© 2016 Youngshil Paek

PSEUDO-LIKELIHOOD ESTIMATION OF MULTIDIMENSIONAL POLYTOMOUS ITEM RESPONSE THEORY MODELS

BY

YOUNGSHIL PAEK

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Educational Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Doctoral Committee:

Professor Carolyn J. Anderson, Chair
Professor Jeffrey A. Douglas
Associate Professor Jinming Zhang
Assistant Professor Steven Culpepper

Abstract

Log-multiplicative association (LMA) models, special cases of log-linear models, can be used as multidimensional item response theory (MIRT) models for polytomous items (Anderson, Verkuilen and Peyton, 2010; Anderson, 2013). LMA models do not require numerical integration for their estimation nor do they require assumptions regarding the marginal distribution of the latent variables. However, maximum likelihood estimation (MLE) of LMA models requires iteratively computing fitted values for all possible response patterns. Standard estimation methods for large numbers of items fail because the number of possible response patterns increases exponentially as the number of items and response options per item increase. In this study, a new algorithm is proposed to solve this estimation problem.

Anderson, Li and Vermunt (2007) proposed using pseudo-likelihood estimation (PLE); however, their solution only applies to models in the Rasch family, which exploits the relationship between log-linear and logistic regression models. Their method is extended to more general models by adding an additional step that estimates slope (item discrimination) parameters for the latent variables.

The new algorithm has two basic steps and simplifies for special cases. In Step 1, a (multinomial) logistic regression model is fit by MLE to one item using rest-scores as an explanatory variable to get new estimates of item slopes that are used in the rest-score for the next item. This process is repeated for each item until all item slopes have been up-dated. Step 2 involves fitting a single conditional logistic regression model for a data set formed by stacking the conditional logistic regressions for each item. This yields new estimates of location (item difficulty) parameters and the covariance matrix for the latent variables. Steps 1 and 2 are repeated until all parameter estimates converge.

The results of simulation and empirical studies with real data show that the proposed algorithm successfully estimates parameters in more general LMA models with both location and slope parameters as MIRT models.

To My Parents

Acknowledgments

I would like to express my deepest gratitude to my academic advisor Dr. Carolyn J. Anderson. Dr. Anderson has always been an admirable and brilliant mentor to me, has shown substantial support for my study and work with her inspiration and encouragement, and has strengthened my determination in pursuit of my dream. Her persistent efforts and patience have assisted me overcome many crisis situations and finish this dissertation.

I also wish to give my sincere appreciation to the members of my dissertation committee, Dr. Jeffrey A. Douglas, Dr. Jinming Zhang, and Dr. Steven Culpepper for their constructive comments and insightful suggestions to help me improve my dissertation. I am honored to have each of them on my committee.

I am also grateful to my dear friends, Jin-hee, Jong-ok, Jeong-rae, Sung-sook, Eun-young, and Yeonsook. Their support, care, and encouragement that warms my heart helped me overcome setbacks and stay focused on my graduate study. I greatly value their friendship and I deeply appreciate their belief in me. They are not only my friends but also my sisters.

Most importantly, none of this would have been possible without the love and patience of my family. My parents to whom this dissertation is dedicated to, have been a constant source of love, concern, support, and strength all these years. I would like to express my heart-felt gratitude to my supportive brothers and their families. Finally, I am deeply thankful to my husband, Taeho and my wonderful sons, Wonwoo and Junwoo, who have stayed with me throughout this journey. I couldn’t have made it without their constant love, dedication, and positive energy.

Table of Contents

List of Tables

List of Figures

Figure Page

Chapter 1

Introduction

Log-multiplicative association (LMA) models, special cases of log-linear models, are implied by different underlying structures. Although multivariate normality implies LMA models for data, the emphasis in this thesis is on LMA models as multidimensional item response theory (MIRT) models. One hindrance to more widespread use of LMA models as MIRT models is that current estimation methods are limited relatively small numbers of items. An algorithm to overcome this limitation is proposed and its performance is evaluated in this thesis.

Estimation of MIRT Models for Polytomous Items

Questionnaire or test items with more than two response options (i.e., polytomous items) are frequently administered to examinees in educational and psychological settings. Item response theory (IRT) models have been developed for polytomous items. Depending on the restrictions on slope (item discrimination) parameters of the latent trait in parameterizations of the models, polytomous IRT models may be classified into either Rasch family models or more general models where the slope parameters are fee to vary across items. Unidimensional polytomous IRT models where slope parameters vary across items include Samejima (1969)'s graded response model (GRM), Muraki (1992)'s generalized partial credit model (GPCM) for ordered responses, and Bock (1972)'s nominal response model (NRM) for items with a non-specified response order. The slope parameters of the GRM and GPCM are constant over the response options; whereas, in the NRM, the slope parameters may vary. In this thesis, research interest lies in estimating slope parameters that may vary across response categories within an item and over items; that is, Bock (1972)'s NRM.

Multidimensional item response theory (MIRT) has been developed, incorporating multiple latent traits into IRT models. It is regarded as a useful tool for exploring the underlying dimensionality of an IRT model. There have been several multidimensional extensions of traditional IRT models for polytomous items (Reckase, 2009). These include the multidimensional graded response model (Muraki & Carlson,1993), the multidimensional partial credit model (Kelderman & Rijkes, 1994), and, more recently, the multidimensional generalized partial credit model (Yao & Schwarz, 2006). Although the usefulness of MIRT has been known for many years in the psychological and educational literature (Ackerman, 1994; Embretson, 1991; Reckase, 1985; Reckase & McKinley, 1991), the estimation of the parameters for MIRT models is challenging.

The parameters of MIRT models can be estimated by the marginal maximum likelihood estimation (MMLE), which was developed by Bock and Lieberman (1970) and elaborated with EM algorithm by Bock and Aitkin (1981). The MMLE procedure regards the observed response patterns as random samples drawn from a population and assumes the distribution of the latent variables. By numerically integrating out the person parameters, marginal likelihood functions in terms of the item parameters are obtained and then item parameters are estimated without dependence on latent variables ($\boldsymbol{\theta}$) of individual examinee.

The MMLE is preferred over other estimation methods because it yields consistent item parameter estimates and can be applied to all of uni- and multidimensional IRT models. Its popularity can be found by many computer programs employing the procedure for MIRT models such as TESTFACT (Bock, Gibbsons, Schilling, Muraki, Wilson, & Wood, 2003), flexMIRT (Cai, 2013), LISREL (Jöreskog & Sörbom, 2004), and Mplus (Muthén and Muthén, 2012). The MMLE approach is also used in PROC/NLMIXED in SAS when the parameters of MIRT models are estimated as nonlinear mixed models (De Boeck & Wilson, 2004; Rijmen, Tuerlinckx, De Boeck, & Kuppens, 2003; Sheu, Chen, Su, & Wang, 2005).

MMLE requires the user to assume the marginal distribution of the latent variable and involves numerically integrating the latent variable out of the model for parameter estimation. This method becomes problematic for multiple latent variables because it requires multiple numerical integrations. Bock, Gibbons, and Muraki (1988) report in their study on full information item factor analysis that the number of dimensions was limited to five factors because of the heavy computational work in MMLE/EM algorithm.

As an alternative for higher dimensionality, Bayesian estimation procedure with Markov chain Monte Carlo (MCMC) methods is used for estimating parameters in MIRT models, but it is extremely time consuming and requires highly advanced computer programming skills with mathematical knowledge.

Estimation of LMA Models as MIRT Models

To alleviate these problems, an easier and more flexible way for parameter estimation in MIRT models can be provided by log-multiplicative association (LMA) models. LMA models are special cases of log-linear models where all two-way interaction terms between pairs of variables (i.e., items) are replaced by products of category scales values and an association parameter (Anderson & Vermunt, 2000). LMA models have a number of advantages as MIRT models: They do not require numerical integration for their estimation nor do they require assumptions regarding the marginal distribution of the latent variables. Covariates can be included in the model and they can be estimated quickly in SAS.

There are at least two derivations of LMA models as item response models. One derivation is due to Holland (1990) through his Dutch Identity and latter extensions of the model to polytomous items[1] (Hessen, 2012; Li, 2010). The other derivation was proposed in Anderson and Yu (2007) for dichotomous items based on fully conditionally specified logistic models using a rest-score in lieu of the latent variable. Of the two derivations, this study focuses more on the fully conditional specification derivation of LMA models. Anderson and Yu (2007) proposed to use a rest-score as an estimate of the latent variable based on the precedence and justification for it in the literature on classical test theory and IRT as mentioned in Junker and Sijtsma (2000). In this approach, logistic regression models are specified for each item conditional on responses to all others. They also showed that the set of fully conditionally specified models uniquely implies an LMA model for the joint distribution based on a proof given by Joe and Liu (1996). The fully conditional derivation was later generalized to polytomous items and multidimensional models (Anderson, Li, & Vermunt, 2007; Anderson, Verkuilen, & Peyton, 2010; Anderson, 2013).

The parameters in LMA models are typically estimated by maximum likelihood estimation (MLE) using computer programs such as ℓEM (Vermunt, 1997), SAS nonlinear programming procedure (PROC/NLP), R, and MatLab. When estimating the parameters in LMA models with MLE, it requires iteratively computing fitted values for all possible response patterns. Parameter estimates of LMA models for small numbers of items can be obtained by MLE easily because the number of all possible response patterns is reasonable. For large numbers of items, however, the standard estimation methods of LMA models fail because the number of possible response patterns increases exponentially as the number of items and response options per item increase. More recently, pseudo-likelihood estimation (PLE) was

---

[1] The derivation by Dutch Identity is formally equivalent to graphical model derivation (Anderson & Vermunt, 2000; Anderson & Böckenholt, 2000; Anderson, 2002). The graphical derivation is more general.

proposed to solve the problem for LMA Rasch models with large numbers of items (Anderson, Li, & Vermunt 2007). It is reported that PLE can fit the models to data with large numbers of items successfully and fast, handle models with multiple latent variables and covariates, yield consistent estimates, and is easy to implement.

Pseudo-likelihood Estimation for LMA Models and Its Limitation

Pseudo-likelihood estimation (PLE) was first introduced by Besag (1974) as an approach to the specification and analysis of spatial interaction. The basic idea behind PLE is to replace numerically challenging problems with more tractable ones by simplifying them with conditional specification approach so that computational demands of fitting models to data.

Anderson, Li, and Vermunt (2007) implemented PLE for LMA Rasch models with polytomous items and multiple correlated latent variables. Following Anderson and Yu (2007)'s fully conditional specification approach, they specified conditional models corresponding to each item using rest score in lieu of the latent variable and defined the pseudo-likelihood as the product of the likelihoods of the conditional multinomial logistic regressions. The whole set of fully conditionally specified logistic regression models were "stacked" into a large design matrix and the model parameters were estimated by fitting a conditional multinomial logistic regression model to the data. The maximum value of the likelihood of the model fit to the stacked data equals the pseudo-likelihood. Based on their simulation studies on the performance of PLE, the estimates obtained by PLE and MLE were almost identical and the parameter recovery of PLE was excellent for large numbers of binary or polytomous items with a single or multiple latent variables (i.e., multidimensional generalizations of Bock's NRM and all special cases).

Although they have shown that parameters in LMA models with large number of items can be estimated very fast and easily by PLE, their application of current PLE is limited to models in the Rasch family (i.e., 1PL model). In this study, their method is extended to more general models such as 2PL model, Bock’s nominal response model, and multidimensional generalizations of these models by adding an additional step that estimates the slope parameters for the latent variables.

Research Objectives

Throughout this study, the performance of the proposed PLE algorithm for more general LMA models as MIRT models are examined. The three main goals of this study are: (1) how well does the newly proposed step for estimating slope parameters perform?; (2) how well does PLE of LMA models using the new two-step algorithm perform relative to MLE of LMA models?; and lastly, (3) how well and fast does the algorithm of PLE perform for LMA models as MIRT models with large numbers of items?

The remainder of this thesis is structured as follows. Chapter 2 provides an overview of the development of the LMA models as multidimensional item response theory (MIRT) models for polytomous items, along with two derivations of LMA models as IRT models. Chapter 3 presents two estimation procedures for LMA models as MIRT models, maximum likelihood estimation (MLE) and pseudo-likelihood estimation (PLE), followed by its implementation for LMA Rasch models (i.e., linear-by-linear models). Chapter 4 introduces a new estimation algorithm for more general LMA models where both location and slope parameters are included, along with the implementation of the estimation method in SAS. Chapter 5 describes the methodology for simulation studies conducted to investigate the performance of the proposed algorithm, followed by possible ways to obtain correct standard errors of pseudo-likelihood estimates. Chapter 6 provides the detailed results of simulation studies in terms of item parameter recovery. Chapter 7 describes the results of empirical studies conducted to provide the practical use of PLE using real data. Lastly, Chapter 8 provides the findings and their implications, along with the possible further research.

Chapter 2

LMA Models as Multidimensional Item Response Theory (MIRT) Models

The purpose of this chapter is to provide an overview of the development of the LMA models as multidimensional item response theory (MIRT) models for polytomous items. After a brief review of IRT models for polytomous items, compensatory MIRT models for nominal responses are discussed. In the subsequent section, two derivations of LMA models as IRT models are presented with the connection between them, followed by research showing the flexibility of the approach. Of the two derivations, this study focuses more on the fully conditional specification derivation of LMA models (Anderson & Yu, 2007; Anderson, Verkuilen, & Peyton, 2010; Anderson, 2013).

Multidimensional Item Response Models for Polytomous Items

Although most IRT models assume unidimensionality (i.e., all of the items on a test are measuring only one latent trait or ability), there are situations where this assumption does not hold. For example, questionnaires or tests are often designed to measure multiple skills/abilities and more than one latent trait may underlie responses to items. Ackerman (1994) states that the assumption of unidimensionality must be considered very carefully and should always be verified when modeling a set of items.

Multidimensional item response theory (MIRT) has been developed, incorporating multiple latent traits into IRT models. It is regarded as a useful tool for exploring the underlying dimensionality of an IRT model. There have been several multidimensional extensions of traditional IRT models for polytomous items (Reckase, 2009). These include the multidimensional graded response model (Muraki & Carlson, 1993), the multidimensional partial credit model (Kelderman & Rijkes, 1994), and, more recently, the multidimensional generalized partial credit model (Yao & Schwarz, 2006).

Multidimensional Compensatory IRT Model for Nominal Responses

One purpose of this study is to estimate the parameters of LMA models that correspond to slope parameters of MIRT models when the slopes vary across response categories within an item and over items. When only one latent variable is considered in the model, Bock (1972)'s NRM is the model of interest. In this section, Bock's NRM will be reviewed, followed by a multidimensional compensatory polytomous IRT model for nominal responses that is a generalization of Bock's NRM.

Bock (1972)'s nominal response model was designed for polytomous items where all of the items are reflecting a single latent variable and the responses of the items do not (necessarily) have a pre-specified order. NRM is a multinomial logistic model that specifies the probability that an examinee with a given value of the latent variable (i.e., θ) selects the response option *j* on item *i*. Formally, Bock's NRM[2] is

$$\mathrm{P}(Y_i = j|\theta) = \frac{\exp(\lambda_{ij} + \nu_{ij}\theta)}{\sum_h \exp(\lambda_{ih} + \nu_{ih}\theta)}\ , \tag{2.1}$$

where ν_{ij} is an unknown slope (i.e., item discrimination) parameter for response *j* of item *i*, and λ_{ij} is a location parameter (i.e., item difficulty) for response *j* on item *i*. The sum in the denominator ensures that the sum of probabilities over all response options on item *i* equals 1.

Special cases of the NRM include the Rasch model for polytomous responses (Andersen, 1995) and two-parameter logistic (2PL) model for dichotomous responses (Alasuutari, Bickman,

---

[2] Note that the notation differs from more standard notation so that connections with other models are more transparent.

& Brannen, 2008; Bartholomew & Knott, 1999; Heinen, 1993, 1996). Anderson, Verkuilen, and Peyton (2010) showed that Bock (1972)'s NRM leads to LMA models where a rest-score is substituted for θ in (2.1). They specified a multinomial logistic regression model for each item and showed that the set of multinomial logistic regression models yields LMA models for the joint distribution of observed responses to all items (i.e., response patterns).

When multiple latent variables underlie responses to nominal items, the unidimensional model in equation (2.1) can be extended to a multidimensional model. The multidimensional model is

$$\mathrm{P}(Y_i = j|\theta_1, \dots \theta_M) = \frac{\exp(\lambda_{ij} + \sum_m \nu_{ijm}\theta_m)}{\sum_h \exp(\lambda_{ih} + \sum_m \nu_{ihm}\theta_m)} \ , \qquad (2.2)$$

where ν_{ijm} is an unknown slope or discrimination parameter for response *j* on item *i* on latent variable θ_m, λ_{ij} is the location or difficulty parameter, and the sum in the denominator ensures that the sum of probabilities over all response options on item *i* equals 1. Given values on the *M* latent traits $\boldsymbol{\theta} = (\theta_1, \dots \theta_M)$, this model specifies the probability that an examinee selects the response option *j* of item *i*.

Model (2.2) includes many well-known special cases. If responses are dichotomous, model (2.2) is equivalent to a multidimensional compensatory version of the 2PL model as presented by McKinley and Reckase (1983). When the slope or discrimination parameters are fixed or assumed to be known, Bock's NRM and its multidimensional models are corresponding to a Rasch model for polytomous responses (Andersen, 1995) and its multidimensional extension (Fischer, 1995).

