1]]+1/3*nu[smem1[j,2]])
            +equals(sch[j],5)*(1/3*nu[smem1[j,1]]+1/3*nu[smem1[j,2]]
            +1/3*nu[smem1[j,3]])
    }
}
for (j in 1:J) {
    for (t in 1:T) {
        for (i in 1:I) {
            logit(prob[j,(t-1)*I+i]) <- theta[j,t]-delta[i]
            resp[j,(t-1)*I+i] ~ dbern(prob[j,(t-1)*I+i])
        }
    }
}

for (s in 1:S) {
    nu[s] ~ dnorm(mu3, iv31)
}

mu3 <- 0
iv31 ~ dgamma(0.001, 0.001)
tau <- 1/iv31
for (j in 1:J) {
    zeta[j] ~ dnorm(mu2, iv21)
}

mu2 <- 0
iv21 ~ dgamma(0.001, 0.001)
psi <- 1/iv21
for (j in 1:J) {
    for (t in 1:T) {
        epsilon[j,t] ~ dnorm(mu1, iv11)
    }
}

mu11 <- 0
iv11 ~ dgamma(0.001, 0.001)
sigma <- 1/iv11
for (i in 1:(I-1)) {
    delta[i] ~ dnorm(0, 1)
}
delta[I] <- -sum(delta[1:(I-1)])
b0 ~ dnorm(0, 1)
b1 ~ dnorm(0, 1)
### Appendix B. Design Matrix of the Verbal Aggression Data

\[
X = \begin{bmatrix}
1 & 0 & 1 & 0.5 & 0.5 \\
1 & 0 & 1 & 0.5 & -1 \\
1 & 0 & 1 & -1 & 0.5 \\
1 & 0 & 1 & 0.5 & 0.5 \\
1 & 0 & 1 & 0.5 & -1 \\
1 & 0 & 1 & -1 & 0.5 \\
1 & 0 & 0 & 0.5 & 0.5 \\
1 & 0 & 0 & 0.5 & -1 \\
1 & 0 & 0 & -1 & 0.5 \\
1 & 0 & 0 & 0.5 & 0.5 \\
1 & 0 & 0 & 0.5 & -1 \\
1 & 0 & 0 & -1 & 0.5 \\
1 & 1 & 1 & 0.5 & 0.5 \\
1 & 1 & 1 & 0.5 & -1 \\
1 & 1 & 1 & -1 & 0.5 \\
1 & 1 & 1 & 0.5 & 0.5 \\
1 & 1 & 1 & 0.5 & -1 \\
1 & 1 & 1 & -1 & 0.5 \\
1 & 1 & 0 & 0.5 & 0.5 \\
1 & 1 & 0 & 0.5 & -1 \\
1 & 1 & 0 & -1 & 0.5 \\
1 & 1 & 0 & 0.5 & 0.5 \\
1 & 1 & 0 & 0.5 & -1 \\
1 & 1 & 0 & -1 & 0.5 \\
1 & 1 & 0 & 0.5 & 0.5 \\
1 & 1 & 0 & 0.5 & -1 \\
1 & 1 & 0 & -1 & 0.5 \\
1 & 1 & 0 & 0.5 & 0.5 \\
1 & 1 & 0 & 0.5 & -1 \\
1 & 1 & 0 & -1 & 0.5 \\
\end{bmatrix}
\]
Appendix C. WinBUGS Code for the MixLLTM and MixRWLLTM

1. MixLLTM

model{
    for (p in 1:P){
        for (i in 1:I){
            r[p,i] <- resp[p,i]
        }
    }

    for (g in 1:G){
        for (i in 1:I){
            for (k in 1:K){
                b[g,i,k] <- q[i,k]*beta[g,k]
            }
            be[g,i] <- sum(b[g,i,])
        }
    }

    # likelihood
    for (p in 1:P){
        for (i in 1:I){
            logit(prob[p,i]) <- theta[p]-be[gmem[p],i]
            r[p,i] ~ dbern(prob[p,i])
        }
    }

    # Prior for ability
    for (p in 1:P){
        theta[p] ~ dnorm(mu[gmem[p]], tau[gmem[p]])
        gmem[p] ~ dcat(phi[1:G])
    }
    mu[1] <- 0
    mu[2] <- 0

    for (g in 1:G){
        tau[g] ~ dgamma(1, 1)
    }
}
\[
\text{var}[g] \leftarrow 1/\tau[g]
\]

\# Prior for mixture probabilities
phi[1:G] ~ ddirch(alpha[])

\# Prior for item difficulty
for (g in 1:G)
  for (k in 1:K)
    beta[g,k]~dnorm(0,1)

2. MixRWLLTM

model{
  for (p in 1:P)
    for (i in 1:I)
      r[p,i]<- resp[p,i]

  for (g in 1:G)
    for (i in 1:I)
      for (k in 1:K)
        b[g,i,k] <- q[i,k]*beta[g,k]
      be[g,i] <- sum(b[g,i,])

  for (p in 1:P)
    for (i in 1:I)
      for (d in 1:D)
        theta2[p,i,d] <- theta1[p,d]*equals(t[i,d],1)
      theta[p,i] <- sum(theta2[p,i,])

# likelihood
for (p in 1:P){
  for (i in 1:I){
    logit(prob[p,i]) <- theta[p,i] - be[gmem[p],i]
    r[p,i] ~ dbern(prob[p,i])
  }
}

# Prior for ability
for (p in 1:P){
  theta1[p,1:2] ~ dmnorm(mu[gmem[p],1:2], tau[gmem[p],1:2, 1:2])
  gmem[p] ~ dcat(phi[1:G])
}

mu[1,1] <- 0
mu[1,2] <- 0
mu[2,1] <- 0
mu[2,2] <- 0

tau[1, 1:2, 1:2] ~ dwish(R[1:2, 1:2], 2)
tau[2, 1:2, 1:2] ~ dwish(R[1:2, 1:2], 2)
var[1, 1:2, 1:2] <- inverse(tau[1, 1:2, 1:2])
var[2, 1:2, 1:2] <- inverse(tau[2, 1:2, 1:2])
corr1 <- var[1,2,1]/(sqrt(var[1,1,1]*var[1,2,2]))
corr2 <- var[2,2,1]/(sqrt(var[2,1,1]*var[2,2,2]))

# Prior for mixture probabilities
phi[1:G] ~ ddirch(alpha[])

# prior for item difficulty
for (g in 1:G){
  for (k in 1:K){
    beta[g,k]~dnorm(0,0.001)
  }
}
}
Appendix D. WinBUGS Code for the SCM 1 of the ADM Analysis

model {
  for (j in 1:J) {
    for (i in 1:16) {
      logit(prob[j,i]) <- theta[j,1]+beta1[i]
      resp[j,i]~-dbern(prob[j,i])
    }
    for (i in 1:18) {
      logit(prob[j,16+i]) <- theta[j,2]+delta1*step(theta[j,1]-0)
                         +delta2*step(0-theta[j,1])-beta2[i]
      resp[j,16+i]~-dbern(prob[j,16+i])
    }
  }
  for (j in 1:J) {
    theta[j, 1:2] ~ dnorm(mu[1:2], R[1:2,1:2])
  }

  mu[1] <- 0
  mu[2] <- 0

  R[1:2,1:2] ~ dwish(Omega[1:2,1:2], 2)
  IR[1:2,1:2] <- inverse(R[1:2,1:2])
  corr <- IR[1,2]/(sqrt(IR[1,1]*IR[2,2]))

  delta1 ~ dnorm(0, 0.001)
  delta2 ~ dnorm(0, 0.001)
}

91