

EXPLORING THE ESTIMATION OF EXAMINEE LOCATIONS
USING MULTIDIMENSIONAL LATENT TRAIT MODELS
UNDER DIFFERENT DISTRIBUTIONAL ASSUMPTIONS

By

HYESUK JANG

A DISSERTATION

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

Measurement and Quantitative Methods - Doctor of Philosophy

2014

ABSTRACT

EXPLORING THE ESTIMATION OF EXAMINEE LOCATIONS USING MULTIDIMENSIONAL LATENT TRAIT MODELS UNDER DIFFERENT DISTRIBUTIONAL ASSUMPTIONS

By

Hyesuk Jang

This study aims to evaluate a multidimensional latent trait model to determine how well the model works in various empirical contexts. Contrary to the assumption of these latent trait models that the traits are normally distributed, situations in which the latent trait is not shaped with a normal distribution may occur (Sass et al, 2008; Woods & Thissen, 2006). As a result when studies construct evaluations or comparisons in order to determine the appropriate estimation method and to avoid inefficient ones, the distribution or distributional statistics of the latent trait are considered as a key assumption. This study explores the performance of parameter estimation using a bifactor model, a type of multidimensional latent trait model in order to provide information of the effects of violations of the distributional assumptions.

The effects of the distributional assumptions are evaluated using simulation studies. A two-parameter logistic bifactor model with three factors: one general and two specific factors, is used as a basic multidimensional latent model. Simulation studies construct eight distributional conditions based on the degree of skewedness of the general factor, the directions of skewedness of the specific factors, the correlation between specific factors and four types of item parameter conditions.

The results showed that item parameter estimation was affected by the degree of skewedness of the general factor, the directions of skewedness of the specific factors, and the correlation between specific factors. These conditions of the latent trait distributions had

different effects on item parameter estimation depending on the type of item parameter. Based on the variances of the mean biases and correlations between generated and estimated parameters, the most important condition of the latent trait distribution for θ parameter estimation was the correlation between the specific factors. With the increasing number of studies and practical need for multidimensional structures of latent traits, this research provides useful guidelines for constructing appropriate multidimensional models.

Key words: Multidimensional latent trait model, bifactor model, latent trait distribution, simulation study

ACKNOWLEDGMENTS

I would like to deeply thank my advisor and committee members for supporting me throughout my doctoral study. Dr. Mark Reckase always encouraged and supported me with not only his professional knowledge and skills but also sincere words and emotional care. With his advice and care, I completed my doctoral study and dissertation. Dr. Kenneth Frank gave me unstinting support all the time. He motivated me to go forward and guided me with valuable directions and warm messages. I really appreciate Dr. Kimberly Maier. With thoughtful advice and support, she helped and guided me about both academic and personal issues during the doctoral study. I am very grateful to Dr. Chae Young Lim for helping and guiding me with my doctoral study and personal plan. She always helped me with her professional knowledge and kind directions.

My acknowledgement goes to my mentors, friends and colleagues. Without their generous sharing and hearty support, it would have been impossible for me to complete my doctoral studies. All of the things that I have shared with them helped me to go through my doctoral study and they are valuable memories for me.

Lastly, I am very grateful to my parents Daeyong Jang and Jinsook Jeong, my brother Hyungoo Jang, my sister in law Soyeon Han, and my beloved nephews Youwhan Jang and Youhyun Jang. I was able to accomplish my doctoral study with their unconditional love, trust and support. From the bottom of my heart, I would like to thank my entire family.

TABLE OF CONTENTS

LIST OF TABLES.....	vii
LIST OF FIGURES	x
1. Introduction.....	1
1.1 Multidimensional Latent Trait	1
1.2 Multidimensional Latent Trait Models	3
1.3 Latent Trait Distributions	5
1.4 Simulation Study	6
1.5 Research Question.....	7
2. Literature Review.....	10
2.1 Bifactor Model	10
2.2 Previous Research on Latent Trait Distributions	12
2.3 Simulation Study	15
2.3.1 Simulation as a Research Methodology	15
2.3.2 Simulation Studies of Latent Trait Models.....	18
a. Distribution of Latent Trait	18
b. Intercorrelation between Latent Traits	18
c. Discrimination Parameter	19
d. Difficulty Parameter	19
e. Replication.....	20
f. Number of Items	20
g. Number of Examinees	21
h. Estimation Methods.....	21
i. Computer Programs for Parameter Estimation	21
2.4 Empirical Data Analysis	22
2.4.1 Distribution of Latent Traits in PISA 2003, 2006, and 2009	22
2.4.2 Sub-domain Proficiency Levels in PISA.....	25
3. Method.....	29
3.1 Data Generation.....	29
3.1.2 Data Generating of Latent Trait Distributions.....	29
3.1.2 Data Generating of Item Parameters	35
3.1.3 Data Generating of Examinees' Responses.....	37
3.2 Evaluation Methods.....	38
4. Results.....	40
4.1. Data Generation.....	40
4.1.1 Item Parameter Generation	40
4.1.2 θ Parameter Generation	41
4.2. Bifactor Analysis.....	47
4.2.1 Item Parameters	47

4.2.2 θ Parameters	52
a. Mean of Mean Biases	52
b. Variances of Mean Biases	55
c. Correlations between Generated and Estimated Parameter Distributions.....	56
d. Kolmogorov-Smirnov Test (KS test)	57
5. Discussion.....	64
5.1 Summary of the Results	64
5.2 Implications	65
APPENDICES	70
Appendix A	71
Appendix B	75
Appendix C	79
Appendix D	83
Appendix E.....	85
BIBLIOGRAPHY	93

LIST OF TABLES

Table 2-1. Comparison of Three Research Methodologies	16
Table 2-2. Cut scores of Reading Literacy Proficiency Levels in PISA 2009	23
Table 2-3. Percentage Distribution of Proficiency Level Scores in PISA 2000, 2003 and 2009 .	24
Table 3-1. Simulation Conditions for Data Generation	31
Table 3-2. Simulation Combinations of Latent Trait Distributions.....	32
Table 3-3. Parameters for Generating Distributions of Simulation Combinations of Item Parameters	36
Table 4-1. Descriptive Statistics of Generated Item Parameters	42
Table 4-2. Descriptive Statistics of General θ Parameters Generated.....	44
Table 4-3. Descriptive Statistics of First Specific θ Parameters Generated	45
Table 4-4. Descriptive Statistics of Second Specific θ Parameters Generated.....	46
Table 4-5. Correlations Between Two Specific θ Parameters Generated	47
Table 4-6. Means of Item Parameter Mean Bias	48
Table 4-7. Variances of Item Parameter Mean Bias	49
Table 4-8. Means of θ Parameter Mean Bias	53
Table 4-9. Variances of θ Parameter Mean Bias	54
Table 4-10. Mean of the Correlations of the General Factors	58
Table 4-11. Correlation Means of the First Specific Factors.....	59
Table 4-12. Correlation Means of the Second Specific Factors	60
Table 4-13. Frequency of Significant Differences between Distributions of Generating and Estimated General Factor θ Parameters	62
Table A-1. Parameter Estimates of Quadratic Regression Function for Positively Skewed Distribution with 2,000 examinees.....	71

Table A-2. Parameter Estimates of Cubic Regression Function for Positively Skewed Distribution with 2,000 examinees	71
Table A-3. Parameter Estimates of Quadratic Regression Function for Positively Skewed Distribution with 10,000 examinees	72
Table A-4. Parameter Estimates of Cubic Regression Function for Positively Skewed Distribution with 10,000 examinees	72
Table A-5. Parameter Estimates of Quadratic Regression Function for Negatively Skewed Distribution with 2,000 examinees	73
Table A-6. Parameter Estimates of Cubic Regression Function for Negatively Skewed Distribution with 2,000 examinees	73
Table A-7. Parameter Estimates of Quadratic Regression Function for Negatively Skewed Distribution with 10,000 examinees	74
Table A-8. Parameter Estimates of Cubic Regression Function for Negatively Skewed Distribution with 10,000 examinees	74
Table B-1. Mean and Variance of Item Parameter Biases under Disc of 1.3 and Diff of -0.5	75
Table B-2. Mean and Variance of Item Parameter Biases under Disc of 1.3 and Diff of 0.5	76
Table B-3. Mean and Variance of Item Parameter Biases under Disc of 1.8 and Diff of -0.5	77
Table B-4. Mean and Variance of Item Parameter Biases under Disc of 1.8 and Diff of 0.5	78
Table C-1. Mean and Variance of θ Parameter Biases under Disc. of 1.3 and Diff. of -0.5	79
Table C-2. Mean and Variance of θ Parameter Biases under Disc. of 1.3 and Diff. of 0.5	80
Table C-3. Mean and Variance of θ Parameter Biases under Disc. of 1.8 and Diff. of -0.5	81
Table C-4. Mean and Variance of θ Parameter Biases under Disc. of 1.8 and Diff. of 0.5	82
Table D-1. Correlations between Generated and Estimated Factors with Discrimination Parameters from mean of 1.3	83
Table D-2. Correlations between Generated and Estimated Factors with Discrimination Parameters from mean of 1.8	84
Table E-1. Numbers of Frequencies Significant by KS Test under Condition 1	85
Table E-2. Numbers of Frequencies Significant by KS Test under Condition 2	86
Table E-3. Numbers of Frequencies Significant by KS Test under Condition 3	87

Table E-4. Numbers of Frequencies Significant by KS Test under Condition 4.....	88
Table E-5. Numbers of Frequencies Significant by KS Test under Condition 5.....	89
Table E-6. Numbers of Frequencies Significant by KS Test under Condition 6.....	90
Table E-7. Numbers of Frequencies Significant by KS Test under Condition 7.....	91
Table E-8. Numbers of Frequencies Significant by KS Test under Condition 8.....	92

LIST OF FIGURES

Figure 1-1. Examples of Multidimensional Latent Variable Models (Reise et al, 2007; Reise et al, 2010).....	3
Figure 2-1. Percentage distribution of proficiency level in PISA 2000, 2003 and 2006.....	25
Figure 2-2. Percentage Distribution of Proficiency Level in PISA 2009	25
Figure 2-3. Percentage Distribution of Sub-domains in PISA 2000, 2003, 2006, and 2009.....	26

1. Introduction

1.1 Multidimensional Latent Trait

The work on latent traits started in the 1950s. According to Gifford (1978), the word ‘latent trait’ was mentioned in Lazarsfeld (1950), and latent trait theory was first developed by Lord (1952, 1953a, 1953b). Latent traits are unobservable, and cannot be measured directly. In latent trait theory, the latent trait is portrayed as underlying participants’ performance on sets of test items, which is why it is called a “latent” trait or ability (Gifford, 1978). Test items are used to collect participants’ responses to particular stimuli, and based on the response features from the collected data, the characteristics of the participants and items may be estimated by using a latent trait model.

In order to have a basis in scientific methods, item response theory (IRT) as a latent trait theory has been developed to describe the relationship between participants’ responses and their level of abilities by a mathematical function (Lord, 1980). The models used for item response theory can be distinguished as unidimensional item response theory (UIRT) or multidimensional item response Theory (MIRT) models depending on whether the number of latent traits modeled is one, or more than one. According to the book by Reckase (2009), work in fields such as education, psychology, and statistics suggests that the structure of human knowledge is complicated, and that the processes that produce observed responses to test items are often complex and varied. As a result, multidimensional item response theory (MIRT) has been developed to better fit reality. Chalmers (2012) also suggested that even though unidimensional models can be useful, in order to adequately specify the nature of measures with complicated structures, it is essential to consider their dimensionality.

Many researchers also have considered multidimensionality in measuring particular constructs of interest. To produce test items that follow an expected factor structure, item construction studies have constructed and analyzed test item data using multidimensional latent trait models, for example in educational assessment (OECD, 2007a & 2007b; von Davier, 2008; Hichendorff, 2013) and psychological or sociological constructs (Capella & Turner, 2004; Yoshida & James, 2010; Eboli & Mazzulla, 2007; Martin, 2007; Duncan-Jones, 1981a & 1981b). In educational assessment of science literacy in PISA 2006, for example, the test consisted of three content areas: earth and space systems, living systems, and physical systems. According to OECD (2007a; 2007b), the average score on science questions from different content areas for a particular country tend to vary. This suggests that even though the test examines science literacy as a general latent trait, a different pattern of the students' ability distribution can exist depending on the sub-latent traits (OECD, 2007a). In this case, modeling the total latent trait, ignoring the sub-domains, could result in scores that are not easily interpretable or policy decisions that are erroneous because of the lack of information about the student latent trait.

Capella and Turner (2004) developed an instrument of customer satisfaction in the vocational rehabilitation services. In this research, the customer satisfaction survey considered four components of satisfaction: counselor interpersonal factors, counselor job effectiveness, satisfaction with the services, and satisfaction with the agency. Confirmatory factor analysis indicated that the satisfaction instrument consisted of three dimensions that reflected the counselor that the customers interacted with, the services that the customers received, and the agency that provided the services. The research showed that customer satisfaction can be described as consisting of multiple latent traits.

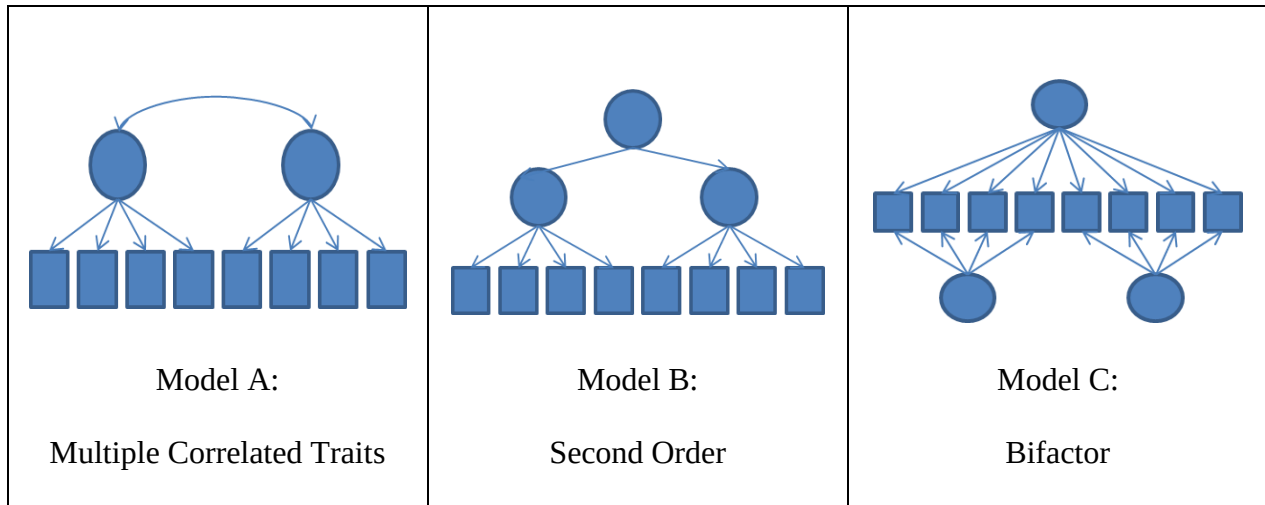


Figure 1-1. Examples of Multidimensional Latent Variable Models (Reise et al, 2007; Reise et al, 2010)

Multidimensional measurement and analytical methods have been used in social network analysis of human interactions. One of the efforts in the measurement of interactions that has attracted interest was the construction of the Interview Schedule for Social Interaction (ISSI), which was developed by the Social Psychiatry Research Unit at the Australian National University (Duncan-Jones, 1981a & 1981b). The survey consisted of 50 items asking about the availability or adequacy of social interaction and attachment, and about acquaintances, friends, attachment, opportunities for nurturance and reassurance of worth, and reliable alliances. Duncan-Jones (1981) evaluated and characterized the structure of the ISSI according to subdomains of social relationships by using confirmatory factor analysis.

1.2 Multidimensional Latent Trait Models

To describe various item content, formats, and relationships between multiple factors, various latent trait models have been developed and used. Reise et al. (2007) provides examples of multidimensional latent models, three of which are shown in Figure 1-1. Circles represent dimensions or latent factors, and rectangles represent items used for measuring the factors.

Model A is a typical multidimensional correlated traits model. Each latent trait is related to some of the items, and it is assumed that there is a correlation between the factors. Model B shows a multidimensional model with a higher order structure of latent traits, which is often referred to as a second-order factor model. As in Model A, the latent traits at the lower level are measured by some of the items. The difference is the presence of a second order trait explaining the correlation between the first level traits. Although Models A and B measure certain factors related to common parts of the items and show relationships among the latent traits, they do not include a general factor directly related to all of the items. Model C shows the structure of a bifactor model. The bifactor model has two kinds of factors: a 'general' factor connected to the all items accounting for the item intercorrelations, and several 'group' or 'specific' factors connected to some of the items representing additional covariance unexplained by the general factor.

The bifactor model has a mathematical relationship with other models specifying multidimensional structures of test items. In Rijmen (2010), the bifactor model was compared to two other multidimensional IRT models: the testlet model and the second-order model. The research demonstrated that while all three models take account of item clusters, there are some differences in the consideration of specific factor loadings. The testlet model has a constraint on the loadings of the specific factors, and they are estimated from the general factor loading in a proportional way within each testlet. Under the assumption that the second order model also has proportional specific factor loadings, the second order models can be described as restricted forms of the bi-factor model. Therefore, research using the bifactor model can be viewed as relevant to multidimensional models in general, including those described above that take item clusters into account.

1.3 Latent Trait Distributions

Many studies using latent trait models focus on the latent trait estimate. To estimate the latent trait means to locate each examinee somewhere on each continuum scale, allowing us to investigate an examinee's status on each latent trait, or to compare examinees' relative statuses (Reckase, 2009). Hambleton et al. (1991) emphasized the importance of the latent trait in demonstrating that the IRT models are based on two postulates; one about whether participants' performances on test items can be explained by their latent traits, and the other about whether the relationship between their item performance and the traits can be modeled by a particular family of item characteristic functions. Among the parameters in latent trait models including item difficulty, item discrimination and item guessing parameters, and person latent trait parameters, Sass et al. (2008) suggested that the latent trait parameters be considered the most important to estimate, because the latent trait estimates can be used to determine an examinee's proficiency classification or standing on a psychological construct.

The estimation of the latent traits is important not only to providing examinees' proficiency classification or measurement of psychological constructs but also because it reflects to a central assumption of latent trait models. Most of factor models assume normal distribution of the latent trait; however, situations in which the latent trait is not normally distributed may occur in reality (Sass et al, 2008; Woods & Thissen, 2006). Violation of the assumption that the latent trait is normally distributed is a critical issue because it affects confidence in the estimates from statistical models, which have desirable regularity properties only under conditions consistent with the assumptions. Researchers who are interested in studies related to parameter estimation of latent trait models have used various estimation methods, such as maximum likelihood, least squares, and Bayesian estimators. All these estimation methods use the

distribution or distributional statistics of the latent trait, and assumptions about the latent trait distribution are an important issue. Many studies have constructed evaluations or comparisons of estimation methods in order to identify the most appropriate estimation methods and to avoid the inefficient ones (Finch, 2010; Cai et al., 2011; Li & Lissitz, 2012; Woods & Thissen, 2006). As an extension of these studies, my research evaluates the estimation performance of a multidimensional model under conditions characterized by various distributional assumptions about the latent traits.

1.4 Simulation Study

The study of latent trait distributions is significant in order to evaluate the performance of latent trait models and their estimation processes. Many of the studies that evaluate estimation quality use simulation and this study exploring the estimation performance of multidimensional latent trait models also uses simulation. Simulation studies have been popular in various fields of research. The large numbers of simulation studies in certain fields or addressing particular topics shows that many researchers are still using simulation studies (Axelrod, 2005). One reason for using simulation study is a lack of appropriate empirical data. Practically, it is true that collecting data requires a lot of time and effort, and sometimes it is hard to get the data that we really want to analyze. Especially when research conditions do not permit researchers to collect appropriate data to address a particular question, such data can be simulated. However, leaving aside practical limitations that suggest study by simulation, there are significant benefits that simulation study provides.

First, by using a simulation study, we can re-use existing information derived from previous research to conduct deeper research or develop a sequential research line. If we have

parameter estimates in an original dataset related to our research interest, using them to generate additional data samples may be a better use of resources than collecting data again.

Simulation study provides not only the tools to allow us to use available information, but also the opportunity to study unobserved research conditions. Without collecting data, empirical analysis is not able to be used unless the analysis uses second hand or published data. On the other hand, in simulation studies, conditions that cannot easily be created or may never have occurred can be produced. Also, the results from these studies can predict and help prepare for empirical situations that might occur in the future.

Simulation study methods allow us to replicate the statistical analyses in a very efficient way. Replication allows us to confirm that the results from a simulation study are reliable, or to test that the inferences from various models are robust (Axelrod, 2005). For example, in order to compare estimation methods, we need to examine amounts of estimation error or other statistical errors that are not related to one's research interests, but inevitably occur. In this case, one set of empirical data is not enough to compare the differences among all conditions of interest. Finally, the results from the simulation study provide methodological information and can be discussed related to empirical applications, as I will do in this study.