Although the usefulness of MIRT has been known for many years in the psychological and educational literature (Ackerman, 1994; Embretson, 1991; Reckase, 1985; Reckase &

McKinley, 1991), the estimation of the parameters for MIRT models is challenging. The parameters of MIRT models can be estimated as nonlinear mixed models using marginal maximum likelihood method (MMLE), which yields consistent parameter estimates. However, MMLE procedure involves numerical integration of the latent variable and the parameter estimation gets more complicated as the number of the latent variables increases. Markov chain Monte Carlo (MCMC) (e.g., Metropolis-Hastings Robbins-Monro estimation) is one of the methods for estimating parameters of MIRT models, but it is computationally demanding. Another potential solution to the problem is connecting IRT models with log-multiplicative models (LMA), which do not use the numerical integration. In the next section, LMA models as IRT models will be discussed.

Log-Multiplicative Models (LMA) as Item Response Models

As discussed in Chapter 1, there are two basic derivations of LMA models as item response models. In this section, the two derivations will be reviewed.

Holland's Dutch Identity[3]

The first derivation of LMA models as item response models was made by Holland (1990) for dichotomous items. He pointed out that standard IRT models based on marginal maximum likelihood estimation encounter intractable integral problems, which obstruct the further understanding of the models. As a solution to this problem, he introduced the Dutch Identity, which establishes a model for probabilities of response patterns (i.e., log P($\boldsymbol{y}$) where $\boldsymbol{y}$ is a response pattern) for binary item responses. In his approach, the manifest probabilities of response patterns are assumed to follow a multinomial distribution. Under conditional (or local)

---

[3] Holland called the Theorem the "Dutch Identity" because he developed it while in Holland.

independence, the distribution of the manifest probabilities for a response pattern $\boldsymbol{y}$, P($\boldsymbol{y}$) in the standard IRT models is given as below:

$$\begin{aligned}\mathrm{P}(\boldsymbol{y}) &= \int P(\boldsymbol{y}|\theta)f(\theta)d(\theta) \\ &= \int \prod_{i=1} P(y_i = 1|\theta)^{y_i} P(y_i = 0|\theta)^{(1-y_i)} f(\theta)d(\theta)\ ,\end{aligned} \tag{2.3}$$

where $\boldsymbol{y}$ is a response pattern for *I* items, and $y_i = 1$ if the response is correct and $y_i = 0$ if the response is incorrect. The Dutch Identity is restated below.

Theorem 1. (The Dutch Identity; Holland, 1990). If the manifest probabilities P($\boldsymbol{y}$) satisfies (2.3), then for any fixed response pattern $\boldsymbol{y}_J$,

$$\frac{P(\boldsymbol{y})}{P(\boldsymbol{y}_J)} = E\left\{\exp\left[\left(\boldsymbol{y} - \boldsymbol{y_J}\right)^T \boldsymbol{\eta}(\theta)\right] \middle| \boldsymbol{Y} = \boldsymbol{y_J}\right\}, \tag{2.4}$$

where $\boldsymbol{\eta}(\theta) = (\eta_1(\theta), \eta_2(\theta), \ldots, \eta_i(\theta))^T$ and $\eta_i(\theta)$ is the item logit function,

$$\eta_i(\theta) = \log\left(\frac{P_i(\theta)}{Q_i(\theta)}\right) = \log\left[\frac{P(y_i = 1|\theta)}{P(y_i = 0|\theta)}\right] \quad \text{for } i = 1, 2, \ldots, I. \tag{2.5}$$

Holland derived second-order log-linear models (i.e., LMA models) as item response models using a corollary to the Dutch Identity where $\boldsymbol{\theta}$ is a column vector (i.e., multidimensional case). In the corollary, he added two assumptions: posterior normality of the latent variables given the response pattern and the linearity of item logit functions. Using slightly different notation from those used by Holland, his corollary is restated below.

Corollary 1. (Holland, 1990). If, for some choice of $\boldsymbol{y_J}$, the posterior distribution of $\boldsymbol{\theta}|\boldsymbol{Y} = \boldsymbol{y_J}$ is a *D*-dimensional normal, that is,

$$\boldsymbol{\theta}|\boldsymbol{Y} = \boldsymbol{y_J} \text{ is } N_D\left(\boldsymbol{\mu}_{\boldsymbol{y}_J}, \boldsymbol{\Sigma}_{\boldsymbol{y}_J}\right),$$

and if the item logit functions $\eta_i(\boldsymbol{\theta})$ are linear, that is,

$$\eta_i(\boldsymbol{\theta}) = \eta_i\left(\boldsymbol{\mu}_{y_J}\right) - \boldsymbol{a}_i^T\left(\boldsymbol{\theta} - \boldsymbol{\mu}_{y_J}\right),$$

where $\boldsymbol{a}_i^T = (a_{1i}, a_{2i}, \dots, a_{Di})$. Then,

$$\log P(\boldsymbol{y}) = \log P(\boldsymbol{y_J}) + \left(\boldsymbol{y} - \boldsymbol{y_J}\right)^T \boldsymbol{\eta}\left(\boldsymbol{\mu}_{y_J}\right) + \frac{1}{2}\left(\boldsymbol{y} - \boldsymbol{y_J}\right)^T \boldsymbol{A\Sigma}_{y_J}\boldsymbol{A^T}\left(\boldsymbol{y} - \boldsymbol{y_J}\right), \tag{2.6}$$

where $\boldsymbol{A}^T = (\boldsymbol{a_1}, \boldsymbol{a_2}, \dots, \boldsymbol{a_I})$is a $D \times I$ matrix.

The above corollary can be directly re-written for unidimensional case (i.e., $D = 1$). If, for some reference response $\boldsymbol{y_J}$**,** the posterior distribution of $\theta|\boldsymbol{Y} = \boldsymbol{y_J}$ is normal with mean $\mu_{\boldsymbol{y_J}}$ and variance $\sigma^2_{\boldsymbol{y_J}}$, that is,

$$\theta|\boldsymbol{Y} = \boldsymbol{y_J} \text{ is } N\left(\mu_{\boldsymbol{y_J}}, \sigma^2_{\boldsymbol{y_J}}\right),$$

and if the item logit functions $\eta_i(\theta)$ are linear, that is,

$$\eta_i(\theta) = \eta_i\left(\mu_{\boldsymbol{y_J}}\right) - a_i\left(\theta - \mu_{\boldsymbol{y_J}}\right).$$

Then,

$$\log P(\boldsymbol{y}) = \log P(\boldsymbol{y_J}) + \left(\boldsymbol{y} - \boldsymbol{y_J}\right)^T \boldsymbol{\eta}\left(\mu_{\boldsymbol{y_J}}\right) + \frac{1}{2}\sigma^2_{\boldsymbol{y_J}}\left[\left(\boldsymbol{y} - \boldsymbol{y_J}\right)^T \boldsymbol{a}_i\right]^2, \tag{2.7}$$

where $\boldsymbol{a}_i = (a_1, a_2, a_3, \dots a_i)^T$ and $\boldsymbol{\eta}\left(\mu_{\boldsymbol{y_J}}\right) = \left(\eta_1\left(\mu_{\boldsymbol{y_J}}\right), \eta_2\left(\mu_{\boldsymbol{y_J}}\right), \eta_3\left(\mu_{\boldsymbol{y_J}}\right), \dots \eta_i(\mu_{\boldsymbol{y_J}})\right)^T$.

He conjectured that the model given in (2.7) is a limiting form for all “smooth” unidimensional IRT models when the number of items is large.

The Dutch Identity provides a simple way for analyzing item response models with the marginal likelihood function of an item response model that does not use numerical integration. This advantage allows the theorem to be applied in several ways for specifying IRT models with large numbers of items, studying the structure of the latent variable models, testing the dimensionality of the latent variables, clearing the problems away in forming item response function (IRF) and latent trait distribution from sample data and so on (Holland, 1990).

A number of studies on the Dutch Identity were performed for the purpose of examining the assumptions or conjectures made by Holland (Chang & Stout, 1993; Chang, 1996; Zhang & Stout, 1997). Chang and Stout (1993) proved that the asymptotic posterior normality of the latent variable given response patterns under nonrestrictive nonparametric assumptions holds for a long test with dichotomously scored items. Chang (1996) extended the results to polytomous IRT models and established that the asymptotic posterior normality of the latent variable could also be assumed in the models.

Zhang and Stout (1997) weakened the two assumptions of posterior normality of the latent variable and linear logit functions (i.e., 2PL). By counterexamples, they demonstrated that the Dutch Identity conjecture does not always hold; however, when the condition of posterior normality was weakened to asymptotic posterior normality and the counterexamples were not likely distributions of theta (θ).

There have also been extensions of the Dutch Identity for dichotomous items to polytomous items (Hessen, 2012; Li, 2010). Hessen (2012) derived the polytomous Dutch Identity theorem to develop polytomous log-linear by linear association models (LLLA), which are special cases of LMA models. Hessen (2012)'s derivation is general but only special cases of responses. The equivalence between LMA models and Bock's NRM was first noted in Anderson and Böckenholt (2000).

Hessen (2012) also presented an extension of the Dutch Identity that can be applied to the multidimensional partial credit model. By using the extension, he derived a conditional multinormal partial credit model (i.e., a special case of LMA model) that does not require numerical integration or assume a marginal multivariate normal distribution of the latent variables in the total population for maximum likelihood estimation. Like Holland, he assumed posterior or conditional normality of the latent variable given a response pattern, $\boldsymbol{y}$ (hence the name "conditional multinormal partial credit model"). He mentioned that his model should be extended to more general models where discrimination parameters are included but parameter estimation under such an extension is "complicated". It will be shown in this study how easily and efficiently parameters can be estimated by LMA models for the more general models.

The Dutch Identity and Statistical Graphical Model Connection

Another derivation of LMA models as item response models was given by Anderson and Yu (2007). Their derivation is based on Anderson and Vermunt (2000)'s LMA model as latent variable models for observed data, which use statistical graphical models for discrete and continuous variables (Lauritzen & Wermuth, 1989). They also showed that the LMA models derived in Anderson and Vermunt (2000) are formally equivalent to models in Holland (1990). For illustration, consider the following two graphs for uni- and multidimensional models presented in Figure 1.

Graph (A) represents a unidimensional model where four items are directly related to only one latent variable, and Graph (B) represents a multidimensional model where each half of four items are directly related to one of two latent variables and those two latent variables are correlated. Each item (i.e., discrete variable) is represented by a square and the latent variables by circles. If two variables are not connected by a line, those variables are independent given all the other variables in the graph. If there is a line connecting two variables, it indicates that they may be (conditionally) dependent. Since no line directly connects any two items in either Graph

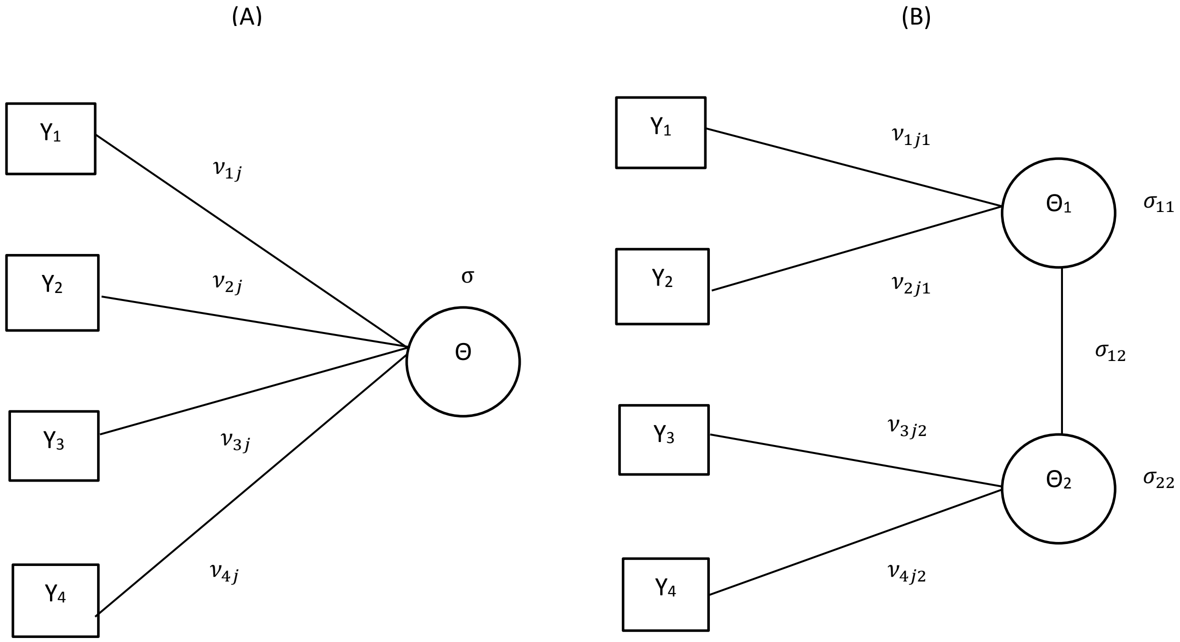


Figure 1. Graphs for (A) one latent variable and (B) two correlated latent variables with four items

(A) or (B), given the latent variables, items are conditionally independent (i.e., local independence).

Anderson and Yu (2007) showed that the assumptions made by Anderson and Vermunt (2000) were the same as those that Holland (1990) made. In addition to assuming the marginal distribution of response patterns is multinomial, there are two major assumptions for the model, which Anderson and Yu (2007) recounted from the perspective of the graphical model:

(a) The responses to items (observed variables) are conditionally independent given the latent continuous ones :

$$p(\boldsymbol{y}|\theta) = \mathrm{P}(\mathbf{Y} = \mathbf{y}|\boldsymbol{\Theta} = \theta) = \prod_{i=1}^{I} p(\boldsymbol{Y}_i = y_i|\theta).$$

(b) The joint distribution of observed and latent continuous variables is a homogenous conditional Gaussian distribution (Lauritzen & Wermuth, 1989). A homogenous conditional Gaussian distribution is the distribution of the continuous variables (i.e., θ) given the response pattern (i.e., $\boldsymbol{y}$) is normal. The mean depends on the response pattern given but the variance remains the same over response patterns :

$$\Theta|\mathbf{Y} = \boldsymbol{y} \text{ is } N\left(\mu_{\boldsymbol{y}},\ \sigma^2\right).$$

Following Anderson and Vermunt (2000) and Anderson and Böckenholt (2000), Anderson and Yu (2007) showed the joint distribution of observed and latent continuous variables, which is restated below.

$$\begin{aligned} f(\boldsymbol{y},\theta) &= f(\theta|\boldsymbol{y})\boldsymbol{P}(\boldsymbol{y}) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{\left(\theta-\mu_{(\boldsymbol{y})}\right)^2}{2\sigma^2}\right]\boldsymbol{P}(\boldsymbol{y}) \\ &= \exp\left[g(\boldsymbol{y}) + h(\boldsymbol{y})\theta - \frac{\theta^2}{2\sigma^2}\right], \end{aligned} \tag{2.8}$$

where $$g(\boldsymbol{y}) = \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) + \log\left(\boldsymbol{P}(\boldsymbol{y})\right) - \frac{\mu_{(\boldsymbol{y})}^2}{2\sigma^2}\ , \tag{2.9}$$

and $$h(\boldsymbol{y}) = \mu_{\boldsymbol{y}}/\sigma^2\ . \tag{2.10}$$

The distribution in (2.8) is a homogeneous conditional Guassian distribution.

In Anderson and Vermunt (2000), $g(\boldsymbol{y})$ represents the dependencies among discrete variables (i.e., item responses to items) given the latent variable and $h(\boldsymbol{y})$ shows the dependencies between discrete variables and the latent variable. By rewriting $g(\boldsymbol{y})$ in equation (2.9) in terms of $\log\left(\boldsymbol{P}(\boldsymbol{y})\right)$, the log manifest probabilities for response pattern are obtained, that is,

$$\log\left(\boldsymbol{P}(\boldsymbol{y})\right) = g(\boldsymbol{y}) + \log\left(\sqrt{2\pi\sigma^2}\right) + \frac{\mu_{(\boldsymbol{y})}^2}{2\sigma^2}\ . \tag{2.11}$$

To derive LMA model for log manifest probabilities for response pattern, Anderson and Vermunt (2000) specified definitions for $g(\boldsymbol{y})$ and $h(\boldsymbol{y})$. By the assumption of conditional independence of item responses given the latent variable, $g(\boldsymbol{y})$ is set equal to the sum of location parameters for each item, that is,

$$g(\boldsymbol{y}) = \sum_{i=1}^{I} \lambda_{ij} \, . \tag{2.12}$$

As presented in equation (2.8), $h(\boldsymbol{y})$ is a function of coefficients that shows the strength of the relationship between item responses to each item and the latent variable and is defined as the sum of category scores for each item, that is,

$$h(\boldsymbol{y}) = \sum_{i=1}^{I} \nu_{ij} \, . \tag{2.13}$$

By the two definitions of $h(\boldsymbol{y})$ in (2.10) and (2.13), the mean of the homogenous conditional Gaussian distribution is defined as a linear expansion of scores, which equals:

$$\mu_{\boldsymbol{y}} = \sigma^2 \sum_{i=1}^{I} \nu_{ij} \, . \tag{2.14}$$

By substituting (2.12) and (2.14) into equation (2.11), the LMA model for the log manifest probabilities of response pattern can be rewritten as

$$\log(\boldsymbol{P}(\boldsymbol{y})) = \lambda + \sum_{i} \lambda_{ij} + \sigma^2 \sum_{i} \sum_{i<k} \nu_{ij} \, \nu_{kj} \, , \tag{2.15}$$

where λ is a normalizing parameter that ensures that the ***P***(***y***) sum to 1 over response patterns, λ_{ij} are the marginal effect term for category *j* on item *i*, ν_{ij} are category scale values or scores for category *j* on item *i*, and σ^2 is an association parameter (i.e., variance of conditional distribution of θ) that shows the strength of the relationship between the items.

Anderson and Yu (2007) showed the equivalence of the LMA model using statistical graphical models to the one derived by Holland (1990) using the Dutch identity by showing that the above assumptions are equivalent to those made by Holland. When deriving the Dutch identity for the manifest probabilities of response patterns, Holland assumed that item logit functions are linear of a latent variable (θ). Anderson and Yu (2007) showed that the same assumption was also made in the LMA model by showing that item discrimination and item difficulty parameters of item logit functions can be rewritten as the difference between the corresponding parameters (i.e., ν_{ij} and λ_{ij}, respectively) of the LMA models. Another assumption made by Holland is that the posterior distribution of θ is normal with the mean and the variance for one response pattern ($\boldsymbol{y}_0$). Anderson and Yu (2007) also proved that if it is the case for $\boldsymbol{y}_0$, then it is true for all response patterns (see also Hessen, 2012). In addition to these two assumptions, conditional independence and a multinomial distribution for the manifest probabilities of response patterns are also assumed in both derivations. The equivalence of the assumptions between both approaches has further established that LMA models can function as item response models.

Fully Conditional Specification Derivation of LMA Models

Anderson and Yu (2007) provided a new derivation of LMA model as item response models based on fully conditionally specified logistic regression models using a rest-score in lieu of θ. The sum of responses to items weighted by category scores (i.e., ν_{ij}) are sufficient statistics for the latent variable (Andersen, 1995; Ostini & Nering, 2006). Adapting the idea with a slight change, Anderson and Yu (2007) (see also Anderson, Li, & Vermunt, 2007; Anderson, Verkuilen, & Peyton, 2010; Anderson, 2013) proposed to use a rest-score as an estimate of the latent variable based on the precedence and justification for it in the literature on classical test theory and IRT as mentioned in Junker and Sijtsma (2000).

A rest-score is the sum of all the item scores except for the item being studied. For example, Graph (A) in Figure 1 represents a case where all of the four items are directly related to the only one latent variable, θ (i.e., unidimentional model). When modeling the response to item 1 in Graph (A), the sum of the responses of Y_2, Y_3, and Y_4 would be used as an estimate of θ; that is,

$$\tilde{\theta}_{-1} = \sigma_{11}\left(\nu_{2j} + \nu_{3j} + \nu_{4j}\right) . \tag{2.16}$$

The symbol $\tilde{\theta}_{-1}$ represents the estimate of θ and is referred to as a rest-score for item 1. The subscript, '-1' indicates that the response of item 1 is not included in the estimate of θ. The sum, $(\nu_{2j} + \nu_{3j} + \nu_{4j})$, is over the category scores of all the other items except for the item that is being modeled (i.e., item 1) and σ_{11} is an association parameter which is the variances of θ within a response pattern.

When defining a rest-score in the case of multiple correlated latent variables, it consists of two components: (a) the one that is directly related to the latent variable and (b) the one that (indirectly) relates to the information from the correlated latent variable(s) with the target latent variable (Anderson, Li, & Vermunt, 2007; Anderson, Verkuilen, & Peyton, 2010; Anderson, 2013). The latter adds to estimation of $\tilde{\theta}_{-i}$. De la Torre and Patz (2005) found that taking into account the correlation between abilities can lead to a great improvement in ability estimation. Wang, Chen, and Cheng (2004) reported that using item responses to other tests as collateral information ca n increase measurement efficiency when the target ability is estimated.

Graph (B) in Figure 1 shows a simple multidimensional structure with four items and two correlated variables. Items 1 and 2 are directly related to θ_1, and items 3 and 4 to θ_2, and two latent variables are correlated. When specifying a model for Y_1, only the response of Y_2 would be taken as an estimate of θ_1. Since two latent variables, θ_1 and θ_2 are correlated and items 3 and 4 are direct indicators of θ_2, it would be possible to get some information about θ_1 from them. Thus, the responses of Y_3 and Y_4 would also be included in modeling θ_1. Therefore, the rest-score for item 1, namely, the estimate of θ_1 in modeling the response of Y_1 under two correlated latent variables,

$$\tilde{\theta}_{1,-1} = \sigma_{11}(\nu_{2j1}) + \sigma_{12}(\nu_{3j2} + \nu_{4j2})\,. \tag{2.17}$$

The more general form for any number of latent variables is

$$\tilde{\theta}_{m,-i} = \sigma_{mm}\sum_{k\neq i}\nu_{kjm} + \sum_{m'\neq m}\sigma_{mm'}\left(\sum_{k}\nu_{kjm'}\right), \tag{2.18}$$

where $\tilde{\theta}_{m,-i}$ indicates the estimate of θ for the latent variable *m* that excludes the response of item *i*, ν_{kjm} is the category score for the response *j* on item *k*, which is a direct indicator of the latent variable *m*, σ_{mm} is a weight (variance) that reflects the scale of the latent variable *m*, and $\sigma_{mm'}$ is a weight that shows the strength of the relationship between latent variables (i.e., covariance).

Replacing θ with its estimator $\tilde{\theta}_{m,-i}$ in equation (2.18) yields a set of fully conditional multinomial logistic regression models, one for each item. For illustration, the example model in Graph (B) in Figure 1 is used;

$$\text{Item 1}: \mathrm{P}\left(Y_1 = j \middle| \tilde{\theta}_{1,-1}\right) = \exp\{\lambda_{1j} + \nu_{1j1}(\sigma_{11}\nu_{2j1} + \sigma_{12}(\nu_{3j2} + \nu_{4j2}))\}/k_1 \tag{2.19}$$

$$\text{Item 2}: \mathrm{P}\left(Y_2 = j \middle| \tilde{\theta}_{1,-2}\right) = \exp\{\lambda_{2j} + \nu_{2j1}(\sigma_{11}\nu_{1j1} + \sigma_{12}(\nu_{3j2} + \nu_{4j2}))\}/k_2 \tag{2.20}$$

$$\text{Item 3}: \mathrm{P}\left(Y_3 = j \middle| \tilde{\theta}_{2,-3}\right) = \exp\{\lambda_{3j} + \nu_{3j2}(\sigma_{22}\nu_{4j2} + \sigma_{12}(\nu_{1j1} + \nu_{2j1}))\}/k_3 \tag{2.21}$$

$$\text{Item 4}: \mathrm{P}\left(Y_4 = j \middle| \tilde{\theta}_{2,-4}\right) = \exp\{\lambda_{4j} + \nu_{4j2}(\sigma_{22}\nu_{3j2} + \sigma_{12}(\nu_{1j1} + \nu_{2j1}))\}/k_4\ , \tag{2.22}$$

where $k_i = \sum_h^j \exp(\lambda_{ih} + \nu_{ihm}\tilde{\theta}_{m,-i})$.

For any number of items and multiple latent variables, the set of fully conditionally specified response functions is defined as

$$\mathrm{P}\left(Y_i = j \middle| \tilde{\theta}_{m,-i}\right)$$

$$= \frac{\exp\left\{\lambda_{ij} + \nu_{ijm}\left(\sigma_{mm}\sum_{k\neq i}\nu_{kjm} + \sum_{m\neq m'}\sigma_{mm'}\left(\sum_k \nu_{kjm'}\right)\right)\right\}}{\sum_h^j \exp\left\{\lambda_{ih} + \nu_{ihm}\left(\sigma_{mm}\sum_{k\neq i}\nu_{kjm} + \sum_{m\neq m'}\sigma_{mm'}\left(\sum_k \nu_{kjm'}\right)\right)\right\}} \quad \text{for } i = 1, 2, .., I. \qquad (2.23)$$

The models above are called ‘fully conditionally specified models’ because the response of each item is modeled conditional on all of the other items. To estimate the parameters of the observed response patterns of items by fitting the set of conditional logistic regression models, the joint distribution for the manifest probabilities of the response patterns, **P(*y*)** will be found (except for λ_{ij}).

A set of fully conditional specification of models (Anderson & Yu, 2007; Anderson, Li, & Vermunt, 2007; Anderson, Verkuilen, & Peyton, 2010) over-determines the joint distribution (Anderson, 2013; Gelman & Speed, 1993). Joe and Liu (1996) used the conditional specification method for multivariate binary response data with covariates and provided necessary and sufficient conditions for compatibility of conditional distributions. According to them, when one binary response variable, Y_i is conditional on Y_j $(i \neq j)$ and vice versa the two-way interaction parameters between two variables, γ_{ij} and γ_{ji} must be equal for the conditional distributions to be compatible or consistent. Extending Joe and Liu’s results to polytomous multidimensional IRT models, Anderson, Li, and Vermunt (2007), Anderson, Verkuilen, and Peyton (2010), and Anderson (2013) showed that this condition also holds for a set of fully conditionally specified models, including those that contained covariates and imposed ordinal constraints or parameters. The set of conditional models uniquely leads to an LMA model for the observed response patterns.

The LMA model for response patterns corresponding to Model (2.19) through (2.22), which corresponds to Graph (B) in Figure 1 is

$$\begin{aligned}\log P(y) = {} & \lambda + \lambda_{1j} + \lambda_{2j} + \lambda_{3j} + \lambda_{4j} \\ & + \sigma_{11}\nu_{1j1}\nu_{2j1} + \sigma_{22}\nu_{3j2}\nu_{4j2} \\ & + \sigma_{12}\left(\nu_{1j1}\nu_{3j2} + \nu_{1j1}\nu_{4j2} + \nu_{2j1}\nu_{3j2} + \nu_{2j1}\nu_{4j2}\right).\end{aligned} \tag{2.24}$$

The more general form of LMA model for the joint distribution is;

$$\log P(\boldsymbol{y}) = \lambda + \sum_{i} \lambda_{ij} + \sum_{i}\sum_{i>k}\sum_{m}\sum_{m\geq m'} \sigma_{mm'}\nu_{ijm}\,\nu_{kjm'} \ , \tag{2.25}$$

where $\boldsymbol{y}$ is the response pattern on *I* items, λ ensures that the sum of the probabilities over all the possible response patterns equals 1, λ_{ij} represents the marginal effect terms of each category on item *i*, and $\sigma_{mm'}\nu_{ijm}\nu_{kjm'}$ is the multiplicative term of category scores between pairs of items with an association parameter. In the case where item k is not directly related the latent variable m, the corresponding ν_{kjm} will be set to zero. For example, since there is no relationship between item 2 and θ_2 in Graph (B), its category scores for the latent variable θ_2, ν_{2j2} equals zero.

As mentioned earlier, LMA models are special cases of log-linear models (a Possion regression) with only two-way interaction terms. Multinomial logistic regression models with categorical variables for predictors can be written in the form of log-linear models (Agresti, 2002). Therefore, each parameter in LMA model given in equation (2.24) is the same as those in equations (2.19) through (2.22) for multinomial logistic models. That is, the marginal effect terms, λ_{ij} correspond to the location parameters in the multinomial logistic regression models, and the product terms of weights ($\sigma_{mm'}$) and category scores (ν_{ijm}) also appear in both equations.

There are two great advantages in the LMA model as item response model derived by Anderson, Verkuilen, and Peyton (2007). First, since LMA models are equivalent to Possion regression models, not only can the parameters be estimated without using numerical integration, but also explicit assumptions regarding the marginal distribution of the latent variables are not necessary. The LMA models allow IRT models with multiple latent variables to be fit to data using MLE by using computer software such as ℓEM (Vermunt, 1997) and PROC/NLP in SAS.

A second of advantage of the fully conditional specification approach is that it suggests how parameters for large numbers of items could be estimated in an efficient way that overcomes the limitations of MLE of LMA models. The parameters of LMA models are estimated by MLE and it requires iteratively computing fitted values for all possible response patterns. ML estimates of LMA models for small numbers of items can be obtained easily, but the standard estimation methods for large numbers of items fail because the number of possible response patterns increases exponentially as the number of items and response options per item increase. Rather than MLE, pseudo-likelihood estimation can be done and will be discussed in the following chapter.

Chapter 3

Estimation of LMA Models

Log-linear and LMA models are typically estimated by maximum likelihood estimation (MLE). This chapter provides an overview of the estimation procedures of the parameters in LMA models. Two estimation procedures for LMA models as MIRT models will be introduced, maximum likelihood estimation and pseudo-likelihood estimation, followed by its implementation for LMA Rasch models (i.e., linear-by-linear models). The more general algorithm will be presented in the Chapter 4.

Maximum Likelihood Estimation (MLE)

Let C denote the number of possible item response patterns $\boldsymbol{y}$, $P(\boldsymbol{y}_c)$ denote their probabilities, where $\sum_c P(\boldsymbol{y}_c) = 1$, and $n(\boldsymbol{y}_c)$ denote the number of examinees in the sample having a response pattern $\boldsymbol{y}_c$, where $\sum_c n(\boldsymbol{y}_c) = N$, total number of examinees in the sample. Then, $\{n(\boldsymbol{y}_c)\}$ follows a multinomial distribution with parameters N and $P(\boldsymbol{y}_c)$ as below:

$$\begin{aligned} &P[n(\boldsymbol{y}_1), n(\boldsymbol{y}_2), n(\boldsymbol{y}_3), \dots, n(\boldsymbol{y}_c)] \\ &= \left(\frac{N!}{n(\boldsymbol{y}_1)!\, n(\boldsymbol{y}_2)!\, n(\boldsymbol{y}_3)! \dots n(\boldsymbol{y}_c)!}\right) P(\boldsymbol{y}_1)^{n(\boldsymbol{y}_1)} P(\boldsymbol{y}_2)^{n(\boldsymbol{y}_2)} P(\boldsymbol{y}_3)^{n(\boldsymbol{y}_3)} \dots P(\boldsymbol{y}_c)^{n(\boldsymbol{y}_c)} \ . \end{aligned} \tag{3.1}$$

Removing a multiplicative constant, the kernel of the likelihood function for multinomial count data of item response patterns equals

$$L = \prod_{c=1}^{C} \frac{\exp(-P(\boldsymbol{y}_c)) P(\boldsymbol{y}_c)^{n(\boldsymbol{y}_c)}}{n(\boldsymbol{y}_c)!}, \tag{3.2}$$

and the kernel of the log-likelihood function is

$$\log L = \sum_{c=1}^{C} n(\boldsymbol{y}_c) \log P(\boldsymbol{y}_c) - P(\boldsymbol{y}_c)\,. \tag{3.3}$$

Holland (1990) showed that fitting IRT models by MMLE based on an item response matrix which presents all the responses to *I* items of all examinees in the sample (i.e., *N*) is equivalent to fitting a second-order log-linear model for multinomial count data by MLE (Holland, 1990) and this goes the same for LMA models.

Computer Programs for Parameter Estimation in LMA Models

There are a number of computer programs for fitting LMA models and these include *ℓ*EM (Vermunt, 1997), SAS nonlinear programming procedure (PROC/NLP), R, and MatLab. Of them, *ℓ*EM and PROC/NLP in SAS are the most frequently used for fitting LMA models. *ℓ*EM is open-source software for the analysis of categorical data developed by Vermunt (1997). It conducts parameter estimation of LMA models by maximum likelihood using a quasi-Newton algorithm. Uni-dimensional Newton-Raphson algorithm is a variant of Newton-Raphson algorithm that only uses the diagonal elements of the Hessian matrix to update equations for the parameters. Global optimal solutions are not guaranteed but multiple runs with random starts can be used to check convergence. For details, see Anderson and Vermunt (2000). PROC/NLP in SAS provides a variety of ways for estimating parameters in nonlinear statistical models. Both unconstrained and constrained maximization/minimization problems can be handled by the procedure with a set of optimization methods, including Newton-Raphson and quasi-Newton method. The latter is used when non-linear constraints are placed on parameters, otherwise Newton-Raphson can be used, which does give a unique global maximum. LMA models can be fit to data easily with three command statements in PROC/NLP. They are: (a) ‘parms’ statement that specifies the parameters to be estimated, (b) the equation of an LMA model, and (c) the logarithm of the likelihood function to be maximized. With Newton-Raphson, it yields estimates of standard errors and covariance matrices for parameter estimates. These features of PROC/NLP provide an easy and flexible way of fitting LMA models.

Motivation of PLE for LMA models

When estimating parameters in LMA models with MLE, it requires iteratively computing fitted values for all possible response patterns. Parameter estimates of LMA models for small numbers of items can be obtained by MLE easily because the number of all possible response patterns is reasonable. For large numbers of items, however, the standard estimation methods of LMA models fail because the number of possible response patterns increases exponentially as the number of items and response options per item increase. For example, Anderson (2013) reported that the estimation of an LMA model for 8 items with 5 categories using PROC/NLP was successful, but failed for 9 items with the same number of categories. The number of possible item response patterns for 8 items equals $5^8 = 390{,}625$ and it increases to $5^9 = 1{,}953{,}125$ when just one item is added to 8 items, which makes the MLE of LMA models infeasible.

More recently, pseudo-likelihood estimation (PLE) was proposed to solve the problem for Rasch models with large numbers of items (Anderson, Li, & Vermunt 2007; Li 2010). It is reported that PLE can fit the models to data with large numbers of items successfully and fast, handle models with multiple latent variables and covariates, yield consistent estimates, and is easy to implement.

Pseudo-likelihood Estimation (PLE)

Pseudo-likelihood estimation (PLE) was first introduced by Besag (1974) as an approach to the specification and estimation of spatial interaction models. He pointed out that a complicated normalizing function hinders a direct approach to statistical inference through maximum likelihood. To solve this problem, he specified a conditional distribution for spatial data observed at a specific site given the data observed at all the other sites and obtained the parameter estimates by maximizing the product of the conditional likelihood functions for data. As shown in his study, the basic idea behind PLE is to replace numerically challenging problems with more tractable ones by simplifying them with conditional specification approach so that computational demands can decrease in fitting models. For its computational efficiency, PLE has been used as an alternative to maximum likelihood estimation in a number of studies on social networks (Strauss & Ikeda, 1990; Wasserman & Pattison, 1990), multivariate clustered data (Geys, Molenberghs, & Ryan, 1999; Molenberghs & Verbeke, 2005), longitudinal data (Troxel, Lipsitz, & Harrington, 1998; Parzen, Lipsitz, Fitzmaurice, Ibrahim, Troxel, & Molenberghs, 2007), incomplete data (Molenberghs, Kenward, Verbeke, & Birhanu, 2011).