1.5 Research Question

This research evaluates estimation performance of a multidimensional latent trait model for latent trait and item parameters under various latent trait distribution conditions. The multidimensional latent trait model that is used for this research is a two-parameter bifactor model with one general factor and two specific factors. In order to answer the research questions, a simulation study is constructed with conditions representing different latent trait distributions

and sets of item parameters. The latent trait distribution conditions can be characterized according to the degree of skewedness of the distribution, the direction of the distribution skewedness, and the intercorrelations between specific factors. The latent trait distribution conditions includes four conditions combining general and specific latent trait distributions with particular skewedness: 1) normal general factor and non-normal specific factors skewed in the same way; 2) normal general factor and non-normal specific factors skewed in a different way; 3) non-normal general factor and non-normal specific factors skewed in the same way; and 4) non-normal general factor and non-normal specific factors skewed in a different way. The parameter estimation is evaluated under two levels of intercorrelation between specific factors. Further, four different possible sets of item difficulty and discrimination parameters are considered. In total, eight conditions of latent trait distributions and four conditions of item parameters are constructed through simulation. Specific descriptions of the simulation study design are provided in Chapter 3.

This research explores the estimation performance of the bifactor model under different distributional and item parameter conditions. My specific research questions are as follows:

What precision results from the parameter estimation of the bifactor model:

1) depending on each combination of latent trait distributions in terms of (a) the normality of the general factor distribution, (b) direction of the skewed distributions included in the general or specific factors, and (c) the correlations between the specific factors?, and

2) depending on the levels of the item difficulty and discrimination parameter values?

The literature review of the bifactor model, latent trait distribution and related studies are discussed in Chapter 2, and the procedure of the simulation study and methods to generate and analyze the data are provided in Chapter 3.

2. Literature Review

2.1 Bifactor Model

Collected data is not a direct measure of the unobservable latent trait, but rather a proxy for the latent trait. Therefore, before an analyst uses collected data, its measurement properties should be evaluated about whether it represents the construct validly, and measures the construct consistently across the participants. Also, in order to provide precise information for construct analysis, it is important to consider the suitability of the method that we use for the analysis.

In order to measure the sub-structure of a general latent trait, Gibbons and Hedeker (1992) introduced the bifactor model for binary items, which is derived from the ‘bifactor’ solution named by Holzinger and Swineford (1937). The model has the constraints that each item has a) a nonzero loading on the general factor, which is the primary dimension; and b) a second loading on no more than one of the specific factors. Also, each specific factor is orthogonal to the general factor and other specific factors. The pattern matrix for a bifactor model of five items, for example, could be shown as

$$\alpha = \begin{bmatrix} \alpha_{10} & \alpha_{11} & 0 \\ \alpha_{20} & \alpha_{21} & 0 \\ \alpha_{30} & \alpha_{31} & 0 \\ \alpha_{40} & 0 & \alpha_{42} \\ \alpha_{50} & 0 & \alpha_{52} \end{bmatrix},$$

where α_{jk} is the factor loading of item j on factor k . The factor loadings indicate the item slope, or “discrimination,” parameters (Cai, Yang, & Hansen, 2011). In the matrix above, the five items are measures for one general factor and two group factors. The loadings on the general factor in the first column are α_{j0} , which should be nonzero, and the loadings of items 1-3 in the second

and 4-5 in the third column are related to specific factors 1 and 2 respectively (Gibbons & Hedeker, 1992; Li & Lissitz, 2012).

The function of the bifactor model can be explained as an IRT model. Compared to the unidimensional two-parameter IRT model, in the bifactor model general and specific latent traits are divided into separate parts with corresponding item parameters.

The functions of the unidimensional two-parameter IRT model and two-parameter bifactor model are as follows (Reckase, 2009; Cai et al, 2011; Li & Lissitz, 2012):

Unidimensional two-parameter IRT model:

$$P(u=1|\theta_0, a_i, d_i) = \frac{1}{1+\exp\{-[d_i+a_i\theta_0]\}}, \text{ and}$$

Two-parameter bifactor model:

$$P(u=1|\theta_0, \theta_s, a_0, a_s, d_i) = \frac{1}{1+\exp\{-[d_i+a_0\theta_0+a_s\theta_s]\}}$$

The left-hand side of each equation represents the probability that an examinee answers a question (or an item) correctly. θ is an individual examinee's ability related to the latent trait, and the value shows the location on the θ continuum. 'a' and 'd' are the item parameters for the i th item. 'a' indicates an item discrimination, and 'd' is calculated from the item difficulty and item discrimination parameters as shown in Equation below: (X):

$$d_i = -b \sqrt{\sum_{v=0}^m a_v^2}$$

where 'b' is an item difficulty parameter and m is the number of dimensions (Reckase, 2009).

The bifactor model has been used for multidimensional item analyses with various purposes. Reise et al (2007) demonstrated the utility of the bifactor model, and according to the research, the bifactor model can inform decisions about the dimensionality of the data and what type of models are appropriate for analysis: a) the bifactor model can be used to check the assumptions of unidimensional IRT models and test the fit of these models to possibly multidimensional data; b) it can be used, like non-hierarchical MIRT models, to form subscales; and c) it can be an alternative to using non-hierarchical multidimensional models for measuring individual differences. As a representative multidimensional latent trait model, the bifactor model is investigated under various distributional conditions of the latent traits. The next section discusses previous research on latent trait distributions.

2.2 Previous Research on Latent Trait Distributions

The estimation of the latent traits is important not only to providing examinees' proficiency classification or measurement of psychological constructs but also because it reflects to a central assumption of latent trait models. Gibbons and Hedeker (1992) explain the assumption of the latent distribution in factor models by using Thurstone's multiple factor model (1947). The multiple factor model is as follows:

$$y_j = \alpha_{j1}\theta_1 + \alpha_{j2}\theta_2 + \alpha_{j3}\theta_3 \dots + \alpha_{jk}\theta_k + \epsilon_j,$$

where y is a latent variable, θ is an underlying ability, α is a factor loading, and ϵ is a residual.

In the multiple factor model, underlying abilities of θ s and the latent trait of y are assumed to follow normal distributions. It implies that the underlying abilities (θ s) are orthogonal, which is an assumption of any bifactor model, and that residuals of ϵ are normally distributed. The assumption that the relations between factors are orthogonal reduces the

complexity of the integration involved in estimating the parameters. The estimation efficiency produced by the assumption of orthogonal latent trait distributions is strength of the bifactor model.

As mentioned above, most of factor models assume normal distribution of the latent trait; however, situations in which the latent trait is not normally distributed may occur in reality. Sass et al (2008) suggested two cases that could result in non-normal distribution of a latent trait: (a) a non-normal sampled distribution, and (b) a non-normal original distribution. A non-normal sampled distribution is derived from a non-randomly sampled population distribution. For example, when the sample is collected from a limited range of the population distribution, for instance, collected only from the low level class or the high level class, the latent trait distribution can be skewed. Also, it may occur that the original latent trait follows a non-normal distribution when a test is very difficult or very easy, or when psychological constructs that have skewed response distributions are observed.

The research on latent trait distributions has been conducted on both the unidimensional and multidimensional latent trait models. Sass et al. (2008)'s research using unidimensional IRT models showed that (a) a positively skewed distribution produces greater latent trait estimation error than a normal distribution does; (b) for extreme examinees, item difficulty estimates produce larger amount of estimation error; and (c) the best latent trait estimation procedure depends on whether a researcher is primarily interested in extreme or non-extreme examinees. Woods and Thissen (2006) introduced the non-parametric estimation of IRT latent distribution using spline-based densities, which they refer to as Ramsay-Curve IRT (RC-IRT). They showed its capability by applying it to normally-distributed and skewed latent distributions in a simulation study.

Finch (2010) compared the estimation methods implemented by NOHARM (unweighted least squares estimation) and Mplus (robust weighted least squares estimation) software using multidimensional confirmatory factor analysis models. The results showed that the estimation methods of NOHARM and Mplus were affected by the distribution of the latent traits, and that item difficulty and discrimination parameters estimated from responses of examinees a skewed latent trait distribution have a larger amount of standard error than those estimated from examinee groups with a normal latent trait distribution. This research added to results that show IRT parameter estimation is affected by the latent trait distribution shape, and can be explained by the fact that item response theory models express the functional relationship between the latent trait and observed score distributions as a normal ogive (McDonald, 1997). Batley and Boss (1993) studied the estimation of latent trait distributions with three levels of intercorrelations between two latent traits in multidimensional two-parameter logistic model. They showed that the both the best estimation of the first latent trait and worst estimation of the second latent trait occurred in the '0' correlation condition. According to their discussion, estimation of the second latent trait was influenced by rescaling of the estimates; as the correlation between the latent traits increases, the model with two latent traits has features closer to those of a unidimensional model. Cai et al. (2011) studied estimation efficiency using full information bifactor analysis. The research was designed to study conditions involving a multigroup bifactor model with normally distributed latent factors, and various types of items, such as dichotomous, ordinal, and nominal items. This study was constructed under the normal distribution assumption; however, the research discussed the possibility of the non-normal distribution of the latent trait and suggested the future research.

2.3 Simulation Study

2.3.1 Simulation as a Research Methodology

In Axelrod (2005), simulation study is identified as one of three major research methodologies: induction, deduction, and simulation; research using induction methodology discovers the patterns that the research is interested in by analyzing empirical data, while research with deduction methodology suggests a set of axioms and proves the consequences from the logical connections between the assumptions. Indicating simulation as a “third way of doing science,” Axelrod (2005) explained its process and theme as different from induction and deduction (See Table 2-1). First, in a simulation study, assumptions of theory are used for data generation, whereas research with the deduction methodology uses assumptions in order to affirm or reject a theorem that the research focuses on. Second, the data set in a simulation study is not only collected empirically as in research with induction methodology, but is also generated with specified conditions based on the assumptions. In the analysis of research conducted by the deduction methodology, consequences can be drawn from logical relationships between assumptions, while in research conducted by the induction methodology, significant patterns in the empirical results can be found. On the other hand, simulation methodology provides a tool to support creation of study designs precisely representing research conditions of theoretical or practical interest.

Küppers and Lenhard (2005) mentioned that computer simulation based on a theoretical or experimental framework might rarely be successful because reality is too complicated to be explained only by the theorem or by experiments. Computer simulation consists of numerical solutions and imitations of empirical situations. The quality of a numerical solution solutions

Table 2-1. Comparison of Three Research Methodologies

	Deduction	Induction	Simulation
Using assumptions	To affirm or reject a theorem	To ground the theorem	To generate data sets
Data	Used for construction of the theorem	Collected empirically	Specified and generated with the assumptions
Analysis and results	Drawing consequences from the relationships between assumptions	Finding the significance in data	Providing a tool to be able to use intuitional methods

* This is tabulated by using Axelrod (2005).

and imitations of empirical situations. The quality of a numerical solution depends on knowing how to control inevitable statistical or calculation errors, and validating an imitation of an empirical situation in order to reproduce the results from empirical analysis. If a theorem to specify phenomena is established, the validity of computer simulation is related to whether the study represents the empirical situation or reality accurately. As a result, Küppers and Lenhard (2005) argued that simulation modeling can be considered an attempt to imitate reality, and can be validated not by theoretical arguments but by using experience or existing data because the simulation study is an “experiment with theories.”

The fact that a simulation study is an imitation and representation of reality means that judgments about its validity can depend on epistemology. Schmid (2005) discussed the truth of simulation as connected to three philosophical theories. First, every simulation study should have a corresponding counterpart in reality, which is called a property of correspondence. Once the property of correspondence is met, there are sequential questions of how to demonstrate the relation between the statements to be described as assumptions, and reality, to exist, and how to define “reality” itself. Second, another philosophical theory related to simulation study is consensus, which means that simulation studies should be accepted by a community perspective. This implies that in addition to having an objective connection to reality, simulation studies should have subjective rationales in their context. Last, simulation studies should have coherence, which means their design is believed to be consistent with other theorems. However, a coherent situation does not guarantee a true relationship between reality and the simulation study. Referring to “sufficient accuracy and specific purpose” as the important points in evaluating the validity of simulation studies (Robinson, 2004, p210), Schmid (2005) delineated a validation process that determines the sufficient level of accuracy of a simulation, and constructs the simulation model to represent the real world system for a specific purpose.

To sum up, constructing a simulation study should be based on theory and empirical evidence to support the validity of the design. Review of previous studies and empirical analyses for determining the conditions of this simulation study will be provided subsequently.

2.3.2 Simulation Studies of Latent Trait Models

In order to answer questions about estimation of performance, many studies have been constructed using simulation. Based on previous simulations studies, simulation conditions and parameters were reviewed for selecting the simulation conditions of this study.

a. Distribution of Latent Trait

Finch (2010) constructed a simulation study to compare unweighted least squares (ULS) and robust weighted least squares (RWLS) estimation. Latent trait distributions were generated as normal, or skewed with skewness of -1.5 and kurtosis of 3.0. Cai et al. (2011) generated a general factor and specific dimensions, which were set to be jointly normally distributed and mutually orthogonal. Li and Lissitz (2012) also generated their general latent traits from normal distributions with means of -0.5, 0, and 1, and a variance of 1, and specific latent trait values from a standard normal distribution. In Woods and Thissen (2006), three kinds of latent trait distributions were constructed from 1) a normal distribution with skewness of 0 and kurtosis of 3, b) a platykurtic distribution with skewness of 0 and kurtosis of 2.53, and c) a positively skewed distribution with skewness of 1.57 and kurtosis of 6.52.

b. Intercorrelation between Latent Traits

In Batley and Boss's (1993) study, three levels of intercorrelations (0, 0.25 and 0.5) were constructed between two latent traits. Gosz and Walker (2002) used three intercorrelations (0.5, 0.75, and 0.9), two of which were higher than those of Batley and Boss. Finch's (2010) research used four levels of intercorrelations in order to evaluate the accuracy of item parameter estimation: a '0' correlation as no correlation, a 0.3 correlation as a low level of intercorrelation, 0.5 as a medium level, and 0.8 as a fairly large correlation. The research concluded that with a

high level of correlation between the latent traits, there is great bias in item parameter estimation, regardless of the estimation method used.

c. Discrimination Parameter

In order to generate the item parameters for a simulation, previous research used either a population distribution or a specific value of each parameter. The study of Finch (2010) generated discrimination parameters from a normal distribution with an estimated mean of 0.9657 and a standard deviation of 0.3161; the simulated discrimination parameters ranged between 0.3736 and 2.0158. Woods and Thissen (2006) also generated their discrimination parameters from a normal distribution, but with a mean of 1.7 and a standard deviation of 0.3, based on analysis of existing psychological scales. Cai et al. (2011) and Li and Lissitz (2012) used specific parameter values for the simulation data, which were values ranging from 1 to 2.

d. Difficulty Parameter

Difficulty parameters are usually generated from a normal distribution. Finch (2010) generated difficulty parameters from the standard normal distribution. Li and Lissitz (2012), who studied the bifactor model in vertical scaling, used difficulty parameters for non-common items from normal distributions with means between -0.5 to 0.5 and variance of 1, and as common items from a uniform distribution with a range of 1.5, for example, -1 to 0.5 or -0.5 to 1. Woods and Thissen (2006) generated difficulty parameters from a truncated standard normal distribution ranging from -2 to 2. Cai et al. (2011) selected specific values for difficulty parameters, for example, -1, -.25, .25, 1.

e. Replication

Various simulation studies with IRT models have used between 100 and 1000 replications. Finch (2010) and Woods and Thissen (2006) completed their study with 1000 replications of the simulation. The number of replications in Cai et al. (2011) was 500, and Li and Lissitz (2012) and Sass et al. (2008) replicated their simulations 100 times.

An appropriate or adequate number of replications depends on what kinds of parameter property are of primary research interest, because each parameter property needs a different level of replication to obtain stable estimation results. Once the parameter property of interest is determined, depending on the number of replications, we can evaluate the precision (stability) of the simulation. If the stability of estimation looks good above a certain minimum number of replications, we do not need to replicate the studies many times unnecessarily. On the other hand, if a large number of replications are necessary in order to get stable results, the appropriate number of simulations should be determined, and conducted. At this point, it is important to think of simulation efficiency or to construct an efficient algorithm of simulation because in case a large volume of data simulation is necessary, it is crucial to do the simulation study in a speedy and simple way.

f. Number of Items

With a small number of items, less than 30, the precision of item parameter estimation is influenced mainly by the latent trait distribution, whereas the impact of the number of items on estimation is small (Finch, 2010). This result is consistent with Stone (1992), who reported that the calibration results from at least 40 items are robust.

g. Number of Examinees

For ULS and RWLS estimation methods, even though the estimation precision slightly increases when the number of examinees is increased, there is no significant effect of the number of examinees on the estimation results if the number is greater than 250 examinees (Finch, 2010). Studying estimation of the bifactor model in vertical scaling, Li and Lissitz (2012) generated 1,000, 2,000, and 4,000 examinees. The research noted that estimation results showed that the accuracy and stability of estimation increased with the sample size, and that especially the results from the sample sizes of 2,000 and 4,000 had lower root mean squared error and standard errors than those from the sample size of 1,000.

h. Estimation Methods

Finch (2010)'s results show that using ULS estimation in NOHARM software provided better precision of item parameters, than RWLS estimation in Mplus, but he points out that ULS with NOHARM should not be used when the models have pseudo-guessing parameters and high correlations between latent traits. Also, both of the estimation methods tended to underestimate the item parameters when latent trait distributions were skewed. In order to conduct multigroup concurrent calibration during vertical scaling, Li and Lissitz (2012) implemented marginal maximum likelihood by the EM algorithm using IRTPRO software. Woods and Thissen (2006) also used the EM algorithm for marginal maximum likelihood to compute spline-based densities.

i. Computer Programs for Parameter Estimation

As interest in parameter estimation of latent trait models has increased, various kinds of computer programs have employed and developed program languages and packages for latent trait analysis. Chalmers (2012) introduced the Multidimensional IRT package for R-programming, and Sheng (2010) studied MATLAB programming in order to estimate MIRT

models with general and specific latent traits using Bayesian methods. IRTPRO is equipped with bifactor model analysis, and Mplus also provides bifactor model analysis with maximum likelihood estimators. Seo (2011) estimated the parameters of latent traits for the bifactor model with maximum likelihood and Bayesian estimation methods using the MBICAT algorithm in R. WLS estimators can be utilized by using NOHARM and Mplus with limited-information algorithms, and BMIRT has a Bayesian MCMC estimator (Chalmers, 2012).

2.4 Empirical Data Analysis

In order to show the importance of and provide the rationale for study of skewed latent trait distributions, the proficiency scales of the Program for International Student Assessment (PISA) mathematics, reading, and science tests were investigated. The Organization for Economic Co-operation and Development provides technical and supplementary reports to describe the test construction and report key findings of the assessment, and the National Center for Education Statistics has also published analysis of the PISA results focusing on US students from PISA 2000 to 2009. The data and information used in this part are collected from those reports and modified for this research.

2.4.1 Distribution of Latent Traits in PISA 2003, 2006, and 2009

The Program for International Student Assessment uses proficiency levels to describe student performance. In order to reach a particular level, a student must be able to answer a majority of items correctly at that level. Students are classified into one of the levels according to their scores (OECD, 2001). For example, the reading literacy scale in PISA 2009 has eight cut point scores from Level 1b to Level 6, and students' scores are located on a scale from 0 to 1,000 (NCES, 2010). An example of specific cut scores is shown in Table 2-2.

Table 2-2. Cut scores of Reading Literacy Proficiency Levels in PISA 2009

	Greater than	Less than or equal to
Below level 1b	-	262.04
Level 1b	262.04	334.75
Level 1a	334.75	407.47
Level2	407.47	480.18
Level3	480.18	552.89
Level4	552.89	625.61
Level5	625.61	698.32
Level6	698.32	-

Table 2-3 and Figures 2-1 and 2-2 show the percentage distributions of 15-year-old students in the United States on combined reading, mathematics, and science literacy scales by proficiency level. In the 2000, 2003, and 2006 PISA results, the distribution of reading proficiency followed a negatively skewed distribution, whereas the mathematics and science literacy scales had positively skewed distributions. In PISA 2009, compared to the results in PISA 2000, the reading proficiency scale, which was modified from 6 levels to 8 levels, was generally normally distributed. In PISA 2009, the distribution of mathematics literacy had heavy left tails similar to the distribution from PISA 2003. Although the science literacy distribution

Table 2-3. Percentage Distribution of Proficiency Level Scores in PISA 2000, 2003 and 2009

Reading (%)				Mathematics Science (%)				
2009		2000		2009		2003	2009	2003
Below level 1b	1	Below level 1	4	Below level 1	8	10	4	8
Level 1b	4	Level 1	9	Level 1	15	16	14	17
Level 1a	13	Level 2	20	Level 2	24	24	25	24
Level 2	24	Level 3	27	Level 3	25	24	28	24
Level 3	28	Level 4	24	Level 4	17	17	20	18
Level 4	21	Level 5	16	Level 5	8	8	8	8
Level 5	8	-	-	Level 6	2	2	1	2
Level 6	2	-	-	-	-	-	-	-

* The table is made by the information from NECS, 2001, 2004, 2007, & 2010

has been getting closer to a symmetric distribution over time, compared to the results from PISA 2006, it still is a little skewed with a heavy left tail.

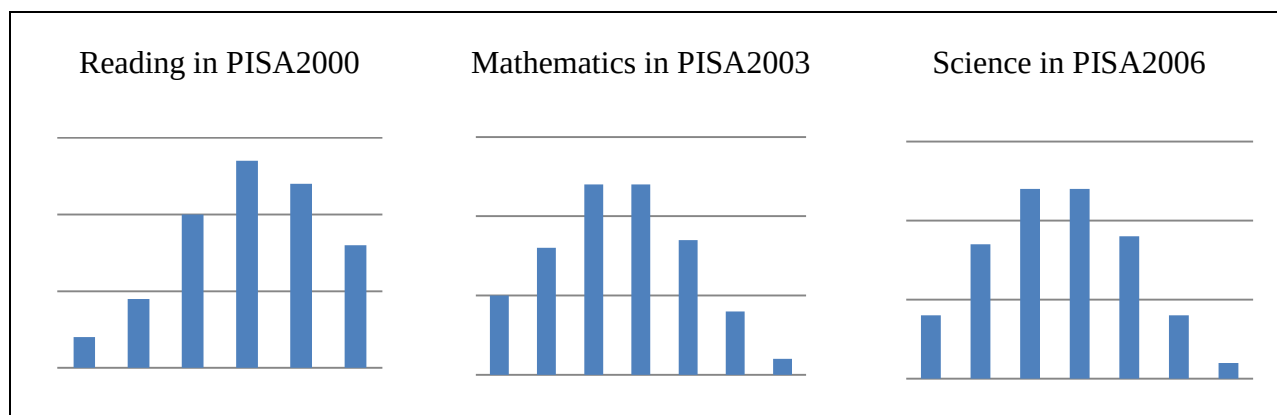


Figure 2-1. Percentage distribution of proficiency level in PISA 2000, 2003 and 2006
(Modified from NCES, 2000, Table A3.7; NCES, 2004, Figure 4; NCES, 2007, Figure 4)

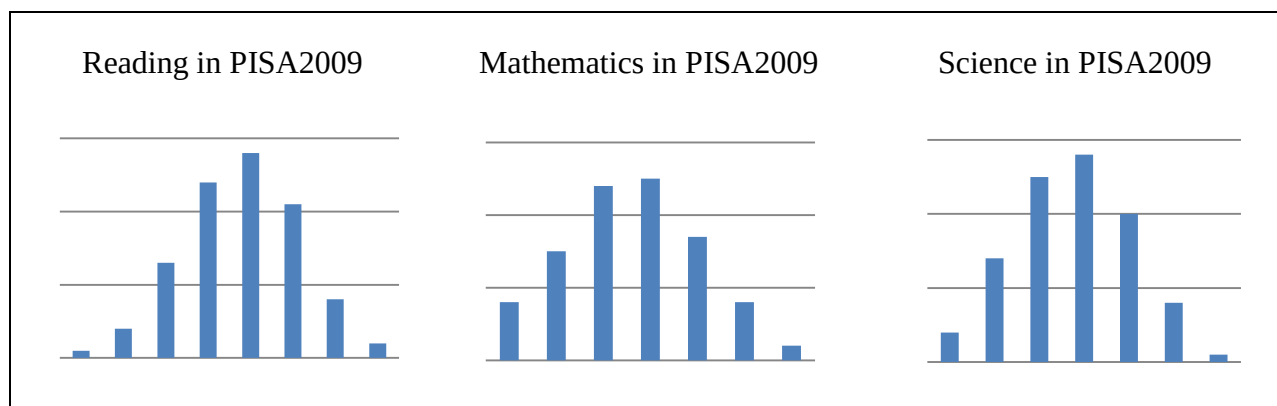


Figure 2-2. Percentage Distribution of Proficiency Level in PISA 2009
(Modified from NCES, 2010, Figure 3, 5, & 7)

2.4.2 Sub-domain Proficiency Levels in PISA

From 2000 to 2009, PISA has measured student's Mathematics and Science literacy with three kinds of sub-domains. The distributions of sub domain performance by proficiency level are shown in Figure 2-3.

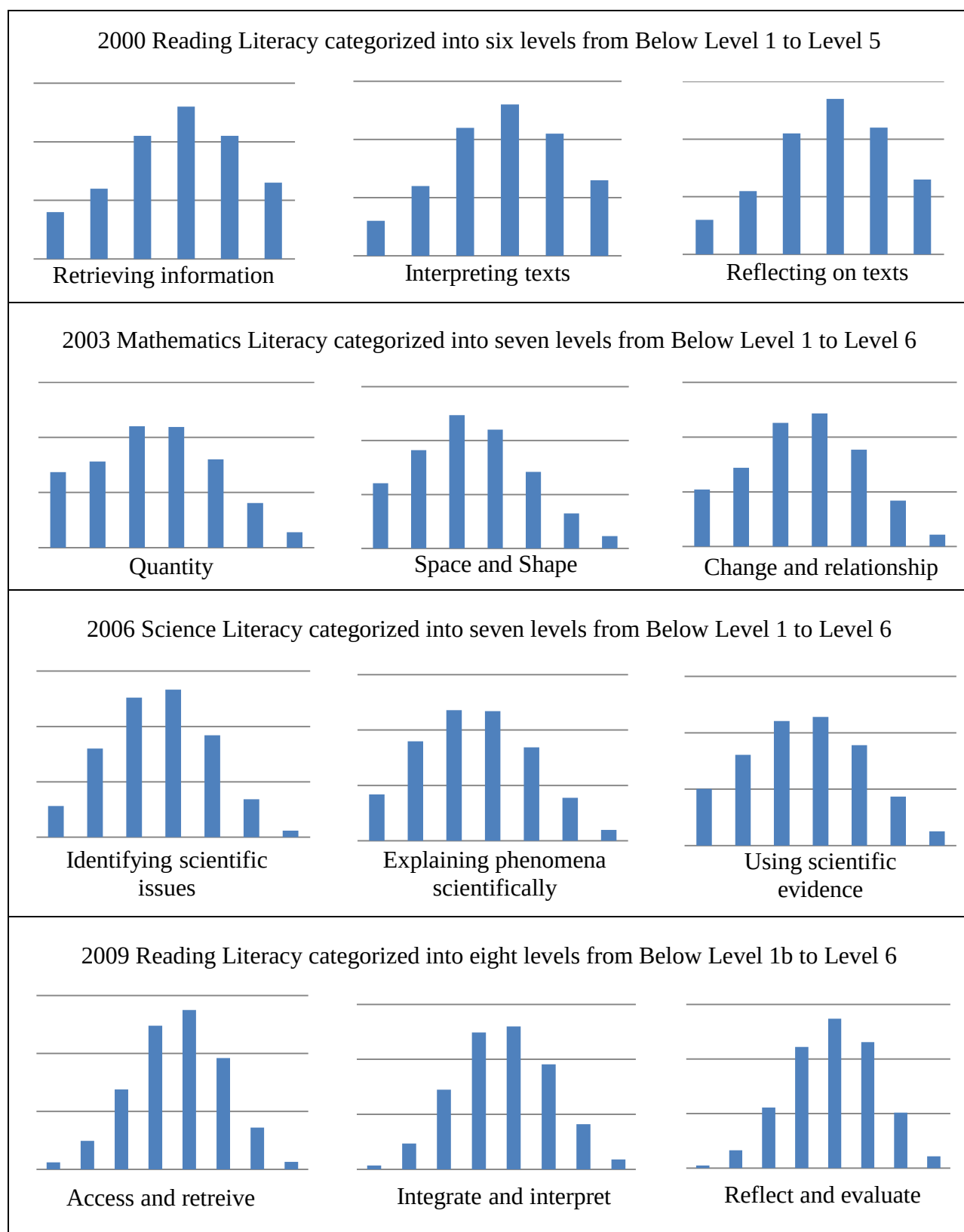


Figure 2-3. Percentage Distribution of Sub-domains in PISA 2000, 2003, 2006, and 2009

The analysis results show that the distributions of sub domain proficiency levels for each subject show different patterns. Reading literacy in PISA 2000 had negatively skewed distributions with long tails to the left for all three sub-domains. Mathematics literacy in PISA 2003 had a positively skewed distribution with a heavy tail in the lower level of proficiency for the three domains, whereas they had peak points at different proficiency levels. Science literacy in PISA 2006 showed similar positively skewed distributions for the three domains, but with slightly different kurtosis. Reading literacy in PISA 2009, with eight levels of proficiency, showed almost a symmetric distribution, and it showed quite a different distribution from reading literacy in PISA 2000, which had seven levels of proficiency.

In summary, in PISA 2003, 2006, and 2009 the distributions of reading are negatively skewed, with a heavy tail on the high levels of proficiency. The distributions of mathematics and science are positively skewed, with the heavy tails toward the low levels of proficiency.

In this analysis of sub-domain proficiency levels in PISA, the distributions of the sub-domains show different patterns by subject, and the three domains in each subject show slightly different distributional properties. While reading in 2000 had negatively skewed distributions for the three domains, each distribution had different levels of skewedness and thickness of its tails. The subdomain distributions for math in 2003 had common heavy tails on the lower level; however, each distribution had different points with the highest frequencies. Science in 2006 had a similar pattern for the three domains, but with different levels of kurtosis, and reading in 2009 also had a similar pattern of symmetric distribution for three domains, with the different level of kurtosis.

The results of this PISA score distribution analysis show that the distributions of the latent trait scores can have different shapes for each subject matter, and distributions within the sub-domains of each subject can have different properties. Within the same general construct, such as math, science or reading, their sub-domain proficiency level scores had different distributional properties, especially related to the skewedness and kurtosis of the distributions.

3. Method

3.1 Data Generation

In order to study the effect of different distributional assumptions when a multidimensional latent trait model is estimated, I generated a set of latent trait distributions and item parameter sets corresponding to the simulation conditions. Previous literature and empirical analyses were used for selecting the true item and person parameters in order to allow the simulation data to be representative of the reality. Table 3-1 shows the simulation conditions for data generation. The model used for estimating the item and latent trait parameters is a two-parameter logistic bifactor model. There are three latent traits in each model, one general factor and two specific factors. For the latent trait distributions, normal and skewed distributions are generated. Simulation conditions related to the item parameters consist of two levels of item difficulty and item discriminations.

3.1.2 Data Generating of Latent Trait Distributions

The bifactor model designed for this research has three latent factors including one general factor and two specific factors and three latent trait distributions in every replication of the simulation study. The trait distributions are generated according to combinations characterized by (a) the normality (shape) of the general factor distribution, (b) the direction of skewedness of the skewed distributions included in the general or specific factors, and (c) the correlations between the specific factors. The general factor distribution in each condition is one of two shape types: normal or positively-skewed. Each general factor distribution is paired with two specific factor distributions, which are two positively skewed distributions or one positively skewed distribution and one negatively skewed distribution.

The bifactor model assumes that the specific factors are not only uncorrelated with the general factor, but also uncorrelated with each other. In order to evaluate the estimation quality when this assumption is violated, two levels of intercorrelation between latent traits of 0.2 (barely correlated) or 0.8 (correlated) are given for each paired condition. With the two conditions of normality of the general factor, two conditions of direction of skewedness of the specific factor distributions, and two levels of intercorrelations, a total of eight simulation conditions are assigned to the latent trait distributions.

Table 3-2 shows the specific simulation conditions of the latent traits. Based on the simulation conditions, in order to look at the effects of general factor distribution, the results from conditions 1 to 4 and the results from corresponding conditions 5 to 8 are compared. Similarly, the direction effects of the skewed distributions of specific factors are evaluated by the comparison of Condition 1 vs. Condition 3, Condition 2 vs. Condition 4, Condition 5 vs. Condition 7, and Condition 6 vs. Condition 8. The effect of the extent of correlation between the specific factors on estimation are evaluated with the comparison of Condition 1 vs. 2, Condition 3 vs. Condition 4, Condition 5 vs. Condition 6, and Condition 7 vs. Condition 8.

Because the model has item difficulty and discrimination parameters, but no guessing parameter, the effects found from the negatively skewed distribution will be the same as the positively skewed distribution except opposite in sign. For example, the results from the combination of negatively skewed specific factors with normal distribution of general factor are implied by the results from the combination of positively skewed specific factors with normal distribution in Condition 1 and 2.

Table 3-1. Simulation Conditions for Data Generation

Simulation factors	Condition
Model	Two-parameter multidimensional latent trait model - bifactor model with three factors of one general factor and two specific factors
Distribution of Latent Traits with Directions	Normal distribution from standard normal distribution with mean of 0 and standard deviation of 1. Positively or negatively skewed with a mean of 0.3 or -0.3, and skewedness of 0.8 or -0.8
Discrimination Parameter	2 conditions generated from lognormal distributions with the range from 0.5 to 2.5; Mean of 1.3 with standard deviation of 0.15 and mean of 1.8 with standard deviation of 0.15
Difficulty Parameter	2 conditions generated from normal distributions with the range from -2 to 2; Mean of -0.5 with standard deviation of 0.4 and mean of 0.5 with standard deviation of 0.4
Number of Items	Total 60 items with 30 items for each specific factor
Number of Examinees	2000
Estimation method	Full Information Marginal Maximum Likelihood in IRTPRO
Replications	50

Table 3-2. Simulation Combinations of Latent Trait Distributions

Condition Number	General	Specific 1	Specific 2	Correlation
1	Normal	Skewed (+)	Skewed (+)	0.2
2	Normal	Skewed (+)	Skewed (+)	0.8
3	Normal	Skewed (+)	Skewed (-)	0.2
4	Normal	Skewed (+)	Skewed (-)	0.8
5	Skewed (+)	Skewed (+)	Skewed (+)	0.2
6	Skewed (+)	Skewed (+)	Skewed (+)	0.8
7	Skewed (+)	Skewed (+)	Skewed (-)	0.2
8	Skewed (+)	Skewed (+)	Skewed (-)	0.8

* (+) or (-) means a positively or negatively skewed distribution respectively.

The simulation conditions of the latent trait distributions combine the degree of skewedness of the distributions and the intercorrelations needed to describe multivariate latent trait distributions. For example, assume that one general factor and two positively skewed specific factors need to be generated. The procedure of the data generation is as follows:1) a distribution for the general factor is generated from a normal distribution with mean of 0 and standard deviation of 1 or

mean of -0.3 and skewedness of 0.8 for a positively skewed distribution, 2) in order to generate skewed distributions for specific factors with particular levels of correlation between them, two distributions from multivariate normal distributions are generated with correlations of 0.2 and 0.8, and 3) the distributions with the correlations that are generated in 2) are non-linearly transformed into skewed distributions. In this case, the transformed distribution is going to be skewed compared to the normal distribution, but the correlation between the two specific factors is not changed.

As mentioned above, the generated distributions considering intercorrelations are transformed into skewed distributions. In order to transform the correlated distributions into skewed distributions, I applied the idea of the Copula method (Nelsen, 1999). Relational functions between normal and targeted skewed distributions are estimated by using their cumulative probability distributions, and are used in order to transform the multivariate normal distributions with the targeted intercorrelation into skewed distributions. The steps to generate the distributions are as follows:

Step 1. Generating the targeted skewed distribution

Based on conditions in previous studies, positively skewed distributions with their first four moments having values of: mean of -0.3, variance of 1, skewedness of 0.8, and kurtosis of 3.5, are created. For negatively skewed distributions, a mean of 0.3, variance of 1, skewedness of -0.8 and kurtosis of 3.5 are used.

Step 2. Calculating the cumulative probability distribution of the skewed distribution generated in the step 1

In order to calculate a cumulative probability distribution, R-programming is used with the function of ‘ecdf’, which calculates an empirical cumulative distribution.

Step 3. Estimating the regression function between the normal and the skewed probability distributions.

For each positively skewed or negatively skewed distribution, regression coefficients between two distributions were estimated. In this research, the coefficients were estimated fifty times through simulation analysis estimating the function needed to transform from a normal distribution to a skewed distribution (see Tables in Appendix A). The fifty data sets of 2,000 or 10,000 cases generated from normal and skewed distributions showed similar coefficients, and quadratic and cubic functions were identified as the best functions to use for the transformation based on the variance explained in the model fit. Each coefficient of the polynomial functions had a very small amount of variance, which shows that estimated coefficients for each trial were very similar. Finally, the cubic transformation function was selected. The R-squares of the regression functions between the data of normal distributions and the data of skewed distributions were over .999. The means of the fifty coefficients were used, and by using a regression function, the normal distributions were approximately transformed into the skewed distributions. The regression functions are as follows:

In order to transform the normal distribution values (X) into skewed distribution values (Y), two regression functions are used:

for positively skewed distributions,

$$Y = -0.4508 + 1.0167 \times X + 0.1461 \times x^2 - 0.0136 \times x^3, \text{ and}$$

for negatively skewed distributions,

$$Y = -0.4516 + 1.0135 \times X - 0.1500 \times x^2 - 0.0115 \times x^3$$

Step 4. Generating multivariate normal distributions with correlations of 0.2 or 0.8

By using the function to generate a multivariate normal distribution, sets of distributions are generated with correlations of 0.2 or 0.8 between the specific factor distributions.

Step 5. Calculating the skewed distribution by using the function estimated in step 3

The values of the multivariate distributions generated in step 4 are transformed into skewed distributions by using the one of the formulas estimated in step 3.

Step 6. Checking the descriptive statistics and correlations of the generated distributions

If the descriptive statistics and correlations are not acceptable, the parameters of the normal distributions are modified to reach the targeted values for the distributions.

3.1.2 Data Generating of Item Parameters

For the purpose of stability, but allowing some comparison, the discrimination parameter and difficulty parameters are each generated with two levels. Two sets of discrimination parameters are generated to represent “high” and “low” levels of discrimination in an item set. In order to avoid negative values for discrimination parameters, the parameters are generated from lognormal distributions. The lognormal distribution is a log-transformed distribution from a normal distribution (Hogg & Tanis, 1997), and the probability density function is

$$f(X) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, X > 0$$

with two parameters of μ and σ , and the mean and variance of a lognormal distribution are calculated by

Table 3-3. Parameters for Generating Distributions of Simulation Combinations of Item Parameters

	Discrimination from 0.5 to 2.5		Difficulty from -2 to 2	
	Mean	SD	Mean	SD
1	1.3	0.15	-0.5	0.4
2	1.3	0.15	0.5	0.4
3	1.8	0.15	-0.5	0.4
4	1.8	0.15	0.5	0.4

$$E(X) = e^{\mu + \frac{\sigma^2}{2}} \text{ and } \text{Var}(X) = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}.$$

Therefore, with the specific target value of mean, $E(X)$, and variance (squared standard deviation), $\text{Var}(X)$, the parameters of μ and σ are calculated by using the two formulas:

$$\mu = \ln(E(X)) - \ln\left(\sqrt{\frac{\text{Var}(X)}{E(X)^2} + 1}\right), \text{ and } \sigma = \sqrt{\ln\left(\frac{\text{Var}(X)}{E(X)^2} + 1\right)}$$

For example, in order to generate a distribution of discrimination parameters with a mean of 1.3 and a standard deviation of 0.15 from a lognormal distribution, the μ of -0.13118 and σ of 0.51222 should be used. Similarly, the discrimination parameters in this study are generated from a lognormal distribution with a mean of 1.3 or 1.8 and a standard deviation of 0.15.

For the difficulty parameter, normal distributions with a mean of -0.5 or 0.5 and a standard deviation of 0.5 are used. To avoid violating the latent trait model assumption that the function between the latent trait and the probability of the correct answer is monotonically

increasing, the variances of the distribution for generating item parameters are manipulated by giving the range of the difficulty parameter distribution. The distributions of the created discrimination parameters ranged from 0.5 to 2.5, and those of the created difficulty parameters ranged from -2 to 2. The combinations of item parameters are shown in Table 3-4.

3.1.3 Data Generating of Examinees' Responses

The estimation performance of the bifactor model is evaluated via the comparison between the data generated under the simulation conditions and the data estimated from the bifactor model. In other words, the comparison means to investigate how the estimated parameters from the bifactor model are close to the generated data under the simulation conditions. In order to estimate parameters by using the bifactor model, examinees' responses are generated based on the item and θ parameters generated under the simulation conditions.

Item parameters and θ parameters are plugged in to the function of the bifactor model in order to calculate the probability of answering correctly:

$$P(u=1|\theta_0, \theta_1, \theta_2, a_0, a_1, a_2, d_i) = \frac{1}{1 + \exp\{-[d_i + a_0\theta_0 + a_1\theta_1 + a_2\theta_2]\}}$$

$$, \text{ where } d_i = -b \sqrt{\sum_{v=0}^2 a_v^2}$$

As a result, each examinee has a probability to answer correctly for each item. To add randomness to each value, random variables from uniform distributions ranged from 0 to 1 are generated and assigned to each response probability value. If the probability is higher than the random value, the corresponding response is assigned as 1, which means that the examinee

answers that question correctly. If the probability is lower than the random value, the response is assigned as 0, which means that the examinee responds with the wrong answer.

The total number of items is 60, which are divided so that 30 items are indicators for each of the two specific factors, and the number of examinees in each data set is 2000. The data sets are generated by using R-programing and the full information marginal maximum likelihood estimation (MML) implemented in IPRPRO is used for estimating the parameters. 50 replications of the simulation study are conducted to achieve stable estimation of the results.