Several studies have been performed to estimate parameters of Rasch models with PLE (Arnold & Strauss, 1991; Strauss, 1992; Zwinderman, 1995; Smit & Kelderman, 2000). Following Besag (1974, 1975)'s conditional approach, Arnold and Strauss (1991) and Strauss (1992) provided the definition of pseudo-likelihood for pairs of binary items using Rasch models and mentioned that maximizing the pseudo-likelihood function is equivalent to finding MLE with a logistic regression procedure.

Zwinderman (1995) conducted a simulation study to investigate the consistency and efficiency of PL estimates for Rasch models using responses to pairs of items, irrespective of other items. He compared the estimates from PLE to those from conditional maximum likelihood (CML) and marginal maximum likelihood (MML) estimation methods and showed that PL estimates are consistent and similar in efficiency to CML and MML estimates.

Smit and Kelderman (2000) proposed to apply PLE method to Rasch models for binary items. Unlike the pairwise estimation method used by Arnold and Strauss (1991), Strauss (1992), and Zwinderman (1995), they estimated the parameters for Rasch models based on a set of binary item responses and showed that PLE method is more computationally attractive than CML and its estimates are almost identical to CML and unconditional ML estimates.

To summarize, the studies have shown that PLE can be used for Rasch models as an alternative estimation method to MLE, but their application of PLE was limited to unidimensional binary Rasch models.

PLE for LMA Rasch Models Using Fully Conditionally Specified Models

More recently, Anderson, Li, and Vermunt (2007) proposed PLE for LMA models to solve the problem of MLE for large numbers of items. Like the previous application of PLE to Rasch model, their methodology was also applied to LMA Rasch models, but its application was extended to the models for polytomous items and multiple latent variables. Their implementation of PLE of LMA Rasch models reflects the original idea of PLE introduced by Besag (1974, 1975). As mentioned earlier, fitting LMA models for large numbers of items with MLE is prohibitive due to the exponential increase in the number of all possible item response patterns. To solve this complex problem, they replaced LMA Rasch models for large numbers of items with simpler conditional multinomial logistic models based on fully conditional specification approach (Anderson & Yu, 2007; Anderson, Li, & Vermunt, 2007; Anderson, Verkuilen, & Peyton, 2010). They specified conditional models corresponding to each item using rest score in lieu of the latent variable and defined pseudo-likelihood as the product of the likelihoods of the conditional multinomial logistic regressions. For implementation of PLE, the whole set of fully conditionally specified logistic regression models were “stacked” into a large design matrix and the model parameters were estimated by finding the maximum value of likelihood function with one large stacked conditional multinomial logistic regression model. The maximum value of the likelihood of the model fit to the stacked data equals the maximum of pseudo-likelihood function.

For illustration, let’s take an example model for a LMA Rasch model with *I* polytomous items and one latent variable. The graph for the model looks similar to Graph (A) in Figure 1 represented in the previous chapter.

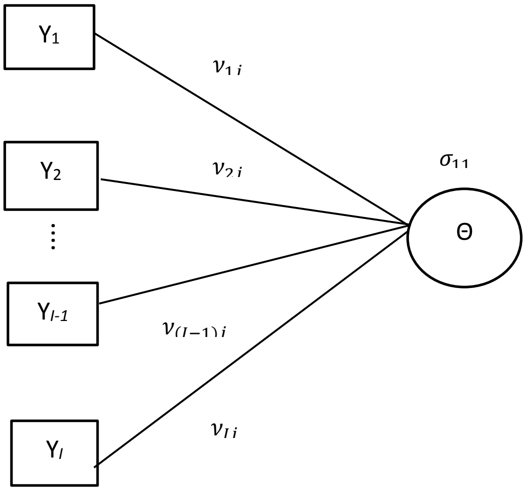


Figure 2. LMA Rasch model with *I* polytomous items and one latent variable

In Anderson, Verkuilen, and Peyton (2010)’s fully conditional specification approach, the rest score for item *i* is defined as

$$\tilde{\theta}_{-i} = \sigma_{11} \sum_{k \neq i} \nu_{kj} \,. \tag{3.4}$$

where $\tilde{\theta}_{-i}$ indicates the estimate of θ (e.g., a rest score) for item *i*, ν_{kj} is the known (or assumed) category score for the response *j* on item *k*, and σ_{11} is a weight (variance) that reflects the scale of the latent variable within response patterns.

Using the defined weighted rest score in equation (3.4), the LMA Rasch model in Figure 2 yields a set of fully conditionally specified multinomial logistic regression models:

$$\mathrm{P}\left(Y_1 = j \middle| \tilde{\theta}_{-1}\right) = \exp\{\lambda_{1j} + \nu_{1j}(\sigma_{11} \sum_{k\neq 1} \nu_{kj})\}/k_1 \tag{3.5}$$

$$\mathrm{P}\left(Y_2 = j \middle| \tilde{\theta}_{-2}\right) = \exp\{\lambda_{2j} + \nu_{2j}(\sigma_{11} \sum_{k\neq 2} \nu_{kj})\}/k_2 \tag{3.6}$$

$$\vdots \qquad = \qquad \vdots$$

$$\mathrm{P}\left(Y_I = j \middle| \tilde{\theta}_{-I}\right) = \exp\{\lambda_{Ij} + \nu_{Ij}(\sigma_{11} \sum_{k\neq I} \nu_{kj})\}/k_I \ , \tag{3.7}$$

where $k_i = \sum_{h=0}^{j} \exp\{\lambda_{ih} + \nu_{ih}(\sigma_{11} \sum_{k\neq I} \nu_{kj})\}$.

As defined earlier, pseudo-likelihood is the product of the likelihoods of the conditional multinomial logistic regressions. Therefore, pseudo-likelihood for person *n* is defined as

$$\begin{aligned} PL\ (\boldsymbol{\lambda}|\boldsymbol{y}_n) &= \prod_{i=1}^{I} \mathrm{P}\left(Y_{ni} = j \middle| \tilde{\theta}_{n(-i)}\right) \\ &= \prod_{i=1}^{I} \frac{\exp\{\lambda_{ij} + \nu_{ij}(\sigma_{11} \sum_{k\neq i} \nu_{kj})\}}{\sum_{h=0}^{j} \exp\{\lambda_{ih} + \nu_{ih}(\sigma_{11} \sum_{k\neq i} \nu_{kj})\}} \ , \end{aligned} \tag{3.8}$$

where $\boldsymbol{\lambda}$ is the vector of parameters in the model and $\boldsymbol{y}_n$ is a response pattern to *I* items of person *n*.

Assuming that each person is independent, pseudo-likelihood for all persons in *N* is expressed as

$$\begin{aligned} PL\ (\boldsymbol{\lambda}|\boldsymbol{y}_1, \boldsymbol{y}_2, \dots \boldsymbol{y}_N) &= \prod_{n=1}^{N} \prod_{i=1}^{I} \mathrm{P}\left(Y_{ni} = j \middle| \tilde{\theta}_{n(-i)}\right) \\ &= \prod_{n=1}^{N} \prod_{i=1}^{I} \frac{\exp\{\lambda_{ij} + \nu_{ij}(\sigma_{11} \sum_{k\neq i} \nu_{kj})\}}{\sum_{h=0}^{j} \exp\{\lambda_{ih} + \nu_{ih}(\sigma_{11} \sum_{k\neq i} \nu_{kj})\}} \ . \end{aligned} \tag{3.9}$$

By taking logarithms on both sides of equation (3.9), log pseudo-likelihood for the whole response patterns of all individuals can be expressed as

$$\log PL = \log\ (PL\ (\boldsymbol{\lambda}|\boldsymbol{y}_1, \boldsymbol{y}_2, \dots \boldsymbol{y}_N))$$

$$= \sum_{n=1}^{N}\sum_{i=1}^{I} \mathrm{P}\big(Y_{ni} = j\big|\tilde{\theta}_{n(-i)}\big)\ . \tag{3.10}$$

Pseudo-likelihood estimates are the values for $\boldsymbol{\lambda}$ and σ_{11} that maximize the pseudo-likelihood function in equation (3.9) and this can be done by maximizing the likelihood of the set of conditional multinomial logistic regression models, (3.5) ~ (3.7). The one large conditional multinomial logistic regression model can be set up by stacking each conditional multinomial logistic regression model in a design matrix and the model fit by MLE. The value of the maximum of the likelihood of the conditional multinomial logistic regression is the value of the PLE.

Table 1 illustrates the design matrix for fitting the set of fully specified conditional multinomial logistic regression models to get pseudo-likelihood estimates of person *1*. For convenience, assume that all items have the same number of response categories. The design matrix consists of $I \times (J + 1)$ rows and $(I \times J) + 4$ columns for one person. Let's suppose that person *1* has the response pattern to *I* polytomous items, $\boldsymbol{y}_1 = (2\ 1 \dots \dots J)$. The first column shows person ID for each person. The next two columns show the item number and the response options for each item, respectively. The fourth column shows the response variable, Y indicating 1 for the selected response option by person *1* and 0 for otherwise. For example, the third response option of item 1 is entered as 1 and the rest of response options as 0 because person *1* selected the third response option for item 1. The next $I \times J$ columns show dummy codes[4] for location parameters of each item, λ_{ij} and λ_{ij} is set to $\lambda_{ij} = 0$ for identification. In each column

---

[4] Other coding schemes can be used.

Table 1. Design matrix for one person for PLE of LMA Rasch model: Polytomous items with one latent variable

Person	Item	Response option	Response variable (Y)	item 1				item 2				…	…	item *I*				Weighted rest-score
				λ_{11}	λ_{12}	…	λ_{1J}	λ_{21}	λ_{22}	…	λ_{2J}	…	…	λ_{I1}	λ_{I2}	…	λ_{IJ}	σ_{11}
1	1	0	0	0	0	…	0	0	0	…	0	…	…	0	0	…	0	0
1	1	1	0	1	0	…	0	0	0	…	0	…	…	0	0	…	0	$\nu_{11}\sum_{k\neq i}\nu_{kj}$
1	1	2	1	0	1	…	0	0	0	…	0	…	…	0	0	…	0	$\nu_{12}\sum_{k\neq i}\nu_{kj}$
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
1	1	*J*	0	0	0	…	1	0	0	…	0	…	…	0	0	…	0	$\nu_{1J}\sum_{k\neq i}\nu_{kj}$
1	2	0	0	0	0	…	0	0	0	…	0	…	…	0	0	…	0	0
1	2	1	1	0	0	…	0	1	0	…	0	…	…	0	0	…	0	$\nu_{21}\sum_{k\neq i}\nu_{kj}$
1	2	2	0	0	0	…	0	0	1	…	0	…	…	0	0	…	0	$\nu_{22}\sum_{k\neq i}\nu_{kj}$
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
1	2	*J*	0	0	0	…	0	0	0	…	1	…	…	0	0	…	0	$\nu_{2J}\sum_{k\neq i}\nu_{kj}$
1	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
1	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
1	*I*	0	0	0	0	…	0	0	0	…	0	…	…	0	0	…	0	0
1	*I*	1	0	0	0	…	0	0	0	…	0	…	…	1	0	…	0	$\nu_{I1}\sum_{k\neq i}\nu_{kj}$
1	*I*	2	0	0	0	…	0	0	0	…	0	…	…	0	1	…	0	$\nu_{I2}\sum_{k\neq i}\nu_{kj}$
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
1	*I*	*J*	1	0	0	…	0	0	0	…	0	…	…	0	0	…	1	$\nu_{IJ}\sum_{k\neq i}\nu_{kj}$

for λ_{ij}, only one element is represented as 1 and all the other elements as 0. The last column shows a rest score weighted by the ν_{ij} for the corresponding item and response option. In Rasch model, λ_{i0} and ν_{i0} are generally set to $\lambda_{i0} = 0$ and $\nu_{i0} = 0$. For this reason, all the elements in the first row of each item equal 0. For total sample size of *N*, there will be $N \times (I \times J)$ rows in the design matrix.

Anderson, Li, and Vermunt (2007) examined the performance of PLE of Rasch models under various situations in the simulation studies, varying the number of items, response options (e.g., binary or polytomous), latent variables (e.g., single or multiple latent variables), and examinees. For small numbers of binary or polytomous items with a single or multiple latent variables, PL estimates were highly correlated with ML estimates for different sample sizes and yielded very similar estimates to true parameters used to simulate the data. The robust or "sandwich" standard errors that were computed were also very similar to those from MLE. The parameter recovery of PLE was excellent for large numbers of binary or polytomous items with a single or multiple latent variables. For large numbers of binary items with a single latent variable, standard errors from PLE and BILOG were compared and it showed that the robust standard errors from PLE were slightly smaller than the standard errors from BILOG. This methodology was further investigated for LMA Rasch models with covariates by Li (2010) and it was shown that the parameters for LMA Rasch models with covariates can be estimated by PLE easily and quickly with a small loss of efficiency relative to MLE.

Motivation of the Extension of PLE to More General Models

As reviewed so far, the use of current PLE is limited to the estimation of location parameters (i.e., models in the Rasch family). In this thesis, PLE will be proposed and developed for more general models including slope (discrimination) parameters.

Chapter 4

Proposed Algorithm for LMA Models

A new estimation algorithm for more general LMA models with both location and slope parameters is introduced. After presenting a general summary of the algorithm, a detailed description of each step involved in the algorithm will be given.

An Overview of the New Algorithm

The previous PLE algorithm for LMA Rasch models (Anderson, Li, & Vermunt, 2007) is extended by adding an additional step that estimates the slope parameters for the latent variables as well as location parameters. The algorithm has two basic steps and simplifies for special cases. In Step 1, a conditional multinomial logistic regression model is fit by MLE to one item using rest-scores as an explanatory variable to get new estimates of slope parameters for the item. Subsequently, the estimated slope parameter of the item is used in the rest-score for the model for the next item. This process is repeated for each item until all slope parameters have been updated. Step 2 involves fitting a single conditional logistic regression model to a data set formed by "stacking" the conditional logistic regressions for each item. This yields new estimates of location parameters and the covariance matrix for the latent variables. Steps 1 and 2 are repeated until all parameter estimates converge.

At convergence, additional iterations result in the situation where (a) log-likelihoods of conditional logit models for items do not change, (b) log-likelihood of a stacked regression does not change, (c) parameter estimates do not change, and (d) location parameters (λ_{ij}) from conditional logit models for each item and from a stacked regression model are identical.

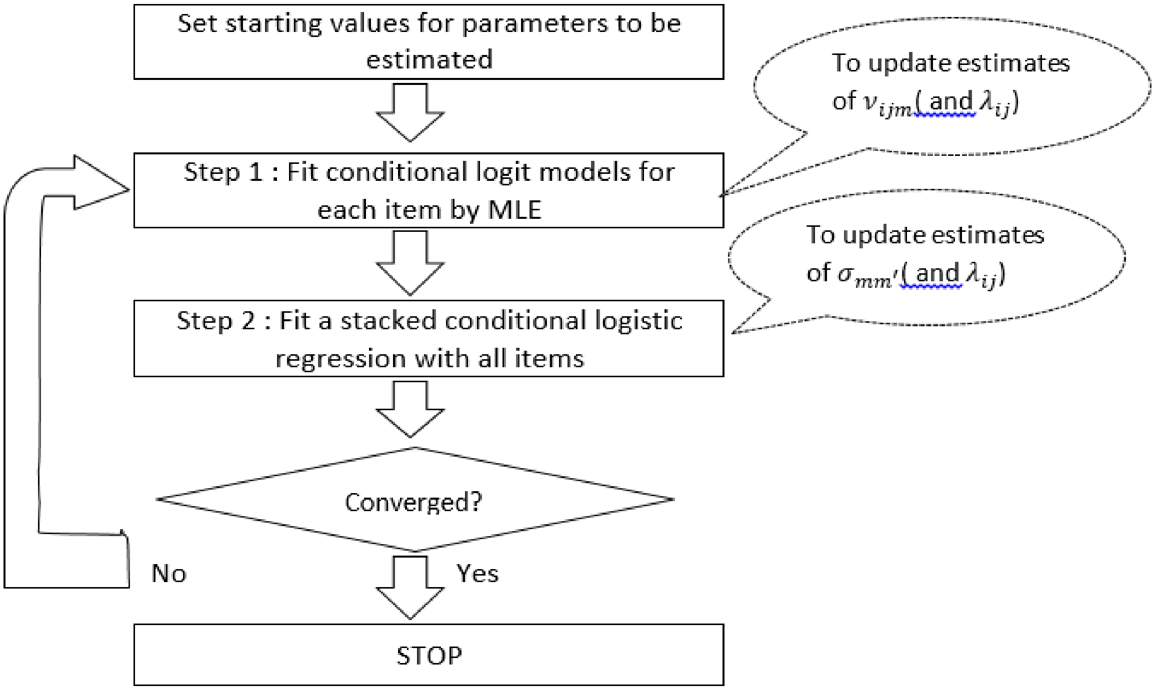


Figure 3. The proposed algorithm

Figure 3 illustrates the algorithm with the parameters to be estimated in each step. In the following section, a detailed description about each step will be given. For illustration, the example model where two latent variables are correlated and two of four items are directly related to one of them and a set of fully conditionally specified models will be used. The example model is illustrated in Figure 1 (B) on page 16.

Step 1: Conditional Multinomial Logit Models for Each Item

The main goal of Step 1 is to estimate (or update) slope (discrimination) parameters for each item. The goal can be achieved by fitting a conditional multinomial logistic regression model for each item with MLE using a weighted rest-score as an explanatory variable.

Estimation Process of Step 1

In Step 1, a set of conditional logistic regression models is fit to data for one item at a time so that the estimated ν_{ijm} can be included in the weighted rest-score for next item and used as an explanatory variable to estimate slope parameters for the item. The location parameters, λ_{ij}'s are also obtained by fitting the models, but the main interest of Step 1 lies on slope parameters. Location parameters are also estimated in Step 2. Let's suppose that the conditional multinomial logistic regression model for item 1 is fit to data. The parameters to be estimated are denoted by $\boxed{\boldsymbol{\lambda}_{\mathbf{1}\boldsymbol{j}}}$ and $\boxed{\boldsymbol{\nu}_{\mathbf{1}\boldsymbol{j}\mathbf{1}}}$ as shown below.

$$\mathrm{P}\left(Y_1 = j \middle| \tilde{\theta}_{1,-1}\right) = \frac{\exp\{\boxed{\boldsymbol{\lambda}_{\mathbf{1}\boldsymbol{j}}} + \boxed{\boldsymbol{\nu}_{\mathbf{1}\boldsymbol{j}\mathbf{1}}}(\sigma_{11}\nu_{2j1} + \sigma_{12}(\nu_{3j2} + \nu_{4j2}))\}}{\sum_h^j \exp(\lambda_{1h} + \nu_{1h1}\tilde{\theta}_{1,-i})} \;.$$

After fitting the model, the MLE of $\boldsymbol{\nu}_{\mathbf{1}\boldsymbol{j}\mathbf{1}}$ is used in the model for item 2; namely,

$$\mathrm{P}\left(Y_2 = j \middle| \tilde{\theta}_{1,-2}\right) = \frac{\exp\{\boxed{\boldsymbol{\lambda}_{\mathbf{2}\boldsymbol{j}}} + \boxed{\boldsymbol{\nu}_{\mathbf{2}\boldsymbol{j}\mathbf{1}}}(\sigma_{11}\boldsymbol{\nu}_{\mathbf{1}\boldsymbol{j}\mathbf{1}} + \sigma_{12}(\nu_{3j2} + \nu_{4j2}))\}}{\sum_h^j \exp(\lambda_{2h} + \nu_{2h1}\tilde{\theta}_{1,-i})} \;.$$

The process continues over all the items in the same way until all of ν_{ijm}s are estimated, including λ_{ij}s. With all 4 items in the example model, the whole process of Step 1 is illustrated as follows:

$$\begin{gathered}
\mathrm{P}\left(Y_1 = j \middle| \tilde{\theta}_{1,-1}\right) = \exp\{\boxed{\boldsymbol{\lambda}_{\mathbf{1}\boldsymbol{j}}} + \boxed{\boldsymbol{\nu}_{\mathbf{1}\boldsymbol{j}\mathbf{1}}}(\sigma_{11}\nu_{2j1} + \sigma_{12}(\nu_{3j2} + \nu_{4j2}))\}/k_1 \\
\downarrow \\
\mathrm{P}\left(Y_2 = j \middle| \tilde{\theta}_{1,-2}\right) = \exp\{\boxed{\boldsymbol{\lambda}_{\mathbf{2}\boldsymbol{j}}} + \boxed{\boldsymbol{\nu}_{\mathbf{2}\boldsymbol{j}\mathbf{1}}}(\sigma_{11}\boldsymbol{\nu}_{\mathbf{1}\boldsymbol{j}\mathbf{1}} + \sigma_{12}(\nu_{3j2} + \nu_{4j2}))\}/k_2 \\
\downarrow \\
\mathrm{P}\left(Y_3 = j \middle| \tilde{\theta}_{2,-3}\right) = \exp\{\boxed{\boldsymbol{\lambda}_{\mathbf{3}\boldsymbol{j}}} + \boxed{\boldsymbol{\nu}_{\mathbf{3}\boldsymbol{j}\mathbf{2}}}(\sigma_{22}\nu_{4j2} + \sigma_{12}(\boldsymbol{\nu}_{\mathbf{1}\boldsymbol{j}\mathbf{1}} + \boldsymbol{\nu}_{\mathbf{2}\boldsymbol{j}\mathbf{1}}))\}/k_3 \\
\downarrow \\
\mathrm{P}\left(Y_4 = j \middle| \tilde{\theta}_{2,-4}\right) = \exp\{\boxed{\boldsymbol{\lambda}_{\mathbf{4}\boldsymbol{j}}} + \boxed{\boldsymbol{\nu}_{\mathbf{4}\boldsymbol{j}\mathbf{2}}}(\sigma_{22}\boldsymbol{\nu}_{\mathbf{3}\boldsymbol{j}\mathbf{2}} + \sigma_{12}(\boldsymbol{\nu}_{\mathbf{1}\boldsymbol{j}\mathbf{1}} + \boldsymbol{\nu}_{\mathbf{2}\boldsymbol{j}\mathbf{1}}))\}/k_4
\end{gathered}$$

where $k_i = \sum_h^j \exp(\lambda_{ih} + \nu_{ihm}\tilde{\theta}_{m,-i})$, $\boxed{\boldsymbol{\lambda}_{\boldsymbol{ij}}}$ and $\boxed{\boldsymbol{\nu}_{\boldsymbol{ijm}}}$ are location and slope parameters to be estimated for item *i*, and $\boldsymbol{\nu}_{\boldsymbol{ijm}}$ is the estimated slope parameter for the response *j* on item *i*, which is a direct indicator of the latent variable *m*.

There are two great things about the estimation of parameters by Step 1. First, it is not sensitive to starting values for slope parameters of LMA models. Second, parameter estimates converge as long as identification constraints of LMA models are imposed. This can be setting $\sigma_{mm} = 1$ or $\sum_j \nu^2{}_{ijm} = 1$ for one item that is directly related to latent variable m. In this thesis, the first one was chosen to simplify computations.

In short, Step 1 involves an iterative process of fitting conditional multinomial logistic regression models for each item to estimate location and slope parameters in the models. The most currently obtained estimates of the slope parameters for previous items are used as explanatory variables for the next item. Once all the slope parameters of all items have been estimated through the repeated process, the algorithm is moving on to Step 2 of a stacked conditional logistic regression to estimate association and location parameters.

Step 2: A Stacked Conditional Logistic Regression Model

The goal of Step 2 is to estimate location (λ_{ij}) and association parameters ($\sigma_{mm'}$) for each item. To achieve the goal, the whole set of conditional logistic regressions for each item that has been fit in Step 1 is stacked and fit as a single "stacked" conditional logistic regression. Step 2 involves the maximization of pseudo-likelihood function by maximizing the likelihood of one large conditional logistic regression with a stacked data and yields new estimates of location and association parameters for latent variables.

Estimation Process of Step 2

One thing that should be noticed is that the definition of a weighted rest-score in Step 2 is different from that in Step 1. Let's take a multinomial conditional logistic regression model for item 1 as an example.

$$\text{Item 1}: \ \mathrm{P}\left(Y_1 = j \middle| \tilde{\theta}_{1,-1}\right) = \frac{\exp\{\lambda_{1j} + \nu_{1j1}(\sigma_{11}\nu_{2j1} + \sigma_{12}(\nu_{3j2} + \nu_{4j2}))\}}{\sum_h^j \exp(\lambda_{1h} + \nu_{1h1}\tilde{\theta}_{1,-i})}.$$

Depending on the step of the algorithm, the conditional logistic model for item 1 takes the different regression coefficients and the weighted rest-score for them. In step 1, the model can be viewed as a regression coefficients, ν_{1j1} for the weighted rest-score, $\sigma_{11}\nu_{2j1} + \sigma_{12}(\nu_{3j2} + \nu_{4j2})$. However, in Step 2, σ_{11} is a regression coefficient of (weighted) rest-score, $\nu_{1j1}\nu_{2j1}$ and σ_{12} is a regression coefficient of (weighted) rest-score, $\nu_{1j1}(\nu_{3j2} + \nu_{4j2})$. Table 3 shows the regression coefficients for the weighted rest-score in each step.

Table 2. Definitions of the weighted rest-score for item 1 in each step

	Parameters to be estimated (regression coefficients)	Weighted rest-score ($\tilde{\theta}_{m,-i}$) (explanatory variables)
Step 1	ν_{ijm}	$\sigma_{mm}\sum_{k\neq i}\nu_{kjm} + \sum_{m'\neq m}\sigma_{mm'}\left(\sum_{k\neq i}\nu_{kjm'}\right)$
Step 2	σ_{mm}	$\nu_{ijm}\sum_{k\neq i}\nu_{kjm}$
	$\sigma_{mm'}$	$\nu_{ijm}\sum_{k\neq i}\nu_{kjm'}$

In Table 2, $\tilde{\theta}_{m,-i}$ indicates the estimate of θ for the latent variable *m* that excludes the response of item *i*, ν_{kjm} is the category score for the response *j* on item *k (k ≠ i)*, which is a direct indicator of the latent variable *m*, σ_{mm} is a weight (variance) that reflects the scale of the latent variable *m*, and $\sigma_{mm'}$ is a weight that shows the strength of the relationship between latent variables (i.e., covariance).

Continuing with the 4 items in the example model, Step 2 starts with the whole set of conditional logistic regression models for each item that has already fit in Step 1 as follows.

$$\mathrm{P}\left(Y_1 = j \middle| \tilde{\theta}_{1,-1}\right) = \exp\{\lambda_{1j} + \boldsymbol{\nu_{1j1}}(\sigma_{11}\boldsymbol{\nu_{2j1}} + \sigma_{12}(\boldsymbol{\nu_{3j2}} + \boldsymbol{\nu_{4j2}}))\}/k_1$$

$$\mathrm{P}\left(Y_2 = j \middle| \tilde{\theta}_{1,-2}\right) = \exp\{\lambda_{2j} + \boldsymbol{\nu_{2j1}}(\sigma_{11}\boldsymbol{\nu_{1j1}} + \sigma_{12}(\boldsymbol{\nu_{3j2}} + \boldsymbol{\nu_{4j2}}))\}/k_2$$

$$\mathrm{P}\left(Y_3 = j \middle| \tilde{\theta}_{2,-3}\right) = \exp\{\lambda_{3j} + \boldsymbol{\nu_{3j2}}(\sigma_{22}\boldsymbol{\nu_{4j2}} + \sigma_{12}(\boldsymbol{\nu_{1j1}} + \boldsymbol{\nu_{2j1}}))\}/k_3$$

$$\mathrm{P}\left(Y_4 = j \middle| \tilde{\theta}_{2,-4}\right) = \exp\{\lambda_{4j} + \boldsymbol{\nu_{4j2}}(\sigma_{22}\boldsymbol{\nu_{3j2}} + \sigma_{12}(\boldsymbol{\nu_{1j1}} + \boldsymbol{\nu_{2j1}}))\}/k_4$$

where $k_i = \sum_h^j \exp(\lambda_{ih} + \nu_{ihm}\tilde{\theta}_{m,-i})$.

The slope parameters highlighted in bold font (i.e., $\boldsymbol{\nu_{ijm}}$) represent that they are the estimates obtained in Step 1. To estimate location and association parameters in the models, the set of conditional logistic models is appropriately formatted for a stacked conditional logistic regression.

Fitting a single conditional logistic regression model with a stacked data set yields the PL estimates for location and association parameters and they are highlighted as below.

$$\mathrm{P}\left(Y_1 = j \middle| \tilde{\theta}_{1,-1}\right) = \exp\{\boldsymbol{\lambda_{1j}} + \boldsymbol{\nu_{1j1}}(\boldsymbol{\sigma_{11}}\boldsymbol{\nu_{2j1}} + \boldsymbol{\sigma_{12}}(\boldsymbol{\nu_{3j2}} + \boldsymbol{\nu_{4j2}}))\}/k_1$$

$$\mathrm{P}\left(Y_2 = j \middle| \tilde{\theta}_{1,-2}\right) = \exp\{\boldsymbol{\lambda_{2j}} + \boldsymbol{\nu_{2j1}}(\boldsymbol{\sigma_{11}}\boldsymbol{\nu_{1j1}} + \boldsymbol{\sigma_{12}}(\boldsymbol{\nu_{3j2}} + \boldsymbol{\nu_{4j2}}))\}/k_2$$

$$\mathrm{P}\left(Y_3 = j \middle| \tilde{\theta}_{2,-3}\right) = \exp\{\boldsymbol{\lambda_{3j}} + \boldsymbol{\nu_{3j2}}(\boldsymbol{\sigma_{22}}\boldsymbol{\nu_{4j2}} + \boldsymbol{\sigma_{12}}(\boldsymbol{\nu_{1j1}} + \boldsymbol{\nu_{2j1}}))\}/k_3$$

$$\mathrm{P}\left(Y_4 = j \middle| \tilde{\theta}_{2,-4}\right) = \exp\{\boldsymbol{\lambda_{4j}} + \boldsymbol{\nu_{4j2}}(\boldsymbol{\sigma_{22}}\boldsymbol{\nu_{3j2}} + \boldsymbol{\sigma_{12}}(\boldsymbol{\nu_{1j1}} + \boldsymbol{\nu_{2j1}}))\}/k_4$$

where $k_i = \sum_h^j \exp(\lambda_{ih} + \nu_{ihm}\tilde{\theta}_{m,-i})$.

Except for the first cycle of the algorithm, the conditional logit models with PL estimates obtained in Step 2 are the starting point of Step 1 for next cycle of the algorithm. The estimates of λ_{ih}, σ_{11}, σ_{22}, and σ_{12} are included in calculating a new weighted rest-score for estimating slope parameters and then using the new value of the weighted rest-score, Step 1 proceeds to update the slope parameters. And again, a new weighted rest-score for Step 2 is computed using the updated slope parameters and Step 2 continues to update the location and the association parameters based on the new values of the weighted rest-score. In this way, Steps 1 and 2 are repeated until all parameter estimates converge.

The Extended Application of PLE to LMA Models as MIRT Models

The proposed algorithm in this study is the extended version of the original application of PLE to LMA Rasch models (Anderson & Yu, 2997; Anderson, Li, & Vermunt, 2007; Li, 2010). The original application of PLE involves only fitting a single stacked conditional logistic model, which corresponds to Step 2. The proposed two-step algorithm can estimate parameters for general LMA models, including MIRT models where both location and slope parameters are included. This is true regardless of the number of latent variables and the response categories. Table 3 summarizes the extended application of PLE to more general LMA models as MIRT models, including unidimensional cases.

In addition to Rasch models, 2PL models for binary items and Bock's nominal response model for polytomous items, the more general LMA models with multiple latent variables can be fit by using the full algorithm, which includes Steps 1 and 2.

Table 3. LMA models as MIRT models covered by the algorithm

Latent variables	Item response	IRT model	The proposed algorithm	
			Step 1	Step 2
			One item at a time	Stacked regression
unidimensional	binary	Rasch	n.a.	X
		2PL	new	n.a.
	polytomous	Bock's NRM	new	n.a.
multidimensional		LMA Rasch models	n.a.	X
		General MIRT models	new	new

Unidimensional models can be fit to data using only Step 1. Either the 2PL or Bock's NRM can be estimated by iteratively fitting conditional logit models for each item.

Implementation of PLE in SAS for More General LMA Models

The proposed algorithm was implemented in SAS using a series of SAS macros. The conditional logit models in Steps 1 and 2 of the algorithm are fit using the MDC (Multinomial Discrete Choice) procedure in SAS. For identification, the sum over response categories of location and slope parameters for each item are set to zero (i.e., $\sum_j \lambda_{ij} = 0$ and $\sum_j \nu_{ij} = 0$). These are consistent with the linear restrictions imposed on location and slope parameters in Bock's nominal response model for polytomous items. For a scaling constraint, σ_{mm} is set equal to a constant (i.e., $\sigma_{mm} = 1$) for all latent variables, while all the association parameters between latent variables (i.e., $\sigma_{mm'}$) are estimated by PLE. In this section, input data sets for PLE are described, followed by SAS macros.

Input Data for PLE

The original PLE for LMA Rasch models was implemented in both R and SAS. The idea of input data described here follows the one employed in the SAS version of the previous application of PLE. Four types of input data are required to conduct PLE with SAS; (1) Response pattern (*Responses*), (2) Item ID (*Items*), (3) Item-by-trait adjacency matrix (*ItemTraitAdj*), and (4) Trait-by-trait adjacency matrix (*TraitAdj*).

"*Response*s" is an item response matrix containing the responses of *S* persons to *I* items (i.e., $S \times I$). Given the number of response categories equals *J*, the responses are represented in the matrix as 1, 2, …, *J*. Table 4 illustrates an example of the input data, 'Responses' from six 3-category items with 200 persons.

"*Items*" is a data set that contains the name of the *I* items from the data set called 'Responses'. Table 5 shows an example of 'Items' dataset. Each of six items in "*Responses*" are named as y1, y2, y3, y4, y5, and y6 in the "*Items*" dataset. These names are used when creating

Table 4. An example of ‘Responses’ dataset from six 3-category items

Person	Item1	Item2	Item3	Item4	Item5	Item6
1	1	3	2	1	1	1
2	2	2	1	2	1	1
3	3	1	3	2	1	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮
200	1	1	1	1	1	1

Table 5. An example of ‘Items” dataset from six items

Item1	Item2	Item3	Item4	Item5	Item6
y1	y2	y3	y4	y5	y6

new variables associated with responses to items in a master dataset by Dataset conversion macro, which will be explained in the next section, “SAS Macros for PLE”.

“*ItemTraitAdj*” is an item-by-latent trait adjacency matrix. If an item is directly related to a latent trait, the element corresponding to the item and the latent trait in the matrix takes 1 and 0 otherwise. Table 6 illustrate examples of “*ItemTraitAdj”* datasets with 6 items.