3.2 Evaluation Methods

The generated data are compared to the true parameters in order to confirm if the generated data sets represent the planned simulation design. Descriptive statistics for the generated data, such as the mean, standard deviation, minimum, maximum, skewedness and kurtosis, will be provided.

In order to evaluate the estimation precision of the model under the designed conditions, θ and item estimated parameters are compared to the true parameters assigned. Means and variances of mean bias were used to judge the precision of parameter estimation. The formula for mean bias is as follows.

$$\text{Mean Bias} = \frac{1}{k} \sum_{i=1}^k (\hat{\theta}_i - \theta_i),$$

where θ is the given parameter, $\hat{\theta}$ is the estimate of the parameter, and k is the number of parameters, for example, $k=30$ for discrimination item parameters of each specific factor. The mean and variance of mean bias were calculated across the replications. Bias is the index showing that the difference between the parameter and its estimates. To judge overall bias across the parameters estimated, mean bias is calculated by the average of the differences.

For the investigation of the distributional difference between the estimated θ distributions and the generated θ distribution, the Kolmogorov-Smirnov (KS) test was utilized. The KS test is used to compare two empirical distributions by using their cumulative functions (Stapleton, 2008; Hogg & Tanis, 1997). In order to compare the generated and estimated θ distribution, the KS test was utilized with the entire θ parameter set, and for specific value ranges of the θ parameters. For the KS test with the entire data set, a total of 2000 parameters for each analysis were tested. For the specific ranges, 2,000 θ parameters were sorted by their locations, and sets of 200 parameters were sequentially assigned to each θ category. Thereby, ten specific θ value categories were constructed for the KS tests. All categories had the same frequency of θ parameters; however, that does not mean that their θ continuums had the same width, because depending on the generated distribution, the frequencies in a certain fixed θ range can be different. Every simulation condition was replicated fifty times, and values showing the numbers of frequencies from the fifty replications that were statistically significant under the significance level of 0.05 will be reported.

4. Results

4.1. Data Generation

For the simulation study, the item and θ parameters were generated, and those values were used to generate response strings through the bifactor models. The descriptive statistics of the generated item and θ parameters are provided.

4.1.1 Item Parameter Generation

The bifactor model used in this study includes three item discrimination parameters related to general or specific factors, and one d parameter calculated using the discrimination and difficulty parameters. Table 4-1 shows the descriptive statistics of the generated discrimination and difficulty parameters. In order to check the range of the generated data, and to avoid extreme values, minimum, mean, and maximum values were calculated.

The discrimination parameters were generated within the range from 0.5 and 2.5 from a lognormal distribution with a mean of 1.3 and 1.8, and a standard deviation of 0.15. The mean, and standard deviation statistics showed that the generated discrimination parameters of the general and the two specific factors had means and standard deviations very close to those of the generating distribution in each simulation condition. For the discrimination parameters generated from a lognormal distribution with a mean of 1.3, the means of three discrimination parameter sets were 1.298, 1.309, and 1.304, which were very close to the simulation condition of 1.3; and for the parameters from a distribution with a mean of 1.8, the three means were 1.802, 1.804, and 1.796. All of the discrimination parameters also had values very close to the parameter of 1.5. No values were out of the range from 0.5 and 2.5. The difficulty parameters were generated from normal distributions with a mean of 0.5 or -0.5, and a standard deviation of 0.4 within the range

of -2 to 2. The generated difficulty parameters showed the proper level of mean and standard deviation for the data sets. The means of the difficulty parameter sets were 0.497 and 0.495 for the mean parameter of 0.5, and they were -0.515 and -0.497 for the mean parameter of -0.5. Their standard deviations were close to 0.4, and there were no extreme values outside the range of -2 to 2.

4.1.2 θ Parameter Generation

There were eight types of ability distributions under the conditions characterized by the degree of skewedness of the distributions, the direction of skewedness, and the correlation between specific factors. The normal distributions were generated from a standard normal distribution with a mean of 0 and a standard deviation of 1. The parameter values for the skewed distributions were means of 0.3 for the negatively skewed distributions, and -0.3 for the positively skewed distributions, standard deviations of 1, skewedness of 0.8, and correlations between the specific factors of 0.2 or 0.8.

The descriptive statistics of the generated θ parameters are provided in Tables 4-2, 4-3, and 4-4. The generated ability distributions showed means standard deviations, skewedness, and correlations near their anticipated values. The general factor distributions were generated from normal distributions or positively skewed distributions (See Table 4-2). The means of the normal distribution means were from -0.002 to 0.010, and the means of their standard deviations were from 0.997 to 1.003. The positively skewed distributions had means that ranged from -0.308 to -0.302, and means of skewedness that ranged from 0.798 to 0.807.

Table 4-1. Descriptive Statistics of Generated Item Parameters

	Mean (SD)	Factors	Mean	SD	Min	Max
Disc. Parameters	1.3 (0.15)	General	Min	1.245	0.128	0.841
			Mean	1.298	0.151	0.994
			Max	1.331	0.188	1.106
		Specific1	Min	1.224	0.096	0.858
			Mean	1.309	0.151	1.028
			Max	1.372	0.196	1.174
		Specific2	Min	1.235	0.114	0.870
			Mean	1.304	0.149	1.018
			Max	1.369	0.217	1.132
	1.8 (0.15)	General	Min	1.764	0.106	1.349
			Mean	1.802	0.148	1.485
			Max	1.840	0.179	1.583
		Specific1	Min	1.739	0.111	1.411
			Mean	1.804	0.151	1.518
			Max	1.880	0.192	1.617
		Specific2	Min	1.734	0.112	1.344
			Mean	1.796	0.156	1.493
			Max	1.888	0.211	1.621
Diff. Parameters	-0.5 (0.4)	Specific1	Min	-0.625	0.281	-1.993
			Mean	-0.515	0.405	-1.330
			Max	-0.382	0.534	-0.901
		Specific2	Min	-0.698	0.323	-1.719
			Mean	-0.497	0.395	-1.315
			Max	-0.287	0.491	-0.937
	0.5 (0.4)	Specific1	Min	0.340	0.301	-0.700
			Mean	0.497	0.399	-0.293
			Max	0.601	0.501	-0.013
		Specific2	Min	0.294	0.289	-0.637
			Mean	0.495	0.386	-0.281
			Max	0.710	0.519	0.044

* Disc. Parameters: Discrimination parameters; Diff. parameters: difficulty parameters

Although this study didn't consider kurtosis as a simulation factor, the skewedness of the distributions affected the kurtosis of the distributions. Standard normal distributions with a mean of 0 and standard deviation of 1 have skewedness of 0 and kurtosis of 3 when the standardized fourth moment is used for the formula of the kurtosis (DeCarlo, 1997). Therefore the kurtosis values of the normal distributions from conditions 1 to 4 in Table 4.2 show values acceptable to be regarded as a standard normal distribution. From conditions 5 to 8, the means of kurtosis values range from 3.551 to 3.646, which are higher than the values for the non-skewed distribution, showing that the kurtosis values are related to the skewedness of the distribution.

Based on the simulation conditions, all of the first specific distributions were positively skewed (See Table 4-3). The generated factor distributions had means of means that ranged from -0.305 to -0.297, means of the standard deviations that ranged from 0.995 to 1.004, and the means of skewedness that ranged from 0.788 to 0.805.

The data sets of second specific factors were generated from positively skewed distributions and negatively skewed distributions (See Table 4-4). For the positively skewed distributions, the means of the distributions were from -0.309 to -0.300, and the means of skewedness were from 0.783 to 0.807. For the negatively skewed distributions, the means were from 0.296 to 0.308, and the means of skewedness were -0.834 to -0.812.

The descriptive statistics for the correlations between specific factors in each of the eight distributional simulation conditions of θ parameters showed that those distributions had values of correlations close to 0.2 or 0.8 (See Table 4-5). The mean correlations of distributions generated with the correlation of 0.2 were 0.192, 0.199, 0.194, and 0.197, and those generated with the correlation of 0.8 were 0.792, 0.794, 0.791, and 0.793.

Table 4-2. Descriptive Statistics of General θ Parameters Generated

Cond.	Distribution	Stat	Mean	SD	Min	Max	Skew	Kurtosis
1	Normal	Min	-0.043	0.960	-4.371	2.845	-0.151	2.841
		Mean	-0.002	0.997	-3.431	3.397	0.006	2.988
		Max	0.048	1.041	-2.889	4.131	0.105	3.335
2	Normal	Min	-0.047	0.962	-4.430	2.775	-0.097	2.733
		Mean	-0.002	1.003	-3.405	3.440	0.005	2.967
		Max	0.046	1.037	-2.807	4.333	0.151	3.239
3	Normal	Min	-0.057	0.958	-4.587	2.878	-0.121	2.751
		Mean	-0.001	0.998	-3.478	3.439	-0.005	3.005
		Max	0.034	1.027	-2.952	3.979	0.111	3.461
4	Normal	Min	-0.035	0.954	-4.149	2.968	-0.094	2.789
		Mean	0.010	1.002	-3.387	3.481	0.004	2.997
		Max	0.063	1.045	-2.929	5.050	0.094	3.264
5	Positively Skewed	Min	-0.347	0.961	-1.868	3.266	0.706	3.288
		Mean	-0.305	0.999	-1.868	4.192	0.807	3.616
		Max	-0.253	1.045	-1.866	5.656	0.938	4.202
6	Positively Skewed	Min	-0.361	0.970	-1.868	3.192	0.683	3.091
		Mean	-0.308	0.996	-1.868	4.243	0.798	3.579
		Max	-0.261	1.036	-1.866	5.838	0.914	4.075
7	Positively Skewed	Min	-0.347	0.964	-1.868	3.427	0.698	3.176
		Mean	-0.302	0.999	-1.868	4.267	0.802	3.602
		Max	-0.260	1.040	-1.864	5.781	0.978	4.425
8	Positively Skewed	Min	-0.350	0.957	-1.868	3.352	0.681	3.119
		Mean	-0.305	0.999	-1.868	4.151	0.798	3.551
		Max	-0.249	1.036	-1.867	5.378	0.923	4.020

* Cond.: Numbers of θ simulation conditions

Table 4-3. Descriptive Statistics of First Specific θ Parameters Generated

Cond.	Distribution	Stat	Mean	SD	Min	Max	Skew	Kurtosis
1	Positively Skewed	Min	-0.363	0.949	-1.868	3.133	0.670	3.028
		Mean	-0.300	0.995	-1.868	4.185	0.788	3.546
		Max	-0.259	1.033	-1.866	5.981	0.939	4.119
2	Positively Skewed	Min	-0.341	0.969	-1.868	3.395	0.669	3.056
		Mean	-0.297	1.002	-1.868	4.262	0.794	3.578
		Max	-0.259	1.035	-1.865	5.544	0.973	4.468
3	Positively Skewed	Min	-0.343	0.967	-1.868	3.182	0.679	3.060
		Mean	-0.303	1.004	-1.868	4.190	0.805	3.591
		Max	-0.247	1.044	-1.867	5.572	0.985	4.569
4	Positively Skewed	Min	-0.364	0.967	-1.868	3.408	0.698	3.177
		Mean	-0.298	0.999	-1.868	4.242	0.803	3.607
		Max	-0.251	1.051	-1.867	5.568	0.939	4.339
5	Positively Skewed	Min	-0.350	0.965	-1.868	3.228	0.613	2.960
		Mean	-0.300	1.000	-1.868	4.172	0.790	3.526
		Max	-0.243	1.041	-1.867	5.467	0.897	3.932
6	Positively Skewed	Min	-0.352	0.970	-1.868	3.461	0.691	3.219
		Mean	-0.301	1.001	-1.868	4.150	0.801	3.563
		Max	-0.254	1.050	-1.867	5.174	0.893	3.969
7	Positively Skewed	Min	-0.352	0.955	-1.868	3.310	0.681	3.097
		Mean	-0.305	0.996	-1.868	4.268	0.799	3.609
		Max	-0.258	1.044	-1.866	5.717	0.912	4.165
8	Positively Skewed	Min	-0.358	0.962	-1.868	3.418	0.727	3.235
		Mean	-0.305	1.000	-1.868	4.266	0.803	3.601
		Max	-0.230	1.041	-1.865	5.572	0.957	4.473

* Cond.: Numbers of θ simulation conditions

Table 4-4. Descriptive Statistics of Second Specific θ Parameters Generated

Cond.	Distribution	Stat	Mean	SD	Min	Max	Skew	Kurtosis
1	Positively Skewed	Min	-0.359	0.966	-1.868	3.262	0.660	3.129
		Mean	-0.309	0.997	-1.868	4.170	0.785	3.518
		Max	-0.270	1.031	-1.866	5.279	0.928	4.313
2	Positively Skewed	Min	-0.348	0.965	-1.868	3.540	0.623	3.127
		Mean	-0.300	1.001	-1.868	4.188	0.788	3.543
		Max	-0.257	1.042	-1.866	5.323	0.929	4.107
3	Negatively Skewed	Min	0.269	0.967	-5.797	1.870	-0.980	3.240
		Mean	0.305	1.000	-4.391	1.871	-0.821	3.664
		Max	0.365	1.035	-3.324	1.871	-0.721	4.501
4	Negatively Skewed	Min	0.253	0.963	-6.221	1.870	-1.012	3.074
		Mean	0.308	0.996	-4.358	1.871	-0.819	3.690
		Max	0.364	1.030	-3.245	1.871	-0.663	4.922
5	Positively Skewed	Min	-0.353	0.966	-1.868	3.127	0.643	3.021
		Mean	-0.303	0.999	-1.868	4.107	0.783	3.534
		Max	-0.252	1.034	-1.868	5.850	0.951	4.200
6	Positively Skewed	Min	-0.354	0.960	-1.868	3.458	0.630	2.979
		Mean	-0.305	0.998	-1.868	4.356	0.807	3.618
		Max	-0.250	1.047	-1.867	5.655	0.939	4.297
7	Negatively Skewed	Min	0.250	0.967	-5.539	1.870	-0.958	3.035
		Mean	0.296	1.004	-4.295	1.871	-0.812	3.607
		Max	0.347	1.040	-3.358	1.871	-0.702	4.140
8	Negatively Skewed	Min	0.242	0.942	-5.740	1.870	-0.965	3.181
		Mean	0.301	1.002	-4.442	1.871	-0.834	3.720
		Max	0.368	1.041	-3.516	1.871	-0.680	4.688

* Cond.: Numbers of θ simulation conditions

Table 4-5. Correlations Between Two Specific θ Parameters Generated

Condition	1	2	3	4	5	6	7	8
Correlation	0.2	0.8	0.2	0.8	0.2	0.8	0.2	0.8
Min	0.133	0.772	0.149	0.781	0.151	0.771	0.158	0.776
Mean	0.192	0.792	0.199	0.794	0.194	0.791	0.197	0.793
Max	0.231	0.828	0.258	0.806	0.241	0.806	0.227	0.807

4.2. Bifactor Analysis

For each condition of the latent trait distributions and item parameters, item and θ parameter estimation was evaluated. The tables in the body text show the mean and variance of the mean bias of the parameter estimates. More details of the descriptive statistics such as minimum and maximum values of the mean and variance are attached in Appendix B, C and D.

4.2.1 Item Parameters

For evaluating the bifactor model under the different distributional conditions, the mean and variance of the mean bias of the estimated parameters were calculated. Table 4-6 and 4-7 show means and variances of item parameter mean bias under each of eight θ simulation conditions, and more detailed statistics are provided in Appendix B. Among the simulation conditions, there were four noticeable patterns of item parameter estimation.

Table 4-6. Means of Item Parameter Mean Bias

Cond.		1	2	3	4	5	6	7	8
θ Cond.	G	(0)	(0)	(0)	(0)	(+)	(+)	(+)	(+)
	S1	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)
	S2	(+)	(+)	(-)	(-)	(+)	(+)	(-)	(-)
	Corr.	0.2	0.8	0.2	0.8	0.2	0.8	0.2	0.8
Disc: 1.3 Diff: -0.5	a0	0.121	0.391	0.137	0.407	0.118	0.382	0.110	0.395
	a1	-0.207	-0.714	-0.190	-0.750	-0.144	-0.715	-0.170	-0.736
	a2	-0.166	-0.715	-0.116	-0.689	-0.188	-0.681	-0.146	-0.717
	d	-0.433	-0.431	0.005	0.023	-0.840	-0.845	-0.420	-0.426
Disc: 1.3 Diff: 0.5	a0	0.143	0.444	0.134	0.413	0.206	0.488	0.171	0.449
	a1	-0.097	-0.678	-0.123	-0.713	-0.109	-0.629	-0.088	-0.671
	a2	-0.125	-0.634	-0.171	-0.728	-0.056	-0.659	-0.210	-0.754
	d	-0.443	-0.438	-0.003	0.008	-0.872	-0.860	-0.428	-0.430
Disc: 1.8 Diff: -0.5	a0	0.216	0.494	0.282	0.538	0.202	0.502	0.185	0.508
	a1	-0.308	-0.978	-0.289	-1.007	-0.284	-0.954	-0.235	-1.000
	a2	-0.321	-0.979	-0.215	-0.957	-0.274	-0.955	-0.290	-0.993
	d	-0.611	-0.611	-0.001	0.054	-1.204	-1.195	-0.602	-0.586
Disc: 1.8 Diff: 0.5	a0	0.305	0.615	0.266	0.542	0.421	0.709	0.334	0.617
	a1	-0.208	-0.891	-0.214	-0.961	-0.145	-0.854	-0.177	-0.921
	a2	-0.208	-0.891	-0.214	-0.961	-0.145	-0.854	-0.177	-0.921
	d	-0.635	-0.615	-0.015	0.009	-1.278	-1.249	-0.623	-0.620

* Cond.: Numbers of θ simulation conditions

* G, S1, & S2: Distributions of general, first specific, and second specific factors

* a0, a1, and a2: Discrimination parameter for general, first and second specific traits; d: d-parameter

* (0): Standard normal distribution; (+): Positively skewed distributions; (-): Negatively skewed distributions

* Corr.: Correlation between specific factor distributions

Table 4-7. Variances of Item Parameter Mean Bias

Cond.		1	2	3	4	5	6	7	8
θ Cond.	G	(0)	(0)	(0)	(0)	(+)	(+)	(+)	(+)
	S1	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)
	S2	(+)	(+)	(-)	(-)	(+)	(+)	(-)	(-)
	Corr.	0.2	0.8	0.2	0.8	0.2	0.8	0.2	0.8
Disc: 1.3 Diff: -0.5	a0	0.019	0.018	0.025	0.021	0.020	0.019	0.025	0.022
	a1	0.013	0.020	0.013	0.026	0.011	0.020	0.011	0.021
	a2	0.012	0.020	0.013	0.020	0.012	0.019	0.013	0.020
	d	0.008	0.008	0.204	0.210	0.009	0.009	0.196	0.209
Disc: 1.3 Diff: 0.5	a0	0.031	0.024	0.022	0.022	0.030	0.021	0.022	0.023
	a1	0.014	0.022	0.014	0.022	0.017	0.020	0.014	0.022
	a2	0.015	0.021	0.012	0.025	0.014	0.023	0.014	0.030
	d	0.011	0.011	0.207	0.213	0.016	0.015	0.209	0.220
Disc: 1.8 Diff: -0.5	a0	0.020	0.023	0.026	0.034	0.028	0.026	0.045	0.030
	a1	0.017	0.023	0.016	0.027	0.015	0.020	0.016	0.025
	a2	0.014	0.023	0.018	0.025	0.015	0.022	0.018	0.022
	d	0.011	0.011	0.415	0.423	0.012	0.011	0.393	0.412
Disc: 1.8 Diff: 0.5	a0	0.031	0.032	0.025	0.033	0.027	0.034	0.036	0.043
	a1	0.016	0.020	0.018	0.025	0.019	0.026	0.019	0.027
	a2	0.017	0.023	0.016	0.031	0.020	0.027	0.017	0.035
	d	0.018	0.018	0.410	0.425	0.028	0.028	0.437	0.449

* Cond.: Numbers of θ simulation conditions

* G, S1, & S2: Distributions of general, first specific, and second specific factors

* a0, a1, and a2: Discrimination parameter for general, first and second specific traits; d: d-parameter

* (0): Standard normal distribution; (+): Positively skewed distributions; (-): Negatively skewed distributions

* Corr.: Correlation between specific factor distributions

First, the degree of skewedness of the general factor was influential in the estimation of d-parameters. Table 4-6 shows the mean of the mean bias related to the item parameters. Under conditions 5, 6, 7, and 8 having the skewed general factor distribution, the d-parameter estimates show a larger amount of the mean bias than under conditions 1, 2, 3, and 4 having the normal distribution of the general factors. Table 4-7 includes the variances of the mean bias of item parameter estimates. When the variances of d-parameter estimates are compared across the conditions, there is no significant pattern with respect to the d-parameter variances, and this result shows that the degree of skewedness of the general factor is influential not in the variances of d-parameters biases but in the means of d-parameter biases.