Table 6. Example of ‘ItemTraitAdj’ datasets from six items

(a) unidimensional

Item	Trait 1
1	1
2	1
3	1
4	1
5	1
6	1

(b) 2-dimensional

Item	Trait 1	Trait 2
1	1	0
2	1	0
3	1	0
4	0	1
5	0	1
6	0	1

(c) 3-dimensional

Item	Trait 1	Trait 2	Trait 3
1	1	0	0
2	1	0	0
3	0	1	0
4	0	1	0
5	0	0	1
6	0	0	1

Table 6 (a) represents an “*ItemTraitAdj*” matrix for a unidimensional model where all of six items are directly to one latent trait, Table 6 (b) for a 2-dimensional model where each half of 6 items are directly related one of two latent traits, and Table 6 (c) for a 3-dimensional model where each one-third of 6 items are directly related to one of three latent traits.

"*TraitAdj*" is a latent trait-by-latent trait adjacency matrix that represents whether or not a pair of latent traits is related to each other. An element involving two different latent traits is represented as 1 if one latent trait is related to the other and 0 otherwise. Table 7 shows examples of "*TraitAdj*" matrix for uni- and multidimensional models.

Table 7. Example of 'TraitAdj' datasets for uni- and multidimensional models

(a) unidimensional

	Trait 1
Trait 1	1

(b) 2-dimensional

	Trait 1	Trait 2
Trait 1	1	1
Trait 2	1	1

(c) 3-dimensional

	Trait 1	Trait 2	Trait 3
Trait 1	1	1	0
Trait 2	1	1	1
Trait 3	0	1	1

Table 7 (a) illustrates a "*TraitAdj*" matrix for a unidimensional model where '1' represents the latent trait itself. Table 7 (b) represents a "*TraitAdj*" matrix for a 2-dimensional model where two latent traits are related to each other. Table 7 (c) is an example for a 3-dimensional model where latent traits 1 and 2, 2 and 3 are correlated while latent traits 1 and 3 are not correlated each other.

SAS Macros for PLE

The extended PLE algorithm consists of basic and execution macros. The PLE basic macro includes 4 separate sub-macros; (1) Dataset conversion macro, (2) computation of weighted rest score macro, (3) Step 1 macro, and (4) Step 2 macro. They can be found in Appendix.

The Dataset conversion macro is to convert a standard item response matrix containing the responses of N persons to I items (i.e., $N \times I$) into a stacked dataset with the length of I items by J categories by S persons (i.e., $I \times J \times N$) and to create variables and initial values necessary for PLE algorithm by using the information provided in input data. Computation of weighted rest score macro is to compute rest-scores for each item and to create a data matrix that contains the scores for each item. The created dataset is used in Step 1 as an explanatory variable for slope parameters of each item. Step 1 and Step 2 macros involve fitting conditional logistic regression models for each item with MLE and a single stacked conditional logistic regression with PLE, respectively.

The PLE execution macro is to conduct PLE for a LMA model selected. Users should define several global macro variables by typing appropriate values or characters when they submit the macro. The global macro variables include ‘masterdata’, ‘cat2’, ‘response’, ‘idnum’, ‘step’, ‘nitems’, ‘ncat’, ‘iterations’, ‘scaling_constraint’, ‘allnphis’, and ‘mymodel’.

The ‘masterdata’ indicates a main data set to be analyzed during PLE execution. The ‘cat2’ variable is used to assign ‘2’ to the second category of the location parameters in “Model” statement of MDC procedure in Step 1. The ‘response’ variable represents an outcome variable in MDC procedure. The ‘idnum’ is a variable that identifies an individual in MDC procedure. The ‘step’ is used to perform only Step 1 (‘step = 1’) for unidimensional models or full steps (‘step = 2’) for multidimensional models. The ‘nitems’, ‘ncat’, and ‘iterations’ specify the number of items, the number of categories per item, and the number of iterations that users want to run the algorithm for convergence, respectively. The following three global variables are associated with a stacked logistic regression in Step 2. The ‘scaling_constraint’ variable is to impose scaling constraints on latent variables for model identification. The ‘allnphis’ is to specify all possible association parameters in an LMA model of interest. The ‘mymodel’ specifies all of the location parameters to be estimated in Step 2.

The values or characters for ‘masterdata’, ‘cat2’, ‘response’, and ‘idnum’ are fixed by default. In other words, users don’t have to make any changes in the values or characters for the variables. For unidimensional models, users should assign appropriate values for four global variables of ‘step’, ‘nitems’, ‘ncat’, and ‘iterations’. For multidimensional models, users need to define three more variables along with those four variables, which are ‘scaling_constraint’, ‘allnphis’, and ‘mymodel’. More detailed descriptions and examples for each global variable can be found in Appendix.

Execution of PLE Algorithm with SAS

PLE algorithm with SAS can be executed as follows: (1) run PLE basic and execution macros, (2) read input data in SAS, and (3) submit PLE execution macro with appropriately defined global variables for a model of interest. The following SAS codes show examples of how PLE execution macro is submitted with the global variables for uni- and 3-dimensional models with six 3-category items.

```
****************************************************
  Set-up global variables for Unidimensional models
****************************************************
%let masterdata= PolyMaster;
%let cat2 = 2 ;
%let response=y ;
%let idnum = caseID ;
%let step = 1 ;
%let nitems = 6 ;
%let ncat = 3 ;
%let iterations = 50 ;

%Execute_PLE

******************************************************
  Set-up global variables for Multidimensional models
******************************************************
%let masterdata= PolyMaster;
%let cat2 = 2 ;
%let response=y ;
%let idnum = caseID ;
%let step = 2 ;
%let nitems = 6 ;
%let ncat = 3 ;
%let iterations = 1 ;
%let scaling_constraint = nphi11=1, nphi22=1, nphi33 =1 ;
%let allnphis = nphi11 nphi12 nphi13 nphi22 nphi23 nphi33 ;
%let mymodel = Lamda12 Lamda13 Lamda22 Lamda23 Lamda32 Lamda33
               Lamda42 Lamda43 Lamda52 Lamda53 Lamda62 Lamda63 ;

%Execute_PLE  ;
```

Chapter 5

Research Methodology

Four sets of simulation studies were conducted to demonstrate the performance of the proposed pseudo-likelihood estimation method. The first two sets of simulation studies were designed to investigate how well item parameters in unidimensional models are estimated by Step 1 of the proposed algorithm, which is the new step that has been added. The next two sets of simulation studies examined the performance of the full algorithm, that is, Step 1 and Step 2 for multidimensional models. For all simulation studies, item parameters were generated from standard normal distributions with $a \sim N(0.1, 1)$ and $b \sim N(0, 1)$. For unidimensional models, item parameters with extreme values that fall outside the limits were excluded (i.e., $-0.4 < a_{ij} < 2.0$ and $-2.5 < b_{ij} < 2.7$). Latent trait values (θ) were drawn from a standard normal distribution $\theta \sim N(0,1)$ for unidimensional models and from a multivariate normal distribution for multidimensional models. Probabilities of response patterns were simulated according to 2PL model and Bock's nominal response model for unidimensional binary and polytomous item response models, respectively, and their generalizations for multidimensional models.

Simulation Studies for Unidimensional Models

Simulation studies 1 and 2 focus on the performance of Step 1 of the proposed estimation algorithm in estimating item parameters of LMA models with one latent variable. The main purpose of Simulation study 1 is to demonstrate that PLE behaves similarly to MLE by comparing the item parameter estimates obtained from the two estimation methods. To achieve the goal, item response datasets with small numbers of items (i.e., 4 and 6) were simulated, varying the number of response categories (i.e., 2, 3, and 5) and sample size (i.e., 200, 500, and 1000). Since the numbers of items in this simulation study are small, it is feasible to get item parameter estimates of LMA models by MLE and it allows to compare the parameter estimates obtained from PLE with those obtained from MLE. Additionally, marginal maximum likelihood estimation (MMLE) was also used to estimate item parameters of 2PL and Bock's nominal response models with small datasets using PROC NLMIXED procedure in SAS and the parameter estimates from MMLE and PLE were compared. This comparison will be reported in Chapter 6 to show that PLE works as well as standard IRT estimation methods even though two models are different, in particular, in terms of distribution assumptions.

Simulation study 2 was designed to examine that PLE overcomes the limitation of MLE when fitting LMA models with large numbers of items. For this simulation study, datasets with 20 and 50 items were simulated, varying the numbers of response categories (i.e., 2, 3, and 5) and sample size (i.e., 200, 500, and 1000). In the same way as Simulation study 1, 2PL and Bock's nominal response models with large datasets were also fit by MMLE using flexMIRT (Cai, 2013) and MULTILOG 7.0 (Thissen, Chen, & Bock, 2003), respectively, and the parameter estimates obtained from MMLE were compared with those obtained from PLE.

The simulation conditions for unidimensional models were designed by the number of items (4, 6, 20, 50) by the number of categories (2, 3, 5) by the number of people (200, 500, 1000); that is, there are 4 × 3 × 3 = 36 different conditions (i.e., 18 conditions for small numbers of items and 18 conditions for large numbers of items).

Identification constraints are required to estimate the parameters of LMA models. These may be setting one of location (λ_{ij}) and slope (ν_{ij}) parameters equal to zero (i.e., dummy coding, $\lambda_{i1} = 0$ and $\nu_{i1} = 0$) or setting the sum equal to zero (i.e., effect coding, $\sum_j \lambda_{ij} = 0$ and $\sum_j \nu_{ij} = 0$). Besides the identification constraints, LMA models require scaling constraints for model identification. Either of two ways may be used for scaling constraints. One possible way is to set $\sum_j {\nu^2}_{ijm} = 1$ for one item that is directly related to latent variable m. Another possible way is to set σ_{mm} to a constant (i.e., $\sigma_{mm} = 1$). In the simulation studies for unidimensional models, zero-sum constraints for location constraints and $\sigma_{11} = 1$ for scaling constraints were used so σ_{11} is not estimated.[5]

Simulation Studies for Multidimensional Models

Simulation studies 3 and 4 aim to evaluate the performance of the full algorithm (i.e., Step 1 and 2) for multidimensional models. Simulation study 3 has the same goal as Simulation study 1 for unidimensional models with small numbers of items, which is how well PLE of LMA models performs relative to MLE of LMA models. Since the simulation study was intended for multidimensional models, the number of latent variables was varied (i.e., 2 and 3) with the correlational structure between latent variables of $r = 0.50$. The number of response categories and sample size varied for simulations are the same as the simulation studies for unidimensional models.

Lastly, Simulation study 4 was conducted to illustrate that PLE works on large numbers of items when fitting LMA models with multiple latent variables. Since Simulation study 4 involves large numbers of items, its simulation design is almost the same as that of Simulation 2, which is for unidimensional models with 20 and 50 items. The only difference between them is that the number of latent variables was considered (i.e., 2, 3, and 4) in Simulation study 4. For correlational structure between latent variables, it was set to $r = 0.50$. Studies 3 and 4 for multidimensional models consist of 81 simulation conditions.

---

[5] Estimates can be rescaled if an estimate of σ_{11} is desired.

For location constraints, zero-sum constraints were placed on the location and slope parameters. For scaling constraints, σ_{mm} was set equal to a constant (i.e., $\sigma_{mm} = 1$) for all latent variables, while all the association parameters between latent variables (i.e., $\sigma_{mm'}$) were estimated by PLE.

Each simulation condition in all studies was replicated 30 times and the parameter estimates from the replications were averaged. The parameter estimates of LMA models obtained from PLE and MLE were transformed so that they would be placed on the same scale as those of standard IRT models (i.e., 2PL and Bock's NRM) and comparable with true parameters used to simulate the data. The transformed estimates were used to compute the mean and standard deviation of the estimates from 30 replications and the evaluation criteria assessing the performance of the estimation methods.

Evaluation Criteria

To evaluate how well PLE performed, 3 criteria were selected: bias, root mean squared error (RMSE), and Pearson product moment correlation coefficient (*r*).

Bias and RMSE were calculated to see how accurately MMLE, MLE, and PLE recovered the parameters used to simulate the data. Bias is defined as the mean difference between the estimated parameter and the parameter used to simulate the data; that is,

$$Bias(\lambda_{ij}) = \frac{\sum_{r=1}^{R}(\hat{\lambda}_{ij} - \lambda_{ij})}{R} \, , \tag{5.1}$$

where λ_{ij}= the location parameter used to simulate the data for category *j* on item *i*, $\hat{\lambda}_{ij}$= the estimated location parameter of category *j* on item *i*, and *R* = the number of replications. Positive bias indicates that the parameter is overestimated, while negative bias reflects the parameter is underestimated.

RMSE is the square root of the mean of the squared difference between the estimated parameter and the parameter used to simulate the data; that is,

$$RMSE\left(\lambda_{ij}\right) = \sqrt{\frac{\sum_{r=1}^{R}\left(\hat{\lambda}_{ij} - \lambda_{ij}\right)^2}{R}} \ . \tag{5.2}$$

Smaller RMSE reflects greater accuracy.

For uni- and multidimensional models with small numbers of items, the Root Mean Squared Difference (RMSDiff) of the parameter estimates between MLE and PLE of LMA models was also computed; that is,

$$RMSDiff\left(\hat{\lambda}_{ij}\right) = \sqrt{\frac{\sum_{p=1}^{P}\left(\hat{\lambda}_{ij}^{MLE} - \hat{\lambda}_{ij}^{PLE}\right)^2}{P}} \ . \tag{5.3}$$

where $\hat{\lambda}_{ij}^{MLE}$ and $\hat{\lambda}_{ij}^{PLE}$ = the estimated location parameter of category *j* on item *i* by MLE and PLE, respectively, and *P* = the number of location parameters to be estimated. The small RMSDiff indicates that the two estimates obtained from MLE and PLE are not only linearly related but also close to each other.

Pearson product moment correlation coefficient (*r*) was calculated to evaluate the accuracy of the parameter estimation by PLE and the equivalence of the parameter estimates between PLE and MLE of LMA models as well.

Estimation of Standard Errors of PL Estimates

Pseudo-likelihood estimation is quite convenient because the maximization of the pseudo-likelihood function is equivalent to the maximization of the likelihood function of a logistic regression model. The standard errors given by pseudo-likelihood estimation, however, are underestimated. In the logistic regression procedure, the parameter estimates and the standard errors of the estimates are obtained based on the assumption that observations are independent.

However, this assumption of independence is not met when MLE procedure is exploited to maximize the pseudo-likelihood function with a stacked dataset because the responses from a single person are set up in the stacked dataset with $I \times J$ rows and thus dependency among observations from the single person exists in the dataset. In pseudo-likelihood estimation using a logistic regression procedure with a stacked dataset, the parameters and standard errors are estimated, ignoring the dependency. Since the observations from the same single person are strongly interrelated, the resulting standard errors obtained by pseudo-likelihood estimation will be considerably small. The problem of ignoring dependency is also occurred when iteratively fitting conditional logit models for each item in Step 1 of the algorithm because they are not independent. To estimate correct standard errors, two possible methods can be considered: jackknife and bootstrap. Both jackknife and bootstrap involve resampling data; that is, repeatedly creating new data sets from the original data.

Jackknife

The jackknife removes one observation from the original sample and calculates an estimate based on the remaining $N - 1$ of them. This process is repeated *N* times, leaving one observation out at a time. Let $\boldsymbol{Y}_{-n} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots \boldsymbol{y}_{n-1}, \boldsymbol{y}_{n+1} \ldots., \boldsymbol{y}_N)$ be a new sample obtained by leaving observation *n* out from the original sample, let $\hat{\lambda}_{ij}^{(n)}$ be the estimate of λ_{ij} from a new sample, and let $\hat{\lambda}_{ij}^{(\cdot)}$ be the mean of *N* estimates of λ_{ij} obtained from *N* new samples, that is, $\hat{\lambda}_{ij}^{(\cdot)} = \frac{1}{N}\sum_{n=1}^{N}\hat{\lambda}_{ij}^{(n)}$. Then the jackknife estimate of standard error of λ_{ij} is defined

$$\sigma_J(\lambda_{ij}) = \sqrt{\frac{N-1}{N}\sum_{n=1}^{N}(\hat{\lambda}_{ij}^{(n)} - \hat{\lambda}_{ij}^{(\cdot)})^2}\ . \tag{5.4}$$

Jackknife has been found to work well for RC(M) association model (Clogg & Eliason, 1987; Wong, 2011).

Bootstrap

The bootstrap method involves drawing random samples of size *N* with replacement repeatedly (e.g., 1,000 times) from the original sample that is also of size *N*. Let $\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_n \dots., \boldsymbol{y}_N)$ be the original sample of size *N*, let $\boldsymbol{Y}^*(m) = (\boldsymbol{y}_1^*, \dots, \boldsymbol{y}_n^*, \dots, \boldsymbol{y}_N^*)$ be the *m*th bootstrap sample, let $\hat{\lambda}_{ij}^{(m)}$ be the estimate of λ_{ij} from the *m*th bootstrap sample, $\boldsymbol{Y}^*(m)$, and let $\hat{\lambda}_{ij}^{(*)}$ be the mean of *M* estimates of λ_{ij} obtained from the *M* bootstrap samples, that is, $\hat{\lambda}_{ij}^{(*)} = \frac{1}{M}\sum_{m=1}^{M} \hat{\lambda}_{ij}^{(m)}$. Then the bootstrap estimate of standard error of λ_{ij} is given by

$$\sigma_B(\lambda_{ij}) = \sqrt{\frac{1}{M-1}\sum_{m=1}^{M} (\hat{\lambda}_{ij}^{(m)} - \hat{\lambda}_{ij}^{(*)})^2} \ . \tag{5.5}$$

Although jackknife and bootstrap procedures are considered as standard ways to estimate standard errors, they are also known to be computationally time consuming procedures. In this thesis, jackknife method was chosen to correct the standard errors given by PLE and performed for 18 unidimensional models with small numbers of items (i.e., Simulation study 1). To apply “leave-one-out” procedure of jackknife, 1) samples were formed from a whole item response pattern matrix containing the responses of *N* persons to *I* items (i.e., $N \times I$), 2) one person’s item response pattern was deleted from $N \times I$ item response pattern, and 3) the $(N-1) \times I$ data matrix was transformed into a stacked data matrix for pseudo-likelihood estimation. The jackknife procedure and PLE with the stacked dataset obtained from the sample was repeated as many times as sample size (*N*) of the sample. Using the *N* estimates obtained by PLE, jackknife estimates of variances of the parameters in LMA models were computed as

$$\sigma^2{}_J(\lambda_{ij}) = \frac{N-1}{N}\sum_{n=1}^{N}(\hat{\lambda}_{ij}^{(n)} - \hat{\lambda}_{ij}^{(\cdot)})^2 \ . \tag{5.6}$$

This computation was repeated with 30 replication datasets and the jackknife estimates of variances of each parameter were averaged across 30 replications. Finally, the jackknife estimates of standard errors were obtained by taking the square root of the mean of 30 jackknife estimates of variances of each parameter in LMA models as below.

$$\sigma_J(\lambda_{ij}) = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\sigma_J^2(\lambda_{ij})} \ , \tag{5.7}$$

where $\sigma_J(\lambda_{ij})$ = the jackknife estimate of standard error of λ_{ij}, $\sigma_J^2(\lambda_{ij})$ = the jackknife estimate of variance of λ_{ij}, and R = the number of replications.

In addition to jackknife estimates of standard errors, two types of standard error estimates were also calculated: the standard deviation of the estimated parameters from 30 replications and the square root of the mean of the squared standard error given by PLE, MLE, and MMLE procedures over the 30 replications.

The first type of standard error estimates is the standard deviation of the estimates from 30 replications. As mentioned earlier, the item parameters of 18 unidimensional models with small number of items were estimated by PLE, MLE, and MMLE and the parameter estimation was repeated 30 times with different datasets to yield 30 sets of parameter estimates. The standard deviations of each of the estimated parameters were calculated from the distribution of the 30 replicate estimates of each parameter and they were used for one of the standard error estimates in this thesis.

The second type of standard error estimates is associated with the original standard errors in the output given by three different estimation procedures across replications when the analysis is done. They are named here as PLE_SE, MLE_SE, and MMLE_SE to represent that they are from original standard errors. The standard error estimates were computed by three steps. First, the original standard errors from PLE, MLE, and MMLE were squared to get the variance estimates of the estimated parameters. Second, the variance estimates were averaged across replications. Third, by taking the square root of the averaged variance estimate of each parameter, the resulting values for PLE_SE, MLE_SE, and MMLE_SE were obtained. They are used to examine the standard error estimates in this study.

Chapter 6

Simulation Studies

Simulation studies are reported here to demonstrate the performance of the extended pseudo-likelihood estimation method, followed by the standard error estimation by jackknife procedure for unidimensional models with small numbers of items, and the computational time of PLE for large numbers of items. As described in research methodology section, each of 117 simulation conditions were replicated 30 times, and the parameter estimates for each condition were averaged across replications. Using the averaged parameter estimates, bias, root mean squared error (RMSE), and correlation coefficients were computed to assess the accuracy of the parameter recovery of PLE and compared with those from MMLE and MLE.

Unidimensional Models with Small Numbers of Items

This section describes the results of the simulation studies focusing on unidimensional LMA models with small numbers of items that only use Step 1 of the PLE algorithm, which is the new step. The performance of PLE for unidimensional LMA models with 4 and 6 items are presented with respect to comparisons of PLE with MLE and parameter recovery. As mentioned earlier, bias, RMSE and correlation coefficients were computed for each item to evaluate parameter recovery. For example, Table 8 contains bias values of parameters for unidimensional models with 4 items and sample size of 1000 by estimation method. In this thesis, the bias and RMSE computed for each item are averaged over items and categories to simplify the interpretation and reported with their standard deviations in the tables.

Comparisons of PLE with MLE

The main purpose of the study of unidimensional models with small numbers of items is to demonstrate that how similarly PLE behaves to MLE. This study shows how well the new

Table 8. Bias of parameters for unidimensional models with 4 items and sample size of 1000, by estimation method

Item	Location			Slope		
	MMLE	MLE	PLE	MMLE	MLE	PLE
			2 categories (N=1000)			
Item 1	0.0005	0.0236	0.0236	0.0313	0.0028	0.0027
Item 2	0.0105	-0.0963	-0.0964	0.0141	-0.0675	-0.0675
Item 3	0.0269	0.4490	0.4494	-0.0109	-0.0336	-0.0338
Item 4	0.0159	0.0410	0.0409	0.0759	-0.0505	-0.0508
			3 categories (N=1000)			
Item 1_2	0.0006	-0.2901	-0.2904	-0.0570	0.1372	0.1355
Item 1_3	0.0081	-0.2204	-0.2230	-0.0022	0.0628	0.0578
Item 2_2	0.0009	-0.1140	-0.1134	-0.0033	-0.0626	-0.0621
Item 2_3	-0.0364	0.2856	0.2840	0.0869	0.0478	0.0492
Item 3_2	0.0306	0.2358	0.2344	0.0343	0.1022	0.0956
Item 3_3	-0.0604	-0.1618	-0.1591	-0.0753	-0.2297	-0.2195
Item 4_2	-0.0026	-0.1711	-0.1719	-0.0029	0.1160	0.1158
Item 4_3	-0.0369	0.0229	0.0233	0.0475	-0.0128	-0.0121
			5 categories (N=1000)			
Item 1_2	0.0326	-0.3256	-0.3267	0.0257	0.1306	0.1310
Item 1_3	0.0299	0.0229	0.0217	-0.0044	0.0942	0.0912
Item 1_4	-0.0040	0.3894	0.3881	0.0549	0.1619	0.1533
Item 1_5	0.0297	0.2325	0.2309	0.0286	0.1216	0.1146
Item 2_2	0.0148	-0.3003	-0.3008	-0.0023	0.0129	0.0160
Item 2_3	0.0265	-0.2195	-0.2199	-0.0260	-0.0336	-0.0297
Item 2_4	-0.0126	0.1323	0.1327	0.0496	0.0352	0.0332
Item 2_5	-0.0425	-0.0424	-0.0422	-0.0442	-0.0380	-0.0369
Item 3_2	-0.0149	-0.3297	-0.3333	0.0211	0.2648	0.2582
Item 3_3	0.0208	0.1217	0.1270	-0.0795	-0.1787	-0.1704
Item 3_4	-0.0567	-0.0169	-0.0123	0.1050	0.0660	0.0731
Item 3_5	0.0359	0.6283	0.6348	-0.0543	-0.2626	-0.2513
Item 4_2	0.0419	0.2429	0.2433	-0.0501	-0.0770	-0.0792
Item 4_3	-0.0535	0.1403	0.1376	0.0289	-0.0288	-0.0246
Item 4_4	0.0110	0.1436	0.1440	0.0014	-0.0626	-0.0647
Item 4_5	-0.0321	-0.0511	-0.0509	0.0132	-0.0876	-0.0868

algorithm is working. Since MLE is feasible to get item parameter estimates of LMA models with 4 and 6 items, it allows to compare the parameter estimates obtained from PLE with those obtained from MLE.

Table 9. Mean bias, RMSE and their (SDs) and correlation coefficients for location parameters of unidimensional models with 4 and 6 items, by estimation method

(a) 2 categories

	4 items					6 items				
	Bias		RMSE		$r_{(\lambda,\ \hat{\lambda})}$	Bias		RMSE		$r_{(\lambda,\ \hat{\lambda})}$
					Sample size (N) =200					
MMLE	0.0220	(0.0606)	0.2243	(0.1038)	0.998	-0.0083	(0.0206)	0.1632	(0.0488)	0.999
MLE	0.1054	(0.2675)	0.4015	(0.2506)	0.998	-0.0373	(0.0278)	0.1852	(0.0649)	0.999
PLE	0.1057	(0.2676)	0.4006	(0.2514)	0.998	-0.0371	(0.0278)	0.1852	(0.0647)	0.999
					Sample size (N) = 500					
MMLE	-0.0116	(0.0117)	0.1360	(0.0542)	0.999	-0.0008	(0.0303)	0.1124	(0.0335)	0.998
MLE	0.0781	(0.1240)	0.2172	(0.1377)	0.999	-0.0355	(0.0394)	0.1114	(0.0521)	0.997
PLE	0.0780	(0.1243)	0.2173	(0.1376)	0.999	-0.0354	(0.0392)	0.1120	(0.0527)	0.997
					Sample size (N) = 1000					
MMLE	0.0134	(0.0110)	0.1028	(0.0526)	0.999	0.0077	(0.0184)	0.0988	(0.0607)	0.999
MLE	0.1043	(0.2378)	0.2437	(0.2006)	0.999	-0.0246	(0.0225)	0.1006	(0.0681)	0.999
PLE	0.1044	(0.2380)	0.2438	(0.2009)	0.999	-0.0245	(0.0225)	0.1006	(0.0682)	0.999

(b) 3 categories

	4 items					6 items				
	Bias		RMSE		$r_{(\lambda,\ \hat{\lambda})}$	Bias		RMSE		$r_{(\lambda,\ \hat{\lambda})}$
					Sample size (N) =200					
MMLE	-0.0135	(0.0310)	0.2762	(0.0921)	0.999	-0.0203	(0.0753)	0.3100	(0.0992)	0.998
MLE	-0.0470	(0.2239)	0.2602	(0.0853)	0.970	-0.0210	(0.1629)	0.3002	(0.0660)	0.982
PLE	-0.0477	(0.2232)	0.2615	(0.0825)	0.971	-0.0233	(0.1623)	0.2952	(0.0636)	0.982
					Sample size (N) = 500					
MMLE	-0.0109	(0.0686)	0.2084	(0.0704)	0.998	-0.0205	(0.0513)	0.2344	(0.0838)	0.999
MLE	-0.0476	(0.2145)	0.2444	(0.0545)	0.973	-0.0305	(0.1683)	0.2481	(0.0461)	0.981
PLE	-0.0477	(0.2148)	0.2454	(0.0538)	0.973	-0.0307	(0.1680)	0.2482	(0.0468)	0.981
					Sample size (N) = 1000					
MMLE	-0.0120	(0.0297)	0.1364	(0.0483)	0.999	-0.0049	(0.0344)	0.1437	(0.0509)	0.999
MLE	-0.0516	(0.2130)	0.2184	(0.0681)	0.973	-0.0385	(0.1789)	0.2109	(0.0556)	0.978
PLE	-0.0520	(0.2126)	0.2189	(0.0679)	0.973	-0.0386	(0.1798)	0.2112	(0.0560)	0.978

Table 9. (cont.)

(c) 5 categories

	4 items					6 items				
	Bias		RMSE		$r_{(\lambda,\ \hat{\lambda})}$	Bias		RMSE		$r_{(\lambda,\ \hat{\lambda})}$
Sample size (N) =200										
MMLE	-0.0326	(0.0992)	0.4306	(0.1956)	0.993	-0.0078	(0.0901)	0.3444	(0.1231)	0.995
MLE	0.0225	(0.2978)	0.4031	(0.1360)	0.933	0.0293	(0.2602)	0.3777	(0.1340)	0.959
PLE	0.0217	(0.2989)	0.4034	(0.1361)	0.933	0.0283	(0.2623)	0.3786	(0.1342)	0.958
Sample size (N) = 500										
MMLE	-0.0100	(0.0643)	0.2366	(0.0997)	0.998	0.0028	(0.0454)	0.2199	(0.1059)	0.999
MLE	0.0452	(0.2688)	0.2974	(0.1223)	0.944	0.0474	(0.2410)	0.2898	(0.1075)	0.964
PLE	0.0455	(0.2689)	0.2970	(0.1222)	0.944	0.0459	(0.2451)	0.2915	(0.1105)	0.963
Sample size (N) = 1000										
MMLE	0.0017	(0.0473)	0.1902	(0.0914)	0.999	-0.0043	(0.0306)	0.1543	(0.0709)	0.999
MLE	0.0480	(0.1326)	0.2759	(0.0928)	0.946	0.0398	(0.2359)	0.2445	(0.1270)	0.967
PLE	0.0484	(0.1284)	0.2768	(0.0933)	0.946	0.0390	(0.2393)	0.2457	(0.1304)	0.966

Table 10. Mean bias, RMSE and their (SDs) and correlation coefficients for slope parameters of unidimensional models with 4 and 6 items, by estimation method

(a) 2 categories

	4 items					6 items				
	Bias		RMSE		$r_{(\nu,\ \hat{\nu})}$	Bias		RMSE		$r_{(\nu,\ \hat{\nu})}$
Sample size (N) =200										
MMLE	0.0633	(0.0732)	0.3183	(0.1028)	0.986	0.0611	(0.0576)	0.2733	(0.0495)	0.994
MLE	-0.0047	(0.1038)	0.4290	(0.0464)	0.985	0.0274	(0.0891)	0.3203	(0.0520)	0.991
PLE	-0.0054	(0.1040)	0.4274	(0.0466)	0.985	0.0294	(0.0888)	0.3242	(0.0527)	0.991
Sample size (N) = 500										
MMLE	0.0354	(0.0542)	0.2000	(0.0456)	0.997	0.0094	(0.0500)	0.1845	(0.0644)	0.994
MLE	0.1076	(0.0568)	0.3079	(0.0371)	0.996	0.0127	(0.0563)	0.2052	(0.0347)	0.994
PLE	0.1067	(0.0569)	0.3066	(0.0377)	0.996	0.0125	(0.0567)	0.2061	(0.0352)	0.994
Sample size (N) = 1000										
MMLE	0.0276	(0.0365)	0.1544	(0.0587)	0.999	0.0237	(0.0277)	0.1326	(0.0430)	0.999
MLE	-0.0372	(0.0300)	0.1785	(0.0390)	0.999	-0.0180	(0.0332)	0.1464	(0.0289)	0.998
PLE	-0.0374	(0.0300)	0.1786	(0.0391)	0.999	-0.0179	(0.0333)	0.1464	(0.0287)	0.998

Table 10. (cont.)

(b) 3 categories

	4 items					6 items				
	Bias		RMSE		$r_{(\nu,\ \hat{\nu})}$	Bias		RMSE		$r_{(\nu,\ \hat{\nu})}$
					Sample size (N) =200					
MMLE	0.0432	(0.0769)	0.3836	(0.1094)	0.999	-0.0093	(0.0789)	0.3751	(0.0796)	0.999
MLE	0.0365	(0.1254)	0.3517	(0.0831)	0.997	0.0166	(0.1584)	0.3687	(0.0911)	0.996
PLE	0.0391	(0.1229)	0.3504	(0.0843)	0.997	0.0145	(0.1404)	0.3593	(0.0810)	0.997
					Sample size (N) = 500					
MMLE	-0.0049	(0.1070)	0.3015	(0.0944)	0.999	-0.0034	(0.0499)	0.3094	(0.0901)	0.999
MLE	0.0199	(0.1557)	0.2673	(0.0569)	0.996	0.0182	(0.1255)	0.2905	(0.0711)	0.998
PLE	0.0201	(0.1545)	0.2669	(0.0575)	0.996	0.0158	(0.1172)	0.2882	(0.0699)	0.998
					Sample size (N) = 1000					
MMLE	0.0035	(0.0532)	0.2304	(0.0541)	0.999	0.0029	(0.0454)	0.1870	(0.0431)	0.998
MLE	0.0201	(0.1210)	0.2056	(0.0488)	0.998	0.0195	(0.1211)	0.1981	(0.0436)	0.997
PLE	0.0200	(0.1169)	0.2070	(0.0491)	0.998	0.0162	(0.1127)	0.1930	(0.0400)	0.997

(c) 5 categories

	4 items					6 items				
	Bias		RMSE		$r_{(\nu,\ \hat{\nu})}$	Bias		RMSE		$r_{(\nu,\ \hat{\nu})}$
					Sample size (N) =200					
MMLE	-0.0393	(0.1044)	0.5334	(0.1908)	0.993	-0.0195	(0.0936)	0.4500	(0.1423)	0.996
MLE	-0.0469	(0.1231)	0.5104	(0.1810)	0.993	-0.0218	(0.1042)	0.4389	(0.1431)	0.995
PLE	-0.0480	(0.1230)	0.5118	(0.1808)	0.993	-0.0214	(0.1010)	0.4378	(0.1437)	0.995
					Sample size (N) = 500					
MMLE	-0.0004	(0.1078)	0.3244	(0.1330)	0.994	0.0117	(0.0702)	0.3017	(0.1138)	0.998
MLE	0.0017	(0.1823)	0.3291	(0.1258)	0.983	0.0084	(0.1450)	0.3057	(0.1175)	0.990
PLE	0.0024	(0.1813)	0.3281	(0.1253)	0.983	0.0088	(0.1355)	0.3021	(0.1162)	0.991
					Sample size (N) = 1000					
MMLE	0.0042	(0.0473)	0.2285	(0.0914)	0.999	0.0033	(0.0558)	0.2118	(0.0847)	0.998
MLE	0.0074	(0.1326)	0.2579	(0.0928)	0.992	0.0099	(0.1047)	0.2199	(0.0734)	0.994
PLE	0.0080	(0.1284)	0.2553	(0.0933)	0.993	0.0102	(0.0954)	0.2162	(0.0716)	0.995

Tables 9 and 10 show mean bias, RMSE and their standard deviations for location (λ_{ij}) and slope (ν_{ij}) parameters of unidimensional models with 4 and 6 items by three different estimation procedures. The correlations (r) between the estimates (i.e., $\hat{\lambda}_{ij}$ and $\hat{\nu}_{ij}$) and the parameters used to simulate the data (i.e., λ_{ij} and ν_{ij}) are also shown in the tables.