Second, the condition of the skewedness directions combined with the skewed specific factor distributions was influential in d-parameter estimations. When two specific factors had distributions with the same direction, for example, two positively skewed distributions or two negatively skewed distributions, the d-parameters had larger biases, the values of which under the conditions 1, 2, 5, and 6 shown in Table 4-6, whereas the d-parameters had smaller amounts of bias when the directions of the skewed distributions were different under the conditions 3, 4, 7, and 8. The amounts of bias increased when the general factor distributions were also skewed. The results for the d parameters under conditions 1 to 4 had smaller amounts of bias than the results under conditions 5 to 8.

Also, the direction of skewedness of the skewed distributions affects the variance of the d parameters. As shown in Table 4-7, the d parameters had large variances under conditions 3, 4, 7, and 8 when the two specific trait distributions had different directions of skewedness. On the other hand, the variance of the discriminations (a_0 , a_1 and a_2) related to the general, first and

second specific factors had no significant patterns depending on the directions of skewedness of the distributions. These patterns were shown across all four item parameter conditions.

Third, the strength of the correlation between the specific factors had a noticeable effect on the estimation of discrimination parameters. The distributional conditions with a high correlation of 0.8 between the specific factors had larger amounts of mean bias in item discrimination parameter estimation than the conditions with a smaller correlation of 0.2 between the specific factors. Table 4-6 shows that the discrimination parameter estimates under conditions 2, 4, 6, and 8 with a high correlation of 0.8 had larger mean biases than the results under conditions 1, 3, 5, and 7 with a lower correlation of 0.2. For example, under the mean discrimination parameter of 1.3 and difficulty parameter of -0.5 condition in Table 4-6, with a low correlation of 0.2 between the specific factors, mean biases of the discrimination parameters related to the general factor range from 0.110 to 0.137 for conditions 1, 3, 5, and 7. However, corresponding the range of the mean biases with the high correlation is from 0.382 to 0.395 under conditions 2, 4, 6, and 8. This pattern was found regardless of the item parameter combination. The high correlation also affects the variance of discrimination parameter estimates for the specific factors. The discrimination parameters had large amounts of variance in the mean bias when there was a high level correlation between the two specific factors under conditions 2, 4, 6, and 8 in Table 4-7.

Lastly, generally the item discrimination parameters related to the general factor were overestimated, whereas the discrimination parameters related to specific factors and d parameters were underestimated. Negative or positive values of bias indicate underestimation or overestimation, respectively, because bias is the result of subtracting a parameter from its estimated value. In Table 4-6, all mean biases of the discrimination parameters related to the

general factors were positive values, which means the parameters tend to be overestimated. Most of the discrimination parameters related to the specific factors and d parameters had negative mean biases, except for some d parameters, especially under conditions 3 and 4.

Based on the results, it was demonstrated that the degree of skewedness of the general factor distributions, the skewedness directions of the specific factor distributions, and the correlation between the specific factor distributions are influential in estimating the item parameters. No noticeable pattern of item parameter estimates across the four item parameter conditions was found.

4.2.2 θ Parameters

Similar to the item parameter estimation, θ parameter estimation was evaluated under the eight distributional conditions across four item parameter conditions. In this section, the mean of the mean biases, variance of the mean biases, and correlation between the generated and estimated trait distributions for the general, first specific and second specific factors are investigated.

a. Mean of Mean Biases

The results for mean bias of the θ parameters are shown in Table 4-8. The θ parameters were generated from a standard normal distribution, from a negatively skewed distribution with a mean of 0.3, or from a positively skewed distribution with a mean of -0.3. The mean values of mean biases in Table 4-8 are very close to 0, -.3, or .3. They are the discrepancies from 0 that is the mean of a standard normal distribution to the simulation parameters. Item response functions are manipulated by the item and θ parameters, but the θ continuum is not a fixed scale, or an arbitrary one. Because of this indeterminacy, the estimation procedure should select the method

Table 4-8. Means of θ Parameter Mean Bias

Cond.		1	2	3	4	5	6	7	8
θ Cond.	G	(0)	(0)	(0)	(0)	(+)	(+)	(+)	(+)
	S1	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)
	S2	(+)	(+)	(-)	(-)	(+)	(+)	(-)	(-)
	Corr.	0.2	0.8	0.2	0.8	0.2	0.8	0.2	0.8
Disc: 1.3 Diff: -0.5	G	0.003	0.006	0.002	-0.010	0.305	0.308	0.303	0.308
	S1	0.301	0.297	0.304	0.298	0.299	0.301	0.305	0.305
	S2	0.309	0.300	-0.305	-0.308	0.304	0.305	-0.295	-0.301
Disc: 1.3 Diff: 0.5	G	0.001	0.004	0.000	-0.010	0.303	0.303	0.301	0.302
	S1	0.301	0.297	0.303	0.298	0.299	0.301	0.303	0.305
	S2	0.307	0.300	-0.306	-0.308	0.302	0.305	-0.296	-0.302
Disc: 1.8 Diff: -0.5	G	-0.001	0.006	0.006	-0.016	0.308	0.309	0.306	0.301
	S1	0.301	0.297	0.306	0.298	0.302	0.301	0.308	0.305
	S2	0.305	0.300	-0.302	-0.308	0.304	0.305	-0.296	-0.301
Disc: 1.8 Diff: 0.5	G	0.001	0.002	0.001	-0.010	0.311	0.308	0.301	0.300
	S1	0.300	0.296	0.304	0.297	0.303	0.301	0.304	0.305
	S2	0.308	0.300	-0.306	-0.308	0.305	0.305	-0.297	-0.302

* Cond.: Numbers of θ simulation conditions

* G: General factor distribution; S1: First specific factor distributions; S2: Second specific distributions

* (0): Standard normal distribution; (+): Positively skewed distributions; (-): Negatively skewed distributions

* Corr.: Correlation between specific factor distributions

Table 4-9. Variances of θ Parameter Mean Bias

Cond.		1	2	3	4	5	6	7	8
θ Cond.	G	(0)	(0)	(0)	(0)	(+)	(+)	(+)	(+)
	S1	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)
	S2	(+)	(+)	(-)	(-)	(+)	(+)	(-)	(-)
	Corr.	0.2	0.8	0.2	0.8	0.2	0.8	0.2	0.8
Disc: 1.3 Diff: -0.5	G	0.383	0.546	0.408	0.570	0.399	0.559	0.413	0.579
	S1	0.471	0.851	0.491	0.889	0.457	0.832	0.471	0.874
	S2	0.469	0.856	0.455	0.816	0.470	0.845	0.454	0.833
Disc: 1.3 Diff: 0.5	G	0.419	0.574	0.403	0.570	0.412	0.558	0.395	0.558
	S1	0.454	0.823	0.448	0.808	0.468	0.803	0.450	0.802
	S2	0.463	0.832	0.490	0.903	0.466	0.813	0.509	0.922
Disc: 1.8 Diff: -0.5	G	0.374	0.542	0.399	0.577	0.396	0.559	0.423	0.591
	S1	0.479	0.852	0.506	0.909	0.470	0.841	0.483	0.894
	S2	0.484	0.862	0.461	0.812	0.476	0.850	0.468	0.835
Disc: 1.8 Diff: 0.5	G	0.415	0.578	0.398	0.575	0.411	0.560	0.390	0.559
	S1	0.461	0.817	0.451	0.800	0.472	0.809	0.455	0.796
	S2	0.466	0.830	0.515	0.927	0.479	0.820	0.531	0.948

* Cond.: Numbers of θ simulation conditions

* G: General factor distribution; S1: First specific factor distributions; S2: Second specific distributions

* (0): Standard normal distribution; (+): Positively skewed distributions; (-): Negatively skewed distributions

* Corr.: Correlation between specific factor distributions

to set up the mean and variance of the θ distribution in order to estimate unique θ parameters (Lord, 1980; Reckase, 2009). The most frequently used method is to appoint a mean of 0 and a variance of 1. The IRTPRO software used for this research sets the mean and variance of the θ distribution to 0 and 1, respectively, as default values, and the results showed the average discrepancies between the generated parameters and 0. Therefore, the mean biases close to 0, -0.3, or 0.3 indicate that the means of the estimated θ parameters were at '0'. This result shows that in order to evaluate the mean of mean bias of the θ distribution, some alternative method for giving the mean value needs to be utilized instead of using the fixed value of 0 for the mean of the distribution. The θ estimates were also consistently centered at '0' regardless of the item condition.

b. Variances of Mean Biases

The condition of the latent trait distributions with the most important effect on the variance of the mean biases was the correlation between specific factors. The variances of θ parameter mean biases are shown in Table 4-9. Different from the results for the means of mean biases, the variance results showed a specific pattern depending on the correlation between the specific factors. The amount of variance in the mean bias increased under conditions 2, 4, 6, and 8 with a high correlation between the specific factors (correlation=0.8), compared to the amount of variance in mean bias under conditions with a low correlation between the specific factors. While the general, first specific and second specific factors all had a large amount of variance in mean bias with the high correlation, the specific factor distributions showed more variance in mean bias than the general factor distributions across all item conditions. In order to investigate information about estimation precision, as a first insight, the correlations between the generated θ parameters and the estimated θ parameters were calculated.

c. Correlations between Generated and Estimated Parameter Distributions

Tables 4-10, 4-11 and 4-12 show the results of the mean correlations between the generated and estimated θ parameters, and more detailed information is provided in Appendix D. Under the various θ and item conditions, no noticeable patterns related to the correlation between the generated and estimated general factor distribution were found. The correlations between the generated and estimated distributions for the general factor are shown in Table 4-10. The estimated general factor scores showed constant high mean correlations regardless of the θ and item conditions, although the correlations were slightly lower when the level of correlations between the specific factors was high. Most of the mean correlations were greater than .77 under conditions 1, 3, 5, and 7, with a low correlation of 0.2 between specific factors, and the mean correlations were greater than .70 under conditions 2, 4, 6, and 8 with a high correlation of 0.8 between the specific factors.

The first and second specific factors showed the estimation precision to be sensitive to the level of correlation between the specific factors. Under the low correlation of 0.2 between the specific factors, the mean correlations of the generated and estimated first specific factors, shown in Table 4-11, and of the generated and estimated second specific factors, shown in Table 4-12, were over 0.7, although those correlations were slightly lower than the correlations between the generated and estimated parameters for the general factor. Whereas the level of correlation between the specific factors was only slightly influential on the observed correlation between the generated and estimated general factor θ parameters, the mean correlations between the generated and estimated parameters for the first and second specific factors were below 0.5 under conditions 2, 4, 6, and 8 with the high correlation between the specific factors.

The correlation between the generated and estimated parameters of the specific factors also showed noticeable patterns according to item condition (see Tables 4-11 and 4-12). While the mean correlations between the generated and estimated parameters of the specific factors did not show a distinguishable difference depending on the level of item discrimination parameters (mean discrimination parameters of 1.3 vs. 1.8), they showed a significant pattern depending on the level of the item difficulty parameters (difficulty parameters of 0.5 and -0.5) especially under conditions 4 and 8, which had specific factors with a high correlation and distribution skewed in opposite directions. The first specific factors had lower correlations between the generated and estimated parameters under conditions 4 and 8 when the item difficulty parameters had a mean of -0.5, whereas the second specific factors had lower correlations under conditions 4 and 8 when the item difficulty parameters had a mean of 0.5. This result implies that the effect of the correlation between the specific factor distributions on the correlation between the generated and estimated parameters for the specific factors is related to not only the direction of skewedness of the distributions but also to the item parameter conditions.

d. Kolmogorov-Smirnov Test (KS test)

In order to compare the generated and estimated θ distributions, the KS test was utilized with the entire θ parameter set, and with specific ranges of the θ parameters. Tables 4-13 shows summary results of the KS test; complete results of the KS tests are included in Appendix E. Every simulation condition was replicated fifty times, and among fifty replications the values in the tables show the numbers of frequencies that were statistically significant under the significance level of 0.05. For example, in Table 4-13, under Condition 1 with the mean of discrimination parameters equal to 1.3 and mean of difficulty parameters equal to -0.5, 18 of the estimated θ distributions among fifty replications were shown to be significantly different from

Table 4-10. Mean of the Correlations of the General Factors

Condition		1	2	3	4	5	6	7	8
θ	G	(0)	(0)	(0)	(0)	(+)	(+)	(+)	(+)
	S1	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)
	S2	(+)	(+)	(-)	(-)	(+)	(+)	(-)	(-)
	Correlation	0.2	0.8	0.2	0.8	0.2	0.8	0.2	0.8
General	(1.3, -0.5)	0.789	0.720	0.775	0.708	0.780	0.711	0.772	0.703
	(1.8, -0.5)	0.767	0.702	0.777	0.709	0.771	0.707	0.782	0.712
	(1.3, 0.5)	0.794	0.724	0.779	0.707	0.782	0.713	0.766	0.699
	(1.8, 0.5)	0.768	0.700	0.779	0.708	0.771	0.704	0.784	0.713

* Condition: Numbers of θ simulation conditions

* G: General factor distribution; S1: First specific factor distributions; S2: Second specific distributions

* (0): Standard normal distribution; (+): Positively skewed distributions; (-): Negatively skewed distributions

* (1.3, -0.5): Discrimination parameters with mean of 1.3 and Difficulty parameters with mean of -0.5

* Correlation: Correlation between specific factor distributions

the generated θ distribution by the p-value for the KS test statistic being less than the significance level of 0.05.

Most of the ten specific categories of the estimated general factor distributions were not significantly different from the generated distributions when they were generated from a standard

Table 4-11. Correlation Means of the First Specific Factors

Condition		1	2	3	4	5	6	7	8
θ	G	(0)	(0)	(0)	(0)	(+)	(+)	(+)	(+)
	S1	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)
	S2	(+)	(+)	(-)	(-)	(+)	(+)	(-)	(-)
	Corr.	0.2	0.8	0.2	0.8	0.2	0.8	0.2	0.8
S1 θ	(1.3, -0.5)	0.726	0.431	0.719	0.386	0.740	0.442	0.728	0.403
	(1.8, -0.5)	0.722	0.434	0.708	0.378	0.730	0.443	0.722	0.395
	(1.3, 0.5)	0.739	0.448	0.747	0.457	0.730	0.465	0.741	0.463
	(1.8, 0.5)	0.732	0.457	0.744	0.465	0.728	0.459	0.736	0.467

* Cond.: Numbers of θ simulation conditions

* G: General factor distribution; S1: First specific factor distributions; S2: Second specific distributions

* (0): Standard normal distribution; (+): Positively skewed distributions; (-): Negatively skewed distributions

* (1.3, -0.5): Discrimination parameters with mean of 1.3 and Difficulty parameters with mean of -0.5

* Corr.: Correlation between specific factor distributions

normal distribution under the conditions 1, 2, 3, and 4, whereas the KS tests on the entire set of parameters more often showed significant differences between the generated and estimated distributions. For example, under Condition 2 with the mean discrimination parameter equal to

Table 4-12. Correlation Means of the Second Specific Factors

	Cond.	1	2	3	4	5	6	7	8
θ	G	(0)	(0)	(0)	(0)	(+)	(+)	(+)	(+)
	S1	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)
	S2	(+)	(+)	(-)	(-)	(+)	(+)	(-)	(-)
	Corr.	0.2	0.8	0.2	0.8	0.2	0.8	0.2	0.8
S2 θ	(1.3, -0.5)	0.729	0.425	0.741	0.448	0.729	0.433	0.743	0.441
	(1.8, -0.5)	0.718	0.423	0.735	0.451	0.726	0.431	0.733	0.442
	(1.3, 0.5)	0.734	0.447	0.718	0.374	0.732	0.449	0.705	0.355
	(1.8, 0.5)	0.730	0.447	0.701	0.359	0.721	0.443	0.692	0.341

* Cond.: Numbers of θ simulation conditions

* G: General factor distribution; S1: First specific factor distributions; S2: Second specific distributions

* (0): Standard normal distribution; (+): Positively skewed distributions; (-): Negatively skewed distributions

* (1.3, -0.5): Discrimination parameters with mean of 1.3 and Difficulty parameters with mean of -0.5

* Corr.: Correlation between specific factor distributions

1.3 and mean difficulty parameter equal to -0.5 in Table 4-13, all of the replications were significant when the entire data set was tested, but few estimated parameter distribution replications (between one and three) were significantly different from the generated true parameter distribution when KS tests were conducted on ten specific categories of the θ values.

When the correlation between the specific factor distributions was high, the frequencies of significant test results for differences between the estimated and generated general factor distributions increased. Compared to conditions 1 and 3, under conditions 2 and 4 significant results under the KS test were found much more frequently. For example, in Table 4-13, with the mean discrimination parameter equal to 1.3 and mean difficulty parameter equal to -0.5, KS test results from all of the fifty replications of the entire data sets showed significant differences between the generated and estimated parameters under the high level correlation between specific factors (Condition 2 and 4) whereas only eighteen or twenty two replications are significant under lower level correlation.

All of the specific factor distributions were positively or negatively skewed, and the results of the KS test showed that the estimated θ distributions were significantly different from the generated θ distributions. According to Stapleton (2008), the KS test is powerful when the tested distributions are away from normality, as long as the sample size is sufficient. That means that when the sample size increases, the sensitivity of the KS test becomes stronger. The tests on the entire parameter distributions that included 2,000 values could have been more sensitive than the tests on the specific categories, which included 200 parameter values. For example, Table 4-13 shows that the KS tests for the entire data set were more frequently significant than the tests within the θ sub-categories.

Table 4-13. Frequency of Significant Differences between Distributions of Generating and Estimated General Factor θ Parameters

Item	Discrimination	Mean:1.8 / SD: 0.15							
	Difficulty	Mean:-0.5 / SD: 0.4				Mean:0.5 / SD: 0.4			
θ	Cond.	1	2	3	4	1	2	3	4
	G	(0)	(0)	(0)	(0)	(+)	(+)	(+)	(+)
	S1	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)
	S2	(+)	(+)	(-)	(-)	(+)	(+)	(-)	(-)
	Correlation	0.2	0.8	0.2	0.8	0.2	0.8	0.2	0.8
1 to 2000		18	50	22	50	34	50	21	49
1 to 200		0	1	0	2	0	5	0	4
201 to 400		0	3	0	3	0	4	0	3
401 to 600		0	3	0	4	0	2	0	4
601 to 800		0	2	0	0	0	4	0	0
801 to 1000		0	2	0	0	0	4	0	0
1001 to 1200		0	2	0	2	0	1	0	4
1201 to 1400		0	2	0	1	0	0	0	2
1401 to 1600		1	2	0	1	1	3	0	2
1601 to 1800		0	2	1	6	0	1	0	6
1801 to 2000		1	3	0	1	0	1	0	3

* Cond.: Numbers of simulation conditions

* G: General factor distribution; S1: First specific factor distributions; S2: Second specific distributions

* (0): Standard normal distribution; (+): Positively skewed distributions; (-): Negatively skewed distributions

* Corr.: Correlation between specific factor distributions

Table 4-13 (cont'd)

Item	Discrimination	Mean:1.8 / SD: 0.15							
	Difficulty	Mean:-0.5 / SD: 0.4				Mean:0.5 / SD: 0.4			
θ	Condition	1	2	3	4	1	2	3	4
	G	(0)	(0)	(0)	(0)	(+)	(+)	(+)	(+)
	S1	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)
	S2	(+)	(+)	(-)	(-)	(+)	(+)	(-)	(-)
	Correlation	0.2	0.8	0.2	0.8	0.2	0.8	0.2	0.8
1 to 2000		28	50	46	50	50	50	46	50
1 to 200		0	1	1	4	1	1	0	5
201 to 400		1	1	0	4	0	3	0	5
401 to 600		2	3	1	6	0	2	1	6
601 to 800		0	2	1	2	0	3	1	0
801 to 1000		0	2	1	2	0	3	1	0
1001 to 1200		0	1	0	5	0	1	0	3
1201 to 1400		0	3	0	1	0	2	0	1
1401 to 1600		1	2	0	1	1	4	1	2
1601 to 1800		1	2	1	6	1	1	0	6
1801 to 2000		0	3	0	0	0	3	0	5

* Cond.: Numbers of θ simulation conditions

* G: General factor distribution; S1: First specific factor distributions; S2: Second specific distributions

* (0): Standard normal distribution; (+): Positively skewed distributions; (-): Negatively skewed distributions

* Corr.: Correlation between specific factor distributions

5. Discussion

5.1 Summary of the Results

Item parameter estimation was affected by the degree of skewedness of general factor, the directions of skewedness of the specific factors, and the correlation between specific factors. These influential conditions of the latent trait distributions had different effects on item parameter estimation depending on the type of item parameter. First, the degree of skewedness of the general factor was influential in the estimation of the d parameters. Second, the direction of skewedness of the specific factor distributions was also influential in d parameter estimation. The skewedness direction affected both the mean and variance of the d parameter mean biases, and the effect on estimation increased when the general factor distribution was also skewed. Third, the correlation between the specific factors had a noticeable effect on the estimation of discrimination parameters. While estimation of discrimination parameters related to both the general and specific factors was affected by the size of the correlation between the specific factors, the discrimination parameters of the specific factors exhibited much more variance in their mean biases as a result than the discrimination parameters corresponding to the general factor. Lastly, generally the item discrimination parameters related to the general factor were overestimated, whereas the discrimination parameters related to the specific factors and d parameters were underestimated.

The estimated θ distributions had means of 0, and so the mean biases of the θ distributions had values close to 0, -0.3, and 0.3, depending on the direction of the generated distribution. Based on the variances of the mean biases and correlations between generated and estimated parameters, the most significant condition of the latent trait distribution in θ parameter estimation was the correlation between the specific factors. The amount of variance in the mean

bias increased under conditions with a high correlation of 0.8 between the specific factors. While all three factors, general, first specific and second specific, had large amounts of variance in mean bias with the high correlation, the specific factor distributions showed much more variance than the general factor distributions across the item conditions.

Whereas only a slightly noticeable pattern was found related to the correlation between the generated and estimated distributions for the general factor, the correlations between the generating and estimated distributions for the first and second specific factors were markedly lower when the correlation between the specific factors was high (0.8) than when it was low (0.2). Also the effect of the correlation between the specific factors depended on the item condition, and this result implies that the effect of the correlation between the specific distributions is related to not only the direction of skewedness of the distributions but also to the item parameter conditions.

By the Kolmogorov-Smirnov test, most of the ten specific categories of the estimated general factor distributions were not found to be significantly different from the generated distributions when the parameters were generated from a standard normal distribution. When the correlation between the specific factor distributions was high, the frequencies of significant test results for the general factor distribution increased. All of the specific factor distributions were positively or negatively skewed distributions, and the results of the KS test showed that the estimated θ distributions were significantly different from the generated θ distributions.

5.2 Implications

The use of measurements based on the concepts of multi-dimensional and non-normal distributions have been increasing in various fields. Latent trait models have been developed in

order to represent these complicated measurement properties. Researchers have studied appropriate estimation methods for each model, and the recommended methods have been evaluated in empirical situations. As an extension of these studies, this research examines the estimation performance of a bifactor model under various distributional conditions of the general and specific factors.

In many cases, the distributions of latent traits represent particular participant characteristics are non-normal. For example, it is not unusual to find that satisfaction measurements from a program evaluation or interaction frequencies in a social networking analysis have a skewed distribution with a long tail or with high kurtosis. When measurement models are used to estimate the parameters from data that do not follow a normal distribution, the normal distribution assumption of the estimation method may be violated. Therefore, new models and estimation methods should be developed in order to solve these problems: how the estimation of the model can be made robust when the normal distribution assumption is violated, or how the empirical data distribution can be substituted for a normal distribution in the estimation procedure. Woods and Thissen (2006) introduced Ramsay-Curve IRT, which is a non-parametric estimation procedure for the IRT latent distribution, and showed the capability of the method with normal and non-normal latent distributions. Also a complex model to allow correlations between the latent trait factors has been studied (Fujimoto, 2014; Cai 2010). For these newly-developed methods, it is necessary to evaluate their capability in different empirical situations to determine their limitations and produce further developments. This research evaluated the estimation quality of the bifactor model and the results showed how conditions of the item and θ parameter distributions affect item and θ parameter estimation under particular

estimation assumptions. As previous research has done, this research is expected to provide information about estimation performance and guidelines for future research.

One of the most important conditions studied in this research was the non-normality of the distribution of the latent traits being measured. Varying the amount and direction of skewedness of the distributions and the correlation between the specific factors, the results showed that in the analysis of data generated from skewed θ distributions, both the item and θ conditions influenced the quality of estimation. Also, the conditions had different effects depending on the type of item parameter estimated. The results from this research showing the effect of skewed latent trait distributions are consistent with the results of previous studies. Sass et al. (2008) demonstrated the effect of skewed distributions on estimating the θ distribution and item parameters using a unidimensional latent trait model. In that study, difficulty parameter estimates were particularly affected by the presence of a skewed latent trait distribution. Similarly, in my research the amount and direction of skewedness of the latent trait distributions had a significant influence on the mean and variance of d parameters' mean biases, which relates to the estimation of difficulty parameters.

The most significant condition of the latent trait distributions for θ estimation was the correlation between the specific factors. The correlation of the specific factors had a remarkable impact on θ estimation not only by itself but also in conjunction with particular item parameter conditions. This result shows that the combination of the item and θ conditions and the distributional assumptions should be considered simultaneously when the model and estimation method are evaluated.

The skewed distributions were transformed from normal distributions via the Copula method, and Kolmogorov-Smirnov tests were used to evaluate the distributional differences between the generated and estimated θ parameter distributions. My application of those methods has suggested some implications for future research. The Copula method requires identification of a transformation function, and in this research two polynomial functions were used for transformation to produce negatively and positively skewed latent trait distributions. Even though the R^2 s of the functions were values very close to 1 (.999), it should be noted that the extreme values were particularly sensitive to the polynomial transformation function selected. Also, Kolmogorov-Smirnov tests showed the differences between the generated and estimated θ parameters in specific θ ranges; however, this method had very high power to detect differences when entire θ distributions were compared. . Especially for the skewed distributions, all cases in each category were significant, using a significance level of 0.05. That shows that the skewed distributional condition tended to have significant differences between the generated and estimated parameter distributions, however, it could not provide specific information and details for each θ range. Therefore, more sophisticated and alternative methods are required for the transformation and the evaluation procedures.

In an effort to measure the structure of complex constructs, multidimensional latent trait models have been developed. The bifactor model is one of those multidimensional models, and is connected mathematically to other major classes of multidimensional measurement models. This research evaluates the bifactor model to determine how well it model works in various empirical contexts. While the distributions of latent traits are often assumed to be normal, the distributions observed in empirical data are not always normal. Also, despite the advantages of the bifactor model, it restricts the latent traits to be orthogonal.

The results from this research provided information about the estimation properties of bifactor models under conditions when their distributional and relational assumptions are not met. Also, the influence of item parameters was shown. Based on this information, the results can be applied to analyses using models of multidimensional latent traits. The study of the effect of the latent trait distribution on parameter estimation is also significant in terms of providing information about measurement error for data analysis. With the increasing number of studies and practical need for multidimensional structures of latent traits, this research is expected to provide useful guidelines for investigating appropriate multidimensional models.

APPENDICES

Appendix A

Table A-1. Parameter Estimates of Quadratic Regression Function for Positively Skewed Distribution with 2,000 examinees

	R Square	Constant	b1	b2	b3
Mean	0.9980	-0.4508	0.9770	0.1461	-
Var	0.0000	0.0006	0.0002	0.0001	-
Min	0.9960	-0.4948	0.9418	0.1282	-
Max	0.9993	-0.3915	1.0128	0.1727	-

*b1, b2, and b3: regression coefficients of linear, quadratic, and cubic terms

Table A-2. Parameter Estimates of Cubic Regression Function for Positively Skewed Distribution with 2,000 examinees

	R Square	Constant	b1	b2	b3
Mean	0.9990	-0.4508	1.0167	0.1461	-0.0136
Var	0.0000	0.0006	0.0003	0.0001	0.0000
Min	0.9978	-0.4948	0.9772	0.1282	-0.0236
Max	0.9997	-0.3915	1.0652	0.1727	-0.0060

*b1, b2, and b3: regression coefficients of linear, quadratic, and cubic terms

Table A-3. Parameter Estimates of Quadratic Regression Function for Positively Skewed Distribution with 10,000 examinees

	R Square	Constant	b1	b2	b3
Mean	0.9986	-0.4434	0.9793	0.1454	-
Var	0.0000	0.0001	0.0001	0.0000	-
Min	0.9978	-0.4701	0.9676	0.1337	-
Max	0.9993	-0.4261	0.9970	0.1557	-

*b1, b2, and b3: regression coefficients of linear, quadratic, and cubic terms

Table A-4. Parameter Estimates of Cubic Regression Function for Positively Skewed Distribution with 10,000 examinees

	R Square	Constant	b1	b2	b3
Mean	0.9995	-0.4434	1.0166	0.1454	-0.0125
Var	0.0000	0.0001	0.0001	0.0000	0.0000
Min	0.9993	-0.4701	0.9989	0.1337	-0.0175
Max	0.9998	-0.4261	1.0366	0.1557	-0.0086

*b1, b2, and b3: regression coefficients of linear, quadratic, and cubic terms

Table A-5. Parameter Estimates of Quadratic Regression Function for Negatively Skewed Distribution with 2,000 examinees

	R Square	Constant	b1	b2	b3
Mean	0.9984	0.4516	0.9801	-0.1500	-
Var	0.0000	0.0006	0.0002	0.0001	-
Min	0.9964	0.3921	0.9453	-0.1785	-
Max	0.9995	0.4950	1.0173	-0.1308	-

*b1, b2, and b3: regression coefficients of linear, quadratic, and cubic terms

Table A-6. Parameter Estimates of Cubic Regression Function for Negatively Skewed Distribution with 2,000 examinees

	R Square	Constant	b1	b2	b3
Mean	0.9991	0.4516	1.0135	-0.1500	-0.0115
Var	0.0000	0.0006	0.0003	0.0001	0.0000
Min	0.9981	0.3921	0.9720	-0.1785	-0.0225
Max	0.9997	0.4950	1.0625	-0.1308	-0.0023

*b1, b2, and b3: regression coefficients of linear, quadratic, and cubic terms

Table A-7. Parameter Estimates of Quadratic Regression Function for Negatively Skewed Distribution with 10,000 examinees

	R Square	Constant	b1	b2	b3
Mean	0.9988	0.4441	0.9798	-0.1462	-
Var	0.0000	0.0001	0.0001	0.0000	-
Min	0.9980	0.4223	0.9646	-0.1566	-
Max	0.9995	0.4705	0.9977	-0.1351	-

*b1, b2, and b3: regression coefficients of linear, quadratic, and cubic terms

Table A-8. Parameter Estimates of Cubic Regression Function for Negatively Skewed Distribution with 10,000 examinees

	R Square	Constant	b1	b2	b3
Mean	0.9996	0.4441	1.0151	-0.1462	-0.0119
Var	0.0000	0.0001	0.0001	0.0000	0.0000
Min	0.9993	0.4223	0.9976	-0.1566	-0.0164
Max	0.9998	0.4705	1.0360	-0.1351	-0.0079

*b1, b2, and b3: regression coefficients of linear, quadratic, and cubic terms

Appendix B

Table B-1. Mean and Variance of Item Parameter Biases under Disc of 1.3 and Diff of -0.5

Cond.		1	2	3	4	5	6	7	8
θ	G	(0)	(0)	(0)	(0)	(+)	(+)	(+)	(+)
	S1	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)
	S2	(+)	(+)	(-)	(-)	(+)	(+)	(-)	(-)
	Corr.	0.2	0.8	0.2	0.8	0.2	0.8	0.2	0.8
Mean									
G Disc.	min	0.0518	0.3207	0.0576	0.3393	0.0374	0.3084	0.0478	0.3176
	mean	0.1214	0.3910	0.1372	0.4072	0.1178	0.3824	0.1098	0.3952
	max	0.2168	0.4773	0.2348	0.5125	0.1995	0.4422	0.1665	0.4613
S1 Disc.	min	-0.5501	-0.8914	-0.5463	-0.9447	-0.3600	-0.8853	-0.5512	-0.9216
	mean	-0.2066	-0.7139	-0.1895	-0.7500	-0.1442	-0.7150	-0.1697	-0.7357
	max	-0.0482	-0.5252	0.0691	-0.6273	0.0698	-0.5660	0.0899	-0.5897
S2 Disc.	min	-0.3516	-0.9029	-0.4893	-0.8175	-0.4804	-0.8331	-0.7114	-0.8767
	mean	-0.1661	-0.7150	-0.1157	-0.6894	-0.1879	-0.6814	-0.1456	-0.7172
	max	0.0331	-0.5773	0.1494	-0.5508	0.0080	-0.5188	0.0821	-0.5423
D	min	-0.5161	-0.5358	-0.0651	-0.1145	-0.9444	-0.9847	-0.4985	-0.5359
	mean	-0.4329	-0.4310	0.0054	0.0234	-0.8401	-0.8448	-0.4201	-0.4264
	max	-0.3546	-0.3103	0.1018	0.1433	-0.7695	-0.7345	-0.3000	-0.2852
Variance									
G Disc.	min	0.0067	0.0107	0.0070	0.0113	0.0071	0.0115	0.0079	0.0147
	mean	0.0188	0.0177	0.0247	0.0212	0.0196	0.0193	0.0245	0.0221
	max	0.0727	0.0296	0.0615	0.0336	0.0686	0.0294	0.1181	0.0313
S1 Disc.	min	0.0060	0.0075	0.0054	0.0125	0.0049	0.0067	0.0049	0.0105
	mean	0.0126	0.0203	0.0129	0.0259	0.0108	0.0203	0.0114	0.0213
	max	0.0369	0.0468	0.0274	0.0457	0.0196	0.0401	0.0249	0.0525
S2 Disc.	min	0.0042	0.0073	0.0056	0.0077	0.0054	0.0094	0.0062	0.0080
	mean	0.0117	0.0195	0.0133	0.0199	0.0122	0.0186	0.0128	0.0198
	max	0.0256	0.0435	0.0343	0.0342	0.0271	0.0339	0.0345	0.0365
D	min	0.0039	0.0043	0.1592	0.1787	0.0059	0.0056	0.1528	0.1778
	mean	0.0077	0.0075	0.2043	0.2096	0.0092	0.0089	0.1964	0.2092
	max	0.0106	0.0115	0.2470	0.2478	0.0147	0.0137	0.2662	0.2436

*Disc.: Discrimination item parameter / Diff.: Difficulty parameter

* Cond.: Numbers of θ simulation conditions

* G: General factor; S1: First specific factor; S2: Second specific

* (0): Standard normal distribution; (+): Positively skewed distributions; (-): Negatively skewed distributions

* Corr.: Correlation between specific factor distributions

Table B-2. Mean and Variance of Item Parameter Biases under Disc of 1.3 and Diff of 0.5

Cond.		1	2	3	4	5	6	7	8
θ	G	(0)	(0)	(0)	(0)	(+)	(+)	(+)	(+)
	S1	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)
	S2	(+)	(+)	(-)	(-)	(+)	(+)	(-)	(-)
	Corr.	0.2	0.8	0.2	0.8	0.2	0.8	0.2	0.8
Mean									
G Disc.	min	0.0464	0.3541	0.0555	0.3390	0.1280	0.3927	0.0635	0.3637
	mean	0.1432	0.4437	0.1344	0.4133	0.2057	0.4875	0.1711	0.4491
	max	0.2692	0.5308	0.2398	0.5052	0.3134	0.5833	0.2394	0.5404
S1 Disc.	min	-0.5176	-0.8712	-0.4041	-0.8545	-0.6482	-0.8471	-0.3627	-0.8388
	mean	-0.0974	-0.6775	-0.1226	-0.7126	-0.1087	-0.6289	-0.0878	-0.6714
	max	0.1969	-0.4658	0.0996	-0.4984	0.2257	-0.4316	0.1478	-0.4996
S2 Disc.	min	-0.5220	-0.9141	-0.4374	-0.9474	-0.3851	-0.8688	-0.7276	-1.0232
	mean	-0.1249	-0.6341	-0.1708	-0.7281	-0.0564	-0.6590	-0.2103	-0.7537
	max	0.1200	-0.4482	0.0475	-0.5688	0.1772	-0.4432	0.0244	-0.6011
D	min	-0.5524	-0.5218	-0.0893	-0.0972	-0.9769	-0.9906	-0.5214	-0.5450
	mean	-0.4432	-0.4382	-0.0034	0.0077	-0.8721	-0.8598	-0.4282	-0.4304
	max	-0.3359	-0.3474	0.0783	0.2179	-0.7575	-0.7560	-0.3112	-0.2353
Variance									
G Disc.	min	0.0084	0.0152	0.0079	0.0109	0.0098	0.0116	0.0088	0.0132
	mean	0.0306	0.0236	0.0218	0.0223	0.0295	0.0213	0.0217	0.0228
	max	0.0985	0.0353	0.1008	0.0341	0.1263	0.0300	0.0619	0.0398
S1 Disc.	min	0.0065	0.0120	0.0050	0.0057	0.0063	0.0077	0.0057	0.0101
	mean	0.0139	0.0224	0.0142	0.0220	0.0172	0.0200	0.0143	0.0223
	max	0.0350	0.0628	0.0336	0.0604	0.0761	0.0388	0.0310	0.0486
S2 Disc.	min	0.0048	0.0081	0.0059	0.0127	0.0059	0.0082	0.0059	0.0115
	mean	0.0146	0.0211	0.0117	0.0246	0.0143	0.0230	0.0143	0.0304
	max	0.0315	0.0430	0.0196	0.0733	0.0309	0.0631	0.0644	0.0617
D	min	0.0065	0.0065	0.1561	0.1804	0.0086	0.0096	0.1767	0.1825
	mean	0.0107	0.0108	0.2068	0.2130	0.0155	0.0147	0.2092	0.2201
	max	0.0152	0.0187	0.2699	0.2423	0.0234	0.0304	0.2582	0.2694

*Disc.: Discrimination item parameter / Diff.: Difficulty parameter

* Cond.: Numbers of θ simulation conditions

* G: General factor; S1: First specific factor; S2: Second specific

* (0): Standard normal distribution; (+): Positively skewed distributions; (-): Negatively skewed distributions

* Corr.: Correlation between specific factor distributions

Table B-3. Mean and Variance of Item Parameter Biases under Disc of 1.8 and Diff of -0.5

Cond.		1	2	3	4	5	6	7	8
θ	G	(0)	(0)	(0)	(0)	(+)	(+)	(+)	(+)
	S1	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)
	S2	(+)	(+)	(-)	(-)	(+)	(+)	(-)	(-)
	Corr.	0.2	0.8	0.2	0.8	0.2	0.8	0.2	0.8
Mean									
G Disc.	min	0.0679	0.3911	0.1290	0.4211	-0.0440	0.4045	-0.0789	0.4002
	mean	0.2163	0.4935	0.2820	0.5378	0.2016	0.5022	0.1850	0.5080
	max	0.3628	0.5746	0.4263	0.6272	0.3604	0.5831	0.3220	0.6128
S1 Disc.	min	-0.4473	-1.0972	-0.3875	-1.1511	-0.6284	-1.1186	-0.5386	-1.1185
	mean	-0.3077	-0.9783	-0.2885	-1.0068	-0.2842	-0.9541	-0.2349	-0.9995
	max	-0.1196	-0.8678	-0.0026	-0.8202	0.0696	-0.7873	0.2381	-0.8592
S2 Disc.	min	-0.5096	-1.1444	-0.3255	-1.0831	-0.6628	-1.1417	-0.6950	-1.1150
	mean	-0.3206	-0.9786	-0.2154	-0.9570	-0.2740	-0.9550	-0.2901	-0.9927
	max	-0.1832	-0.8407	-0.1340	-0.8116	0.0605	-0.7702	0.0484	-0.9018
D	min	-0.7969	-0.7648	-0.1830	-0.1192	-1.4206	-1.3482	-0.8585	-0.8027
	mean	-0.6106	-0.6111	-0.0009	0.0544	-1.2043	-1.1949	-0.6017	-0.5864
	max	-0.4454	-0.4274	0.1481	0.2478	-1.0508	-1.0203	-0.4262	-0.3206
Variance									
G Disc.	min	0.0114	0.0136	0.0139	0.0203	0.0096	0.0136	0.0142	0.0199
	mean	0.0201	0.0225	0.0259	0.0336	0.0283	0.0260	0.0451	0.0300
	max	0.0640	0.0356	0.1055	0.0496	0.1399	0.0418	0.2533	0.0434
S1 Disc.	min	0.0058	0.0109	0.0074	0.0138	0.0077	0.0104	0.0067	0.0122
	mean	0.0168	0.0228	0.0163	0.0270	0.0154	0.0204	0.0157	0.0250
	max	0.0373	0.0416	0.0286	0.0469	0.0336	0.0341	0.0289	0.0547
S2 Disc.	min	0.0068	0.0098	0.0074	0.0132	0.0052	0.0114	0.0069	0.0106
	mean	0.0142	0.0227	0.0183	0.0246	0.0146	0.0215	0.0180	0.0216
	max	0.0263	0.0383	0.0339	0.0429	0.0340	0.0395	0.0751	0.0403
D	min	0.0072	0.0065	0.3004	0.3682	0.0076	0.0075	0.3244	0.3592
	mean	0.0112	0.0110	0.4151	0.4230	0.0119	0.0107	0.3932	0.4115
	max	0.0158	0.0181	0.4952	0.4753	0.0206	0.0150	0.5369	0.5031

*Disc.: Discrimination item parameter / Diff.: Difficulty parameter

* Cond.: Numbers of θ simulation conditions

* G: General factor; S1: First specific factor; S2: Second specific

* (0): Standard normal distribution; (+): Positively skewed distributions; (-): Negatively skewed distributions

* Corr.: Correlation between specific factor distributions

Table B-4. Mean and Variance of Item Parameter Biases under Disc of 1.8 and Diff of 0.5

Cond.		1	2	3	4	5	6	7	8
θ	G	(0)	(0)	(0)	(0)	(+)	(+)	(+)	(+)
	S1	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)
	S2	(+)	(+)	(-)	(-)	(+)	(+)	(-)	(-)
	Corr.	0.2	0.8	0.2	0.8	0.2	0.8	0.2	0.8
Mean									
G Disc.	Min	0.0205	0.5126	0.0211	0.4275	0.2486	0.5698	0.1853	0.5014
	Mean	0.3053	0.6149	0.2661	0.5419	0.4206	0.7092	0.3344	0.6165
	Max	0.4789	0.7022	0.3815	0.6671	0.5687	0.8086	0.4428	0.7829
S1 Disc.	Min	-0.4631	-1.0605	-0.3136	-1.0915	-0.2541	-1.0893	-0.3178	-1.0873
	Mean	-0.2075	-0.8910	-0.2135	-0.9612	-0.1446	-0.8536	-0.1768	-0.9212
	Max	-0.0386	-0.6841	0.0053	-0.8276	0.0221	-0.6149	-0.0555	-0.7876
S2 Disc.	Min	-0.3654	-1.0131	-0.4576	-1.1091	-0.2997	-1.0280	-0.4166	-1.1505
	Mean	-0.1884	-0.8777	-0.3010	-0.9951	-0.1591	-0.8628	-0.3035	-1.0090
	Max	0.1765	-0.7680	-0.2001	-0.8612	-0.0055	-0.6371	-0.1684	-0.9012
D	Min	-0.7658	-0.7621	-0.2331	-0.1928	-1.5176	-1.4091	-0.7523	-0.7552
	Mean	-0.6353	-0.6148	-0.0147	0.0086	-1.2776	-1.2486	-0.6227	-0.6196
	Max	-0.4744	-0.4390	0.2370	0.2027	-1.1127	-1.1132	-0.4510	-0.4308
Variance									
G Disc.	Min	0.0132	0.0217	0.0135	0.0193	0.0142	0.0212	0.0084	0.0235
	Mean	0.0306	0.0321	0.0253	0.0325	0.0270	0.0335	0.0363	0.0432
	Max	0.1981	0.0562	0.0770	0.0599	0.0599	0.0566	0.0779	0.0705
S1 Disc.	Min	0.0052	0.0088	0.0092	0.0113	0.0068	0.0112	0.0100	0.0148
	Mean	0.0162	0.0204	0.0177	0.0245	0.0187	0.0263	0.0189	0.0273
	Max	0.0361	0.0411	0.0328	0.0438	0.0300	0.0574	0.0370	0.0591
S2 Disc.	Min	0.0082	0.0117	0.0074	0.0154	0.0099	0.0119	0.0078	0.0138
	Mean	0.0174	0.0228	0.0157	0.0308	0.0195	0.0270	0.0170	0.0348
	Max	0.0314	0.0487	0.0282	0.0523	0.0300	0.0475	0.0304	0.0666
D	Min	0.0098	0.0094	0.3407	0.3794	0.0172	0.0137	0.3464	0.3523
	Mean	0.0183	0.0181	0.4104	0.4245	0.0283	0.0283	0.4371	0.4490
	Max	0.0358	0.0321	0.4890	0.4770	0.0520	0.0521	0.5564	0.5055

*Disc.: Discrimination item parameter / Diff.: Difficulty parameter

* Cond.: Numbers of θ simulation conditions

* G: General factor; S1: First specific factor; S2: Second specific

* (0): Standard normal distribution; (+): Positively skewed distributions; (-): Negatively skewed distributions

* Corr.: Correlation between specific factor distributions

Appendix C

Table C-1. Mean and Variance of θ Parameter Biases under Disc. of 1.3 and Diff. of -0.5

Cond.		1	2	3	4	5	6	7	8
θ	G	(0)	(0)	(0)	(0)	(+)	(+)	(+)	(+)
	S1	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)
	S2	(+)	(+)	(-)	(-)	(+)	(+)	(-)	(-)
	Corr.	0.2	0.8	0.2	0.8	0.2	0.8	0.2	0.8
Mean									
G	Min	-0.0488	-0.0625	-0.0352	-0.0793	0.2690	0.2224	0.2388	0.2542
	Mean	0.0028	0.0064	0.0021	-0.0096	0.3048	0.3084	0.3029	0.3078
	Max	0.0479	0.0818	0.0513	0.0568	0.3582	0.3697	0.3516	0.3638
S1	Min	0.2641	0.2586	0.2431	0.2510	0.2430	0.2538	0.2513	0.2301
	Mean	0.3011	0.2966	0.3042	0.2976	0.2989	0.3012	0.3046	0.3050
	Max	0.3671	0.3412	0.3503	0.3636	0.3568	0.3521	0.3598	0.3577
S2	Min	0.2681	0.2570	-0.3532	-0.3643	0.2483	0.2501	-0.3484	-0.3677
	Mean	0.3086	0.3003	-0.3051	-0.3079	0.3039	0.3047	-0.2950	-0.3014
	Max	0.3578	0.3479	-0.2684	-0.2527	0.3531	0.3538	-0.2346	-0.2418
Variance									
G	Min	0.3345	0.4924	0.3597	0.5294	0.3627	0.5197	0.3705	0.5306
	Mean	0.3827	0.5464	0.4077	0.5702	0.3990	0.5589	0.4129	0.5786
	Max	0.4680	0.5952	0.4981	0.6341	0.4601	0.6262	0.5272	0.6369
S1	Min	0.4169	0.7023	0.4116	0.7582	0.3874	0.7176	0.4160	0.7748
	Mean	0.4714	0.8506	0.4905	0.8886	0.4569	0.8322	0.4708	0.8738
	Max	0.6123	0.9790	0.6497	1.0351	0.5355	0.9342	0.5497	1.0342
S2	Min	0.4151	0.7733	0.4076	0.7341	0.4108	0.6922	0.3743	0.7218
	Mean	0.4688	0.8557	0.4546	0.8158	0.4698	0.8449	0.4537	0.8330
	Max	0.5209	0.9521	0.5657	0.9319	0.5523	0.9613	0.6878	0.9304

*Disc.: Discrimination item parameter / Diff.: Difficulty parameter

* Cond.: Numbers of θ simulation conditions

* G: General factor; S1: First specific factor; S2: Second specific

* (0): Standard normal distribution; (+): Positively skewed distributions; (-): Negatively skewed distributions

* Corr.: Correlation between specific factor distributions

Table C-2. Mean and Variance of θ Parameter Biases under Disc. of 1.3 and Diff. of 0.5

Cond.		1	2	3	4	5	6	7	8
θ	G	(0)	(0)	(0)	(0)	(+)	(+)	(+)	(+)
	S1	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)
	S2	(+)	(+)	(-)	(-)	(+)	(+)	(-)	(-)
	Corr.	0.2	0.8	0.2	0.8	0.2	0.8	0.2	0.8
Mean									
G	Min	-0.0579	-0.0538	-0.0442	-0.1161	0.2325	0.2271	0.2487	0.2191
	Mean	0.0007	0.0043	0.0000	-0.0104	0.3033	0.3034	0.3006	0.3023
	Max	0.0463	0.0719	0.0656	0.0462	0.3606	0.3632	0.3483	0.3538
S1	Min	0.2534	0.2585	0.2469	0.2508	0.2404	0.2539	0.2572	0.2299
	Mean	0.3006	0.2965	0.3032	0.2975	0.2993	0.3011	0.3027	0.3048
	Max	0.3661	0.3411	0.3538	0.3635	0.3506	0.3515	0.3445	0.3577
S2	Min	0.2548	0.2569	-0.3568	-0.3644	0.2453	0.2500	-0.3389	-0.3678
	Mean	0.3070	0.3002	-0.3060	-0.3080	0.3019	0.3046	-0.2956	-0.3015
	Max	0.3724	0.3483	-0.2682	-0.2528	0.3524	0.3534	-0.2487	-0.2417
Variance									
G	Min	0.3616	0.5200	0.3526	0.5297	0.3571	0.5115	0.3553	0.5122
	Mean	0.4194	0.5741	0.4030	0.5698	0.4123	0.5578	0.3949	0.5581
	Max	0.5173	0.6330	0.4736	0.6306	0.5163	0.6333	0.4665	0.6049
S1	Min	0.3798	0.7138	0.3569	0.7004	0.3959	0.6914	0.3853	0.7350
	Mean	0.4540	0.8229	0.4479	0.8076	0.4676	0.8034	0.4496	0.8016
	Max	0.5882	0.9599	0.5479	0.9231	0.7469	0.9582	0.5219	0.9080
S2	Min	0.3922	0.7491	0.4265	0.8308	0.4104	0.6833	0.4231	0.7808
	Mean	0.4626	0.8316	0.4897	0.9025	0.4658	0.8125	0.5092	0.9221
	Max	0.5590	0.9164	0.5512	0.9843	0.5341	0.9619	0.7431	1.0380

*Disc.: Discrimination item parameter / Diff.: Difficulty parameter

* Cond.: Numbers of θ simulation conditions

* G: General factor; S1: First specific factor; S2: Second specific

* (0): Standard normal distribution; (+): Positively skewed distributions; (-): Negatively skewed distributions

* Corr.: Correlation between specific factor distributions

Table C-3. Mean and Variance of θ Parameter Biases under Disc. of 1.8 and Diff. of -0.5

Cond.		1	2	3	4	5	6	7	8
θ	G	(0)	(0)	(0)	(0)	(+)	(+)	(+)	(+)
	S1	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)
	S2	(+)	(+)	(-)	(-)	(+)	(+)	(-)	(-)
	Corr.	0.2	0.8	0.2	0.8	0.2	0.8	0.2	0.8
Mean									
G	Min	-0.0752	-0.0751	-0.0515	-0.0815	0.2449	0.2489	0.2355	0.2266
	Mean	-0.0006	0.0064	0.0062	-0.0158	0.3080	0.3092	0.3055	0.3014
	Max	0.0770	0.0755	0.0793	0.0325	0.3767	0.3796	0.3938	0.3713
S1	Min	0.2484	0.2587	0.2542	0.2512	0.2558	0.2539	0.2558	0.2299
	Mean	0.3009	0.2967	0.3057	0.2977	0.3018	0.3012	0.3080	0.3050
	Max	0.4024	0.3414	0.3610	0.3636	0.3512	0.3510	0.3857	0.3576
S2	Min	0.2614	0.2571	-0.3461	-0.3644	0.2461	0.2499	-0.3475	-0.3677
	Mean	0.3053	0.3003	-0.3022	-0.3079	0.3040	0.3047	-0.2957	-0.3013
	Max	0.3521	0.3481	-0.2544	-0.2528	0.3740	0.3533	-0.2395	-0.2421
Variance									
G	Min	0.3391	0.4948	0.3606	0.5362	0.3365	0.5172	0.3635	0.5561
	Mean	0.3737	0.5422	0.3989	0.5765	0.3959	0.5593	0.4227	0.5911
	Max	0.4285	0.5954	0.4342	0.6204	0.4531	0.6102	0.5729	0.6315
S1	Min	0.4215	0.7403	0.4629	0.8165	0.4059	0.7850	0.4283	0.8126
	Mean	0.4791	0.8520	0.5061	0.9091	0.4699	0.8405	0.4826	0.8936
	Max	0.5247	0.9537	0.5536	1.0262	0.5151	0.9308	0.5409	0.9735
S2	Min	0.4347	0.7870	0.4223	0.7219	0.4195	0.7551	0.4111	0.6987
	Mean	0.4844	0.8619	0.4608	0.8119	0.4761	0.8497	0.4679	0.8347
	Max	0.5302	0.9652	0.5055	0.8846	0.5481	0.9195	0.6768	0.9165

*Disc.: Discrimination item parameter / Diff.: Difficulty parameter

* Cond.: Numbers of θ simulation conditions

* G: General factor; S1: First specific factor; S2: Second specific

* (0): Standard normal distribution; (+): Positively skewed distributions; (-): Negatively skewed distributions

* Corr.: Correlation between specific factor distributions

Table C-4. Mean and Variance of θ Parameter Biases under Disc. of 1.8 and Diff. of 0.5

Cond.		1	2	3	4	5	6	7	8
θ	G	(0)	(0)	(0)	(0)	(+)	(+)	(+)	(+)
	S1	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)
	S2	(+)	(+)	(-)	(-)	(+)	(+)	(-)	(-)
	Corr.	0.2	0.8	0.2	0.8	0.2	0.8	0.2	0.8
Mean									
G	Min	-0.0565	-0.0658	-0.0730	-0.0768	0.2499	0.2409	0.2534	0.2200
	Mean	0.0005	0.0022	0.0006	-0.0104	0.3108	0.3076	0.3010	0.3001
	Max	0.0640	0.0651	0.0742	0.0379	0.3715	0.3713	0.3475	0.3487
S1	Min	0.2505	0.2576	0.2435	0.2511	0.2358	0.2552	0.2502	0.2299
	Mean	0.2998	0.2963	0.3038	0.2973	0.3033	0.3014	0.3043	0.3048
	Max	0.3546	0.3408	0.3695	0.3636	0.3824	0.3526	0.3547	0.3576
S2	Min	0.2521	0.2559	-0.3656	-0.3646	0.2455	0.2499	-0.3638	-0.3678
	Mean	0.3078	0.3001	-0.3058	-0.3080	0.3052	0.3045	-0.2967	-0.3015
	Max	0.3620	0.3471	-0.2578	-0.2526	0.3549	0.3546	-0.2475	-0.2419
Variance									
G	Min	0.3821	0.5198	0.3686	0.5382	0.3768	0.5213	0.3580	0.4981
	Mean	0.4154	0.5780	0.3983	0.5753	0.4105	0.5599	0.3897	0.5585
	Max	0.5126	0.6286	0.4257	0.6249	0.4413	0.6044	0.4188	0.6138
S1	Min	0.4073	0.7006	0.4196	0.7288	0.4313	0.7346	0.4150	0.6650
	Mean	0.4614	0.8169	0.4509	0.8000	0.4715	0.8088	0.4554	0.7957
	Max	0.5468	0.9201	0.4836	0.8611	0.5162	0.8903	0.5117	0.8618
S2	Min	0.4155	0.7575	0.4742	0.8381	0.4219	0.7322	0.4887	0.8186
	Mean	0.4658	0.8296	0.5145	0.9271	0.4794	0.8197	0.5306	0.9475
	Max	0.4983	0.9384	0.5620	1.0190	0.5316	0.9070	0.5711	1.0542

*Disc.: Discrimination item parameter / Diff.: Difficulty parameter

* Cond.: Numbers of θ simulation conditions

* G: General factor; S1: First specific factor; S2: Second specific

* (0): Standard normal distribution; (+): Positively skewed distributions; (-): Negatively skewed distributions

* Corr.: Correlation between specific factor distributions

Appendix D

Table D-1. Correlations between Generated and Estimated Factors with Discrimination Parameters from mean of 1.3

	Cond.	1	2	3	4	5	6	7	8
θ	G	(0)	(0)	(0)	(0)	(+)	(+)	(+)	(+)
	S1	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)
	S2	(+)	(+)	(-)	(-)	(+)	(+)	(-)	(-)
	Corr.	0.2	0.8	0.2	0.8	0.2	0.8	0.2	0.8
Difficulty parameters with mean of -0.5 and standard deviation of 0.4									
G	Mean	0.7890	0.7200	0.7749	0.7079	0.7800	0.7106	0.7719	0.7033
	Var	0.0002	0.0002	0.0003	0.0002	0.0002	0.0001	0.0002	0.0002
	Min	0.7440	0.6867	0.7219	0.6737	0.7355	0.6736	0.7166	0.6698
	Max	0.8253	0.7504	0.8030	0.7347	0.8046	0.7364	0.7952	0.7329
S1	Mean	0.7260	0.4310	0.7186	0.3859	0.7396	0.4416	0.7276	0.4032
	Var	0.0007	0.0018	0.0007	0.0016	0.0004	0.0011	0.0003	0.0017
	Min	0.6280	0.3317	0.6346	0.2993	0.6904	0.3624	0.6905	0.2815
	Max	0.7670	0.5216	0.7649	0.5044	0.7768	0.5251	0.7674	0.4973
S2	Mean	0.7295	0.4248	0.7414	0.4482	0.7289	0.4329	0.7431	0.4414
	Var	0.0003	0.0011	0.0003	0.0014	0.0004	0.0011	0.0011	0.0013
	Min	0.6969	0.3178	0.6718	0.3558	0.6685	0.3690	0.5572	0.3504
	Max	0.7631	0.4932	0.7702	0.5161	0.7656	0.5059	0.7839	0.5397
Difficulty parameters with mean of 0.5 and standard deviation of 0.4									
G	Mean	0.7668	0.7024	0.7771	0.7085	0.7713	0.7066	0.7819	0.7119
	Var	0.0003	0.0002	0.0002	0.0002	0.0003	0.0002	0.0002	0.0002
	Min	0.7031	0.6677	0.7479	0.6711	0.7182	0.6664	0.7468	0.6867
	Max	0.8047	0.7352	0.8046	0.7312	0.7980	0.7357	0.8037	0.7407
S1	Mean	0.7387	0.4480	0.7466	0.4572	0.7296	0.4650	0.7410	0.4625
	Var	0.0007	0.0019	0.0004	0.0017	0.0023	0.0016	0.0003	0.0015
	Min	0.6492	0.2882	0.7044	0.3705	0.4561	0.3751	0.6983	0.3786
	Max	0.7742	0.5267	0.8015	0.5427	0.7702	0.5451	0.7850	0.5311
S2	Mean	0.7339	0.4470	0.7178	0.3736	0.7324	0.4487	0.7049	0.3549
	Var	0.0004	0.0010	0.0003	0.0015	0.0005	0.0020	0.0012	0.0024
	Min	0.6733	0.3722	0.6874	0.2322	0.6688	0.3681	0.5076	0.2216
	Max	0.7746	0.5015	0.7508	0.4258	0.7721	0.5467	0.7505	0.4712

* Cond.: Numbers of θ simulation conditions

* G: General factor; S1: First specific factor; S2: Second specific

* (0): Standard normal distribution; (+): Positively skewed distributions; (-): Negatively skewed distributions

* Corr.: Correlation between specific factor distributions

Table D-2. Correlations between Generated and Estimated Factors with Discrimination Parameters from mean of 1.8

Condition		1	2	3	4	5	6	7	8
θ	G	(0)	(0)	(0)	(0)	(+)	(+)	(+)	(+)
	S1	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)
	S2	(+)	(+)	(-)	(-)	(+)	(+)	(-)	(-)
	Corr.	0.2	0.8	0.2	0.8	0.2	0.8	0.2	0.8
Difficulty parameters with mean of -0.5 and standard deviation of 0.4									
G	Mean	0.7937	0.7243	0.7785	0.7072	0.7817	0.7125	0.7663	0.6992
	Var	0.0001	0.0002	0.0001	0.0001	0.0001	0.0002	0.0004	0.0001
	Min	0.7658	0.6891	0.7569	0.6724	0.7535	0.6867	0.6900	0.6771
	Max	0.8177	0.7471	0.7993	0.7244	0.8072	0.7399	0.7945	0.7217
S1	Mean	0.7215	0.4336	0.7083	0.3783	0.7304	0.4433	0.7219	0.3947
	Var	0.0002	0.0007	0.0002	0.0009	0.0002	0.0005	0.0002	0.0007
	Min	0.6959	0.3464	0.6830	0.2991	0.6950	0.3829	0.6914	0.3248
	Max	0.7442	0.4907	0.7341	0.4266	0.7731	0.4833	0.7556	0.4463
S2	Mean	0.7184	0.4226	0.7351	0.4508	0.7255	0.4314	0.7333	0.4418
	Var	0.0002	0.0007	0.0002	0.0008	0.0003	0.0007	0.0008	0.0007
	Min	0.6863	0.3561	0.7057	0.3838	0.6624	0.3703	0.5923	0.3486
	Max	0.7569	0.4847	0.7604	0.5123	0.7740	0.4996	0.7745	0.4911
Difficulty parameters with mean of 0.5 and standard deviation of 0.4									
G	Mean	0.7677	0.6997	0.7790	0.7075	0.7710	0.7041	0.7839	0.7129
	Var	0.0002	0.0002	0.0001	0.0002	0.0001	0.0002	0.0001	0.0002
	Min	0.7055	0.6628	0.7563	0.6799	0.7490	0.6830	0.7648	0.6805
	Max	0.7911	0.7315	0.8001	0.7348	0.7939	0.7398	0.8086	0.7482
S1	Mean	0.7323	0.4566	0.7439	0.4649	0.7276	0.4595	0.7362	0.4667
	Var	0.0002	0.0009	0.0001	0.0007	0.0001	0.0007	0.0001	0.0009
	Min	0.6942	0.3662	0.7172	0.4064	0.7069	0.4089	0.7027	0.4047
	Max	0.7568	0.5181	0.7617	0.5211	0.7516	0.5106	0.7577	0.5344
S2	Mean	0.7305	0.4471	0.7007	0.3591	0.7211	0.4433	0.6917	0.3414
	Var	0.0002	0.0006	0.0002	0.0012	0.0002	0.0012	0.0002	0.0010
	Min	0.7053	0.3885	0.6766	0.2989	0.6995	0.3356	0.6651	0.2687
	Max	0.7577	0.4987	0.7279	0.4371	0.7576	0.5178	0.7218	0.4105

* Condition: Numbers of θ simulation conditions

* G: General factor; S1: First specific factor; S2: Second specific

* (0): Standard normal distribution; (+): Positively skewed distributions; (-): Negatively skewed distributions

* Corr.: Correlation between specific factor distributions

Appendix E

Table E-1. Numbers of Frequencies Significant by KS Test under Condition 1

Discrimination	Mean: 1.3 / SD: 0.15						Mean: 1.8 / SD: 0.15					
Difficulty	Mean: -0.5			Mean: 0.5			Mean: -0.5			Mean: 0.5		
	SD: 0.4			SD: 0.4			SD: 0.4			SD: 0.4		
θ	G	S1	S2	G	S1	S2	G	S1	S2	G	S1	S2
1 to 2000	18	50	50	34	50	50	28	50	50	50	50	50
1 to 200	0	50	50	0	50	50	0	50	50	1	50	50
201 to 400	0	50	50	0	50	50	1	50	50	0	50	50
401 to 600	0	50	50	0	50	50	2	50	50	0	50	50
601 to 800	0	50	50	0	50	50	0	50	50	0	50	50
801 to 1000	0	50	50	0	50	50	0	50	50	0	50	50
1001 to 1200	0	50	50	0	50	50	0	50	50	0	50	50
1201 to 1400	0	50	50	0	50	50	0	50	50	0	50	50
1401 to 1600	1	50	50	1	50	50	1	50	50	1	50	50
1601 to 1800	0	50	50	0	50	50	1	50	50	1	50	50
1801 to 2000	1	50	50	0	49	50	0	50	50	0	50	50

* Condition: Numbers of θ simulation conditions

* G: General factor; S1: First specific factor; S2: Second specific factor

* The frequencies under the significance level of .05 were counted.

Table E-2. Numbers of Frequencies Significant by KS Test under Condition 2

Discrimination	Mean: 1.3 / SD: 0.15						Mean: 1.8 / SD: 0.15					
Difficulty	Mean: -0.5			Mean: 0.5			Mean: -0.5			Mean: 0.5		
	SD: 0.4			SD: 0.4			SD: 0.4			SD: 0.4		
θ	G	S1	S2	G	S1	S2	G	S1	S2	G	S1	S2
1 to 2000	50	50	50	50	50	50	50	50	50	50	50	50
1 to 200	1	50	50	5	50	50	1	50	50	1	50	50
201 to 400	3	50	50	4	50	50	1	50	50	3	50	50
401 to 600	3	50	50	2	50	50	3	50	50	2	50	50
601 to 800	2	50	50	4	50	50	2	50	50	3	50	50
801 to 1000	2	50	50	4	50	50	2	50	50	3	50	50
1001 to 1200	2	50	50	1	50	50	1	50	50	1	50	50
1201 to 1400	2	50	50	0	50	50	3	50	50	2	50	50
1401 to 1600	2	50	50	3	50	50	2	50	50	4	50	50
1601 to 1800	2	50	50	1	50	50	2	50	50	1	50	50
1801 to 2000	3	50	50	1	50	50	3	50	50	3	50	50

* Condition: Numbers of θ simulation conditions

* G: General factor; S1: First specific factor; S2: Second specific factor

* The frequencies under the significance level of .05 were counted.

Table E-3. Numbers of Frequencies Significant by KS Test under Condition 3

Discrimination	Mean: 1.3 / SD: 0.15						Mean: 1.8 / SD: 0.15					
Difficulty	Mean: -0.5			Mean: 0.5			Mean: -0.5			Mean: 0.5		
	SD: 0.4			SD: 0.4			SD: 0.4			SD: 0.4		
θ	G	S1	S2	G	S1	S2	G	S1	S2	G	S1	S2
1 to 2000	22	50	50	21	50	50	46	50	50	46	50	50
1 to 200	0	50	50	0	50	50	1	50	50	0	50	50
201 to 400	0	50	50	0	50	50	0	50	50	0	50	50
401 to 600	0	50	50	0	50	50	1	50	50	1	50	50
601 to 800	0	50	50	0	50	50	1	50	50	1	50	50
801 to 1000	0	50	50	0	50	50	1	50	50	1	50	50
1001 to 1200	0	50	50	0	50	50	0	50	50	0	49	50
1201 to 1400	0	50	50	0	50	50	0	50	50	0	50	50
1401 to 1600	0	50	50	0	50	50	0	50	50	1	50	50
1601 to 1800	1	50	50	0	50	50	1	50	50	0	50	50
1801 to 2000	0	50	50	0	50	50	0	50	50	0	50	50

* Condition: Numbers of θ simulation conditions

* G: General factor; S1: First specific factor; S2: Second specific factor

* The frequencies under the significance level of .05 were counted.

Table E-4. Numbers of Frequencies Significant by KS Test under Condition 4

Discrimination	Mean: 1.3 / SD: 0.15						Mean: 1.8 / SD: 0.15					
Difficulty	Mean: -0.5			Mean: 0.5			Mean: -0.5			Mean: 0.5		
	SD: 0.4			SD: 0.4			SD: 0.4			SD: 0.4		
θ	G	S1	S2	G	S1	S2	G	S1	S2	G	S1	S2
1 to 2000	50	50	50	49	50	50	50	50	50	50	50	50
1 to 200	2	50	50	4	50	50	4	50	50	5	50	50
201 to 400	3	50	50	3	50	50	4	50	50	5	50	50
401 to 600	4	50	50	4	50	50	6	50	50	6	50	50
601 to 800	0	50	50	0	50	50	2	50	50	0	50	50
801 to 1000	0	50	50	0	50	50	2	50	50	0	50	50
1001 to 1200	2	50	50	4	50	50	5	50	50	3	50	50
1201 to 1400	1	50	50	2	50	50	1	50	50	1	50	50
1401 to 1600	1	50	50	2	50	50	1	50	50	2	50	50
1601 to 1800	6	50	50	6	50	50	6	50	50	6	50	50
1801 to 2000	1	50	50	3	50	50	0	50	50	5	50	50

* Condition: Numbers of θ simulation conditions

* G: General factor; S1: First specific factor; S2: Second specific factor

* The frequencies under the significance level of .05 were counted.

Table E-5. Numbers of Frequencies Significant by KS Test under Condition 5

Discrimination	Mean: 1.3 / SD: 0.15						Mean: 1.8 / SD: 0.15					
Difficulty	Mean: -0.5			Mean: 0.5			Mean: -0.5			Mean: 0.5		
	SD: 0.4			SD: 0.4			SD: 0.4			SD: 0.4		
θ	G	S1	S2	G	S1	S2	G	S1	S2	G	S1	S2
1 to 2000	50	50	50	50	50	50	50	50	50	50	50	50
1 to 200	50	50	50	50	50	50	50	50	50	50	50	50
201 to 400	50	49	50	50	50	50	50	50	50	50	50	50
401 to 600	49	50	50	50	50	50	50	50	50	50	50	50
601 to 800	49	50	50	49	50	50	50	50	50	50	50	50
801 to 1000	49	50	50	49	50	50	50	50	50	50	50	50
1001 to 1200	49	50	50	49	50	50	50	50	50	50	50	50
1201 to 1400	50	50	50	49	50	50	50	50	50	50	50	50
1401 to 1600	50	50	50	50	50	50	50	50	50	50	50	50
1601 to 1800	50	50	50	50	50	50	50	50	50	50	50	50
1801 to 2000	50	50	50	50	50	50	50	50	50	50	50	50

* Condition: Numbers of θ simulation conditions

* G: General factor; S1: First specific factor; S2: Second specific factor

* The frequencies under the significance level of .05 were counted.

Table E-6. Numbers of Frequencies Significant by KS Test under Condition 6

Discrimination	Mean: 1.3 / SD: 0.15						Mean: 1.8 / SD: 0.15					
Difficulty	Mean: -0.5			Mean: 0.5			Mean: -0.5			Mean: 0.5		
	SD: 0.4			SD: 0.4			SD: 0.4			SD: 0.4		
θ	G	S1	S2	G	S1	S2	G	S1	S2	G	S1	S2
1 to 2000	50	50	50	50	50	50	50	50	50	50	50	50
1 to 200	50	50	50	48	50	50	50	50	50	50	50	50
201 to 400	50	50	50	50	50	50	50	50	50	49	50	50
401 to 600	49	50	50	49	50	50	50	50	50	50	50	50
601 to 800	50	50	50	50	50	50	49	50	50	49	50	50
801 to 1000	50	50	50	50	50	50	49	50	50	49	50	50
1001 to 1200	50	50	50	50	50	50	50	50	50	50	50	50
1201 to 1400	49	50	50	48	50	50	50	50	50	49	50	50
1401 to 1600	50	50	50	50	50	50	50	50	50	50	50	50
1601 to 1800	50	50	50	50	50	50	50	50	50	50	50	50
1801 to 2000	50	50	50	50	50	50	50	50	50	50	50	50

* Condition: Numbers of θ simulation conditions

* G: General factor; S1: First specific factor; S2: Second specific factor

* The frequencies under the significance level of .05 were counted.

Table E-7. Numbers of Frequencies Significant by KS Test under Condition 7

Discrimination	Mean: 1.3 / SD: 0.15						Mean: 1.8 / SD: 0.15					
Difficulty	Mean: -0.5			Mean: 0.5			Mean: -0.5			Mean: 0.5		
	SD: 0.4			SD: 0.4			SD: 0.4			SD: 0.4		
θ	G	S1	S2	G	S1	S2	G	S1	S2	G	S1	S2
1 to 2000	50	50	50	50	50	50	50	50	50	50	50	50
1 to 200	50	50	50	49	50	50	50	50	50	50	50	50
201 to 400	50	50	50	49	50	50	50	50	50	50	50	50
401 to 600	48	50	50	50	50	50	49	50	50	50	50	50
601 to 800	50	50	50	49	50	50	50	50	50	49	50	50
801 to 1000	50	50	50	49	50	50	50	50	50	49	50	50
1001 to 1200	50	50	50	50	50	50	50	50	50	50	50	50
1201 to 1400	50	50	50	50	50	50	50	50	50	50	50	50
1401 to 1600	50	50	50	49	49	50	49	50	50	50	50	50
1601 to 1800	50	50	50	49	50	50	50	50	50	50	50	50
1801 to 2000	50	50	50	50	50	50	50	50	50	50	50	50

* Condition: Numbers of θ simulation conditions

* G: General factor; S1: First specific factor; S2: Second specific factor

* The frequencies under the significance level of .05 were counted.

Table E-8. Numbers of Frequencies Significant by KS Test under Condition 8

Discrimination	Mean: 1.3 / SD: 0.15						Mean: 1.8 / SD: 0.15					
Difficulty	Mean: -0.5			Mean: 0.5			Mean: -0.5			Mean: 0.5		
	SD: 0.4			SD: 0.4			SD: 0.4			SD: 0.4		
θ	G	S1	S2	G	S1	S2	G	S1	S2	G	S1	S2
1 to 2000	50	50	50	50	50	50	50	50	50	50	50	50
1 to 200	50	50	50	50	50	50	50	50	50	49	50	50
201 to 400	50	50	50	50	50	50	50	50	50	50	50	50
401 to 600	50	50	50	48	50	50	50	50	50	49	50	50
601 to 800	50	50	50	50	50	50	50	50	50	50	50	50
801 to 1000	50	50	50	50	50	50	50	50	50	50	50	50
1001 to 1200	50	50	50	49	50	50	50	50	50	50	50	50
1201 to 1400	50	50	50	48	50	50	49	50	50	49	50	50
1401 to 1600	50	50	50	50	50	50	50	50	50	50	50	50
1601 to 1800	49	50	50	50	50	50	50	50	50	49	50	50
1801 to 2000	49	50	50	48	50	50	49	50	50	49	50	50

* Condition: Numbers of θ simulation conditions

* G: General factor; S1: First specific factor; S2: Second specific factor

* The frequencies under the significance level of .05 were counted.

BIBLIOGRAPHY

BIBLIOGRAPHY

- Axelrod, R. (2005). Advancing the art of simulation in the social sciences. In J.-P. Rennard (Ed.), *Handbook of research on nature inspired computing for economy and management* (pp. 90-100). Hersey, PA: Idea Group.
- Batley, R. -M., & Boss, M. W. (1993). The effects on parameter estimation of correlated dimensions and a distribution-restricted trait in a multidimensional item response model. *Applied Psychological Measurement*, 17(2), 131-141.
- Bratley, P., Fox, B., & Schrage, L. (1987). *A guide to simulation*. Second Edition. New York: Springer-Verlag.
- Cai, L. (2010). A two-tier full-information item factor analysis model with applications. *Psychometrika*, 75(4), 581-612.
- Cai, L., Yang, J.S., & Hansen, M. (2011). Generalized full-information item bifactor analysis. *Psychological Methods*, 16(2), 221-248.
- Capella, M. E., & Turner, R. C. (2004). Development of an instrument to measure consumer satisfaction in vocational rehabilitation. *Rehabilitation Counseling Bulletin*, 47(2), 76-85.
- Chalmers, R. P. (2012). A Multidimensional item response theory package for the R Environment. *Journal of Statistical Software*, 48(6), 1-29.
- Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, 41(2), 189-225.
- DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, 2(3), 292-307.
- DeMars, C. E. (2006). Application of the bi-factor multidimensional item response theory model to testlet-based tests. *Journal of Educational Measurement*, 43(2), 145-168.
- Duncan-Jones, P. (1981a). The structure of social relationships: Analysis of a survey instrument: I. *Social Psychiatry*. *Social Psychiatry*, 16(2), 55-61.
- Duncan-Jones, P. (1981b). The structure of social relationships: Analysis of a survey instrument: II. *Social Psychiatry*, 16(3), 143-149.
- Eboli, L., & Mazzulla, G. (2007). Service quality attributes affecting customer satisfaction for bus transit. *Journal of Public Transportation*, 10(3), 21-34.
- Fahrmeir, L., & Tutz, G. (2001). *Multivariate statistical modeling based on generalized linear models* (2nd ed.). New York, NY: Springer.

- Finch, H. (2010). Item parameter estimation for the MIRT model: Bias and precision of confirmatory factor analysis-based models. *Applied Psychological Measurement*, 34(1), 10-26.
- Frank, K. (1998). The social context of schooling: Quantitative methods. *Review of Research in Education*, 23, 171-216.
- Fujimoto, K. A. (2014). Bayesian Extended Two-Tier Full-information Item Factor Analysis Model. Paper presented at the 76th Annual conference of the National Council on Measurement in Education, Philadelphia, PA.
- Gibbons, R. D., & Hedeker, D. R. (1992). Full-information item bi-factor analysis. *Psychometrika*, 57(3), 423-436.
- Gifford, J. A. (1978). Developments in latent trait theory: Models, technical issues, and applications. *Review of Educational Research*, 48(4), 467-510.
- Gosz, J. K., & Walker, C. M. (2002). An empirical comparison of multidimensional item response data using TESTFACT and NOHARM. In annual meeting of the National Council on Measurement in Education, New Orleans, LA.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications, Inc.
- Hichendorff, M. (2013). The language factor in elementary mathematics assessments: Computational skills and applied problem solving in a multidimensional IRT framework. *Applied Measurement in Education*, 26(4), 253-278.
- Hogg, R. V., & Tanis, E. A. (1997). *Probability and statistical inference*. Upper Saddle River, NJ: Prentice Hall.
- Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika*, 2(1), 41-54.
- Jung, I, Choi, S, Lim, C, & Leem, J. (1994). Effects of different types of interaction on learning achievement, satisfaction and participation in web-based instruction. *Innovations in Education and Teaching International*, 39(2), 153-162.
- Küppers, G., & Lenhard, J. (2005). Validation of simulation: Patterns in the social and natural sciences. *Journal of Artificial Societies and Social Simulation*, 8(4)3. (<http://jasss.soc.surrey.ac.uk/8/4/3.html>)
- Lazarsfeld, P. F. (1950). The logical and mathematical foundation of latent structure analysis. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, & J. A. Clausen (Eds.), *Studies in social psychology in World War II: Vol. 4. Measurement and prediction* (pp. 362-412). Princeton, NJ : Princeton University Press.
- Li, Y., Bolt, D. M., & Fu, J. (2006). A comparison of alternative models for testlets. *Applied Psychological Measurement*, 30(1), 3-21.

- Li, Y., & Lissitz, R. W. (2012). Exploring the full-information bifactor model in vertical scaling with construct shift. *Applied Psychological Measurement*, 36(1), 3-20.
- Lord, F. M. (1952). A theory of test scores. *Psychometric Monograph*, No7.
- Lord, F. M. (1953a). An application of confidence intervals and of maximum likelihood to the estimation of an examinee's ability. *Psychometrika*, 18, 57-75.
- Lord, F. M. (1953b). The relation of test score to the trait underlying the test. *Educational and Psychological Measurement*, 13, 517-548.
- Lord, F. M. (1980). Applications of item response theory to practical testing problems. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Marsden, P. V. (2005). Recent Developments in Network Measurement. in P. J. Carrington, J. Scott, & S. Wasserman (Eds.) *Models and Methods in Social Network Analysis* (pp.8-30). New York: Cambridge University Press.
- Martin, A. J. (2007). Examining a multidimensional model of student motivation and engagement using a construct validation approach. *British Journal of Educational Psychology*, 77(2), 412-440.
- McDonald, R. P. (1997). Normal-Ogive multidimensional model. In W. J. van der Linden & P. K. Hambleton (Eds.) *Handbook of modern item response theory* (pp. 257-269). New York: Springer.
- Murphy, K. R., Cronin, B. E., & Tam, A. P. (2003). Controversy and consensus regarding the use of cognitive ability testing in organizations. *Journal of Applied Psychology*, 88(4), 660-671.
- National Center for Education Statistics (2010). Highlights from PISA 2009. (NCES 2011-004). U.S. Department of Education. Retrieved from NCES (<http://nces.ed.gov/pubs2011/2011004.pdf>).
- National Center for Education Statistics (2007). Highlights from PISA 2006. (NCES 2008-016). U.S. Department of Education. Retrieved from NCES (<http://nces.ed.gov/pubs2008/2008016.pdf>).
- National Center for Education Statistics (2004). PISA 2003 results from the U.S. perspective highlights. (NCES 2005-003). U.S. Department of Education. Retrieved from NCES (<http://nces.ed.gov/pubs2005/2005003.pdf>).
- National Center for Education Statistics (2001). Outcomes of learning: Results from the 2000 Program for International Student Assessment of 15-year-olds in reading, mathematics, and science literacy (NCES 2002-115). U.S. Department of Education. Retrieved from NCES (<http://nces.ed.gov/pubs2002/2002115.pdf>).
- Nelsen, R. B. (1999). An introduction to copulas. Springer.
- Organization for Economic Cooperation and Development. (2001). Knowledge and Skills

for Life: First Results from the OECD Programme for International Student Assessment. Paris: Author.

Organization for Economic Cooperation and Development (2007a). PISA 2006: Science Competencies for Tomorrow's World Executive Summary, Paris: Author.

Organization for Economic Cooperation and Development. (2007b). PISA 2006: Science Competencies for Tomorrow's World (Volume I: Analysis.) Paris: Author.

Pommerich, M., & Segall, D. O. (2008). Local dependence in an operational CAT: Diagnosis and implications. *Journal of Educational Measurement*, 45(4), 201-223.

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, 16, 19-31.

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, 92(6), 544-559.

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

Rijmen, F. (2010). Formal relations and an empirical comparison among the bi-factor, the testlet, and a second-order multidimensional IRT model. *Journal of Educational Measurement*, 47(3), 361-372.

Robinson, S. (2004). *Simulation: The practice of model development and use*. Wiley, Chichester, UK.

Sass, D. A., Schmitt, D. A., & Walker, C. M. (2008) Estimating non-normal latent trait distributions within item response theory using true and estimated item parameters. *Applied Measurement in Education*, 21(1), 65-88.

Schumid, A. (2005). What is the truth of simulation? *Journal of Artificial Societies and Social Simulation*, 8(4)5. (<http://jasss.soc.surrey.ac.uk/8/4/5.html>)

Seo, D. G. (2011). *Application of the bifactor model to computerized adaptive testing*. Unpublished doctoral dissertation, University of Minnesota. (ERIC Document Reproduction Service No.ED526366) Retrieved July 27, 2012, from ERIC database.

Sheng, Y. (2010). Bayesian estimation of MIRT models with general and specific latent traits in MATLAB. *Journal of Statistical Software*, 34(3), 1-27.

Singelis, T. M. (1994). The measurement of independent and interdependent self-construals. *Personality and Social Psychology Bulletin*, 20(5), 580-591.

Stapleton, J. H. (2008). *Models for probability and statistical inference: Theory and applications*. Hoboken, NJ: John Wiley & Sons.

Stone, C. A. (1992). Recovery of marginal maximum likelihood estimates in the two-parameters logistic response model: An evaluation of MULTILOG. *Applied Psychological Measurement*, 16(1), 1-16.

Stouffer, S.A. (1950). An overview of the contributions to scaling and scale theory. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, & J. A. Clausen (Eds.), *Studies in social psychology in World War II: Vol. 4. Measurement and prediction* (pp. 3-45). Princeton, NJ : Princeton University Press.

Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago: University of Chicago press.

von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, 61(2), 287-307.

Woods, C. M., & Thissen, D. (2006). Item response theory with estimation of the latent population distribution using spline-based densities. *Psychometrika*, 71(2), 281-301.

Yoshida, M., & James, J. D. (2010). Customer satisfaction with game and service experiences: Antecedents and consequences. *Journal of Sport Management*, 24, 338-361.