Expected from comparing MLE and PLE of LMA models, Tables 8 and 9 show that mean bias, mean RMSE, and correlations of PLE are almost identical to those of MLE across all conditions. The findings are more clearly illustrated in Figures 4 through 6 for location parameters and Figures 7 through 9 for slope parameters. They further confirm that PLE recovered the parameters used to simulate the data with the same amount of errors as MLE in all unidimensional models.

Table 11 provides the correlations between the parameter estimates obtained from MLE and PLE for all of 18 unidimensional models with 4 and 6 items. All of the correlations are 0.999 or 1.000 across all models, indicating the parameter estimates between MLE and PLE are all equivalent.

Table 11. Correlation coefficients (r) between the parameter estimates obtained from MLE and PLE for unidimensional models with 4 and 6 items

	2 categories			3 categories			5 categories		
	Sample size			Sample size			Sample size		
	N=200	N=500	N=1000	N=200	N=500	N=1000	N=200	N=500	N=1000
Location parameters									
4 items	1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.999	0.999
6 items	0.999	0.999	1.000	0.999	0.999	0.999	0.999	0.999	0.999
Slope parameters									
4 items	1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.999	0.999
6 items	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999

Table 12. Mean RMSDiff of parameter estimates between MLE and PLE for unidimensional models with 4 and 6 items.

	2 categories			3 categories			5 categories		
	Sample size			Sample size			Sample size		
	N=200	N=500	N=1000	N=200	N=500	N=1000	N=200	N=500	N=1000
Location parameters									
4 items	0.000	0.000	0.000	0.022	0.015	0.009	0.016	0.009	0.007
6 items	0.001	0.001	0.000	0.035	0.022	0.012	0.017	0.013	0.029
Slope parameters									
4 items	0.001	0.001	0.000	0.016	0.011	0.007	0.014	0.007	0.006
6 items	0.002	0.001	0.001	0.019	0.013	0.008	0.012	0.008	0.014

Table 12 contains the mean RMSDiff of the parameter estimates between MLE and PLE of LMA models. It was computed to assess how similar the parameter estimates from PLE are to those from MLE. The small RMSDiff indicates that the two estimates obtained from MLE and PLE are not only linearly related but also close to each other. The differences range from 0.000 to 0.035 for location parameters and 0.000 to 0.019 for slope parameters. Overall, the mean RMSDiff values tend to decrease as sample sizes increase given the number of items and response categories. All the values in the table show very small mean RMSDiff between MLE and PLE, suggesting that the parameter estimates of PLE are very close to those of MLE, and PLE behaves almost the same as MLE in estimating parameters of LMA models.

Comparisons of MMLE of IRT Models and PLE of LMA Models

Location parameters. Table 9 shows that PLE yields relatively small mean bias for location parameters in each of the 18 unidimensional models with small numbers of items, ranging from -0.0520 to 0.1057. The binary response model with 4 items and 200 persons resulted in the largest amount of bias with a value of 0.1057, while the 5 response categories with the same number of items and sample size yielded the minimum bias value of 0.0217.

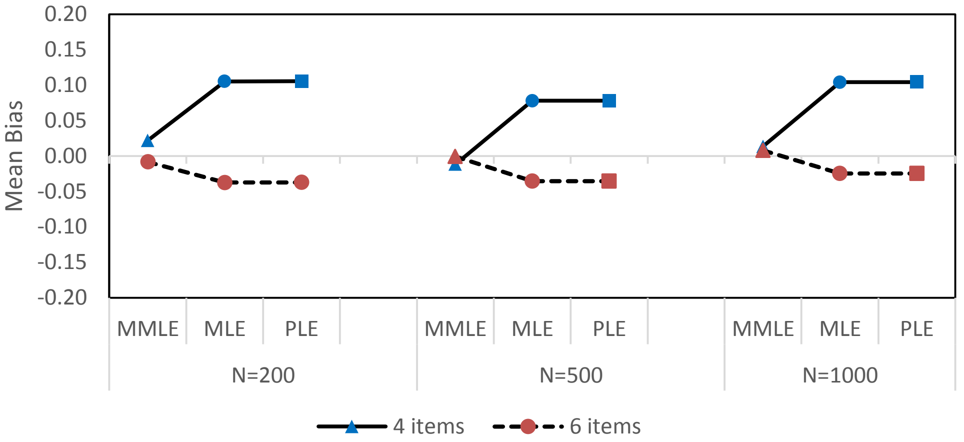
Overall, the mean bias for polytomous response models are smaller than those for binary response models. It suggests that PLE recovered location parameters of polytomous response models with less amount of errors than those of binary response models.

Figure 4 provides the mean bias results found in Table 9. As would be expected, the standard IRT estimation method, MMLE produces the mean bias values that are close to 0 over all conditions. The big differences in mean bias between PLE and MMLE is found in the binary response models with 4 items over all sample sizes. PLE and MMLE yield nearly identical mean bias values for the conditions of 3 categories, 6 items, and sample sizes of 200 and 500, which are close to 0. It indicates that PLE recovered location parameters of the models as well as MMLE did.

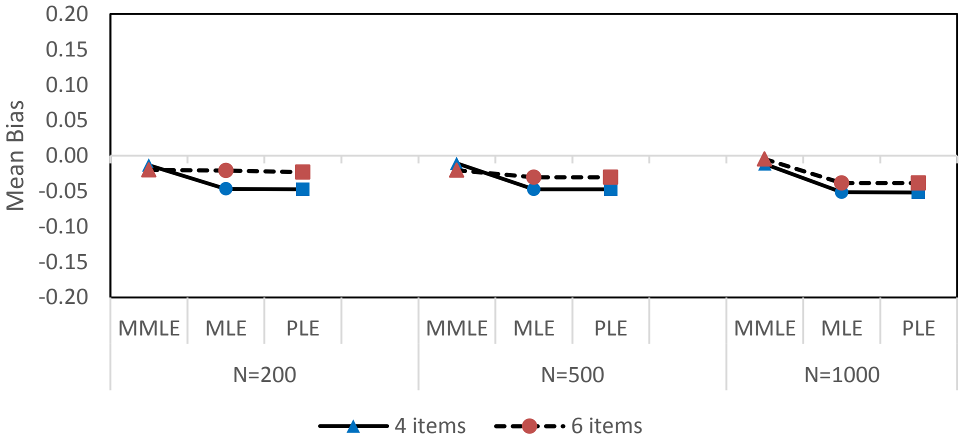
The mean RMSE for location parameters of PLE ranges from 0.1006 to 0.4034. The large RMSE of PLE occur in two conditions with values of RMSE = 0.4006 for a binary response model with 4 items and 200 persons and RMSE = 0.4034 for a 5 response category model with 4 items and 200 persons. The largest mean RMSE value of MMLE is also shown in the 5 response category model with 4 items and 200 persons with a value of 0.4306, which is a little larger than the value of PLE. Given the same number of items and sample sizes, the mean RMSE for location parameters increases as the number of categories increases over all estimation procedures compared. Because the 5 response category model with 4 items and 200 persons resulted in the largest mean RMSE of both PLE and MMLE, this finding indicates that a short test with a small sample size and a large number of categories may result in misleading estimates.

Figure 5 illustrates the mean RMSE results found in Table 9. One noticeable result is that the mean RMSE values of PLE for location parameters are almost equal to those of MMLE

(a) 2 categories



(b) 3 categories



(c) 5 categories

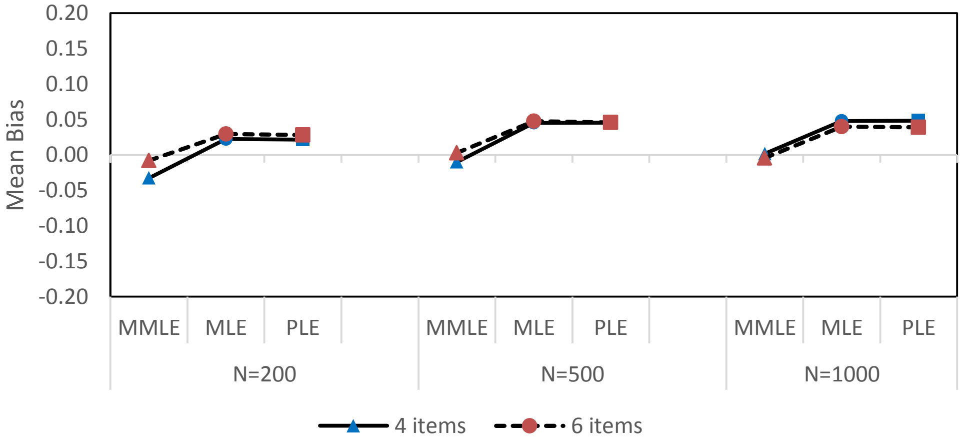
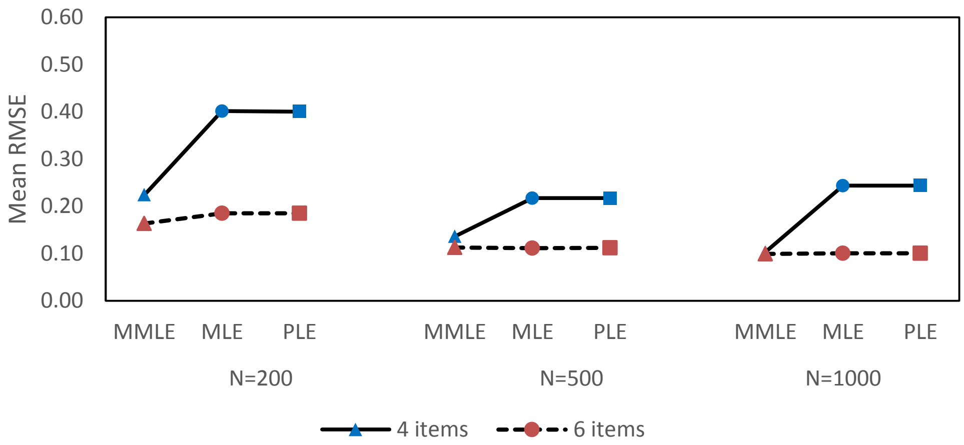
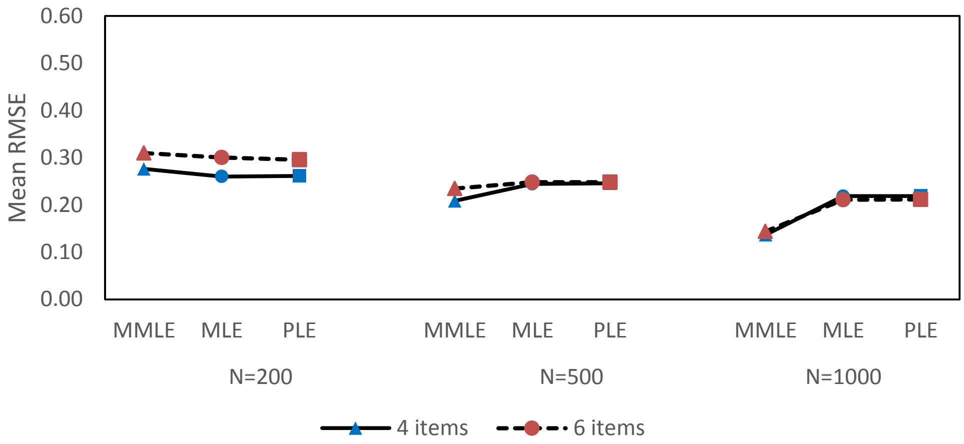


Figure 4. Mean bias for location parameters of unidimensional models with 4 and 6 items

(a) 2 categories



(b) 3 categories



(c) 5 categories

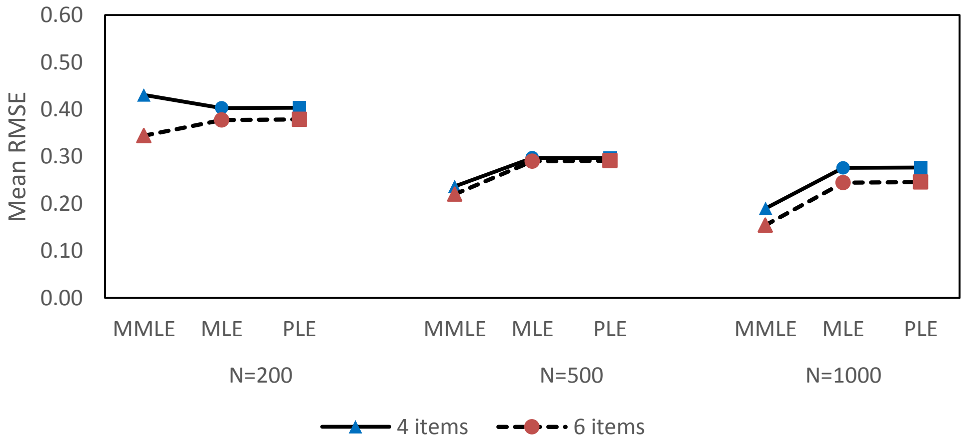


Figure 5. Mean RMSE for location parameters of unidimensional models with 4 and 6 items

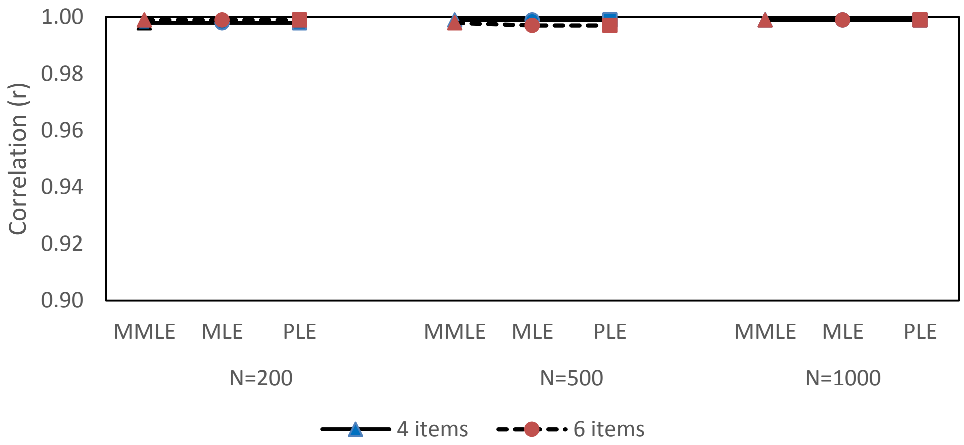
in five conditions associated with 6 items. For binary response models with 6 items, the mean RMSE values of PLE and MMLE are 0.1852 and 0.1632 for sample size of 200, respectively, and the differences in mean RMSE between two estimation procedures decreases as sample size gets large, resulting in almost the same RMSE values. The similar pattern is found in 3 response category models with the sample sizes of 200 and 500. The mean RMSE of PLE with sample of 200, however, is a bit smaller than the value of MMLE (i.e., RMSE = 0.2952 for PLE, and RMSE = 0.3100 for MMLE). The findings from Figure 5 suggest that PLE may behave similarly to MMLE when it recovers true parameters.

Table 9 also shows the correlation coefficients (r) between the estimates ($\hat{\lambda}_{ij}$) and the location parameters used to simulate the data (λ_{ij}). MMLE shows consistently highest correlations across all conditions. PLE also yields high correlations in all conditions, ranging from 0.933 to 0.999.

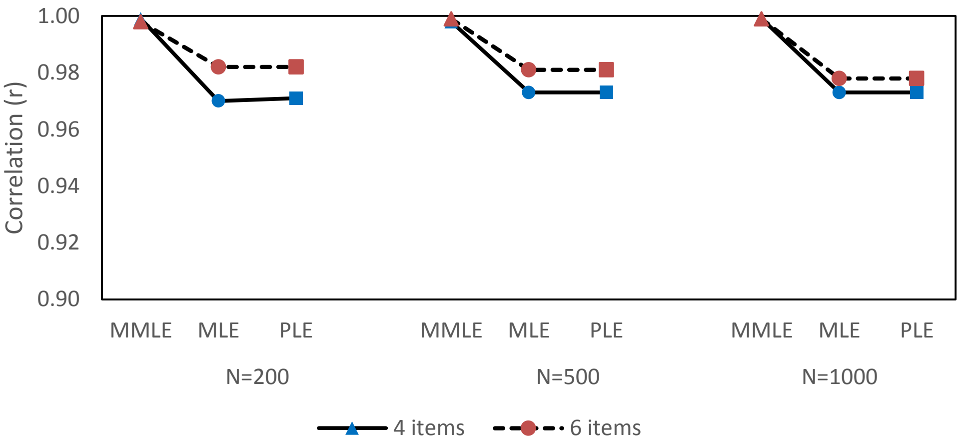
Figure 6 illustrates that, for all binary response models, the correlations of PLE between the estimates and true parameters for location parameters are exactly the same as those of MMLE with coefficients of 0.998 for sample size of 200 and 0.999 for sample sizes of 500 and 1000. The magnitude of the correlations of PLE gets smaller as the number of categories increases to 3 and 5, although the correlations associated with 3 and 5 categories are still high. The correlations for 3 and 5 categories with 6 items, however, are higher than those with 4 items, which indicates that the correlations of PLE between the estimates and true parameters for location parameters may be improved by increasing the number of items.

Slope parameters. Table 10 shows that the mean bias of PLE for slope parameters are relatively small across all unidimensional models with 4 and 6 items, ranging from -0.0054 to 0.1067. The largest mean bias of PLE occurs in the binary response model with 4 items and

(a) 2 categories



(b) 3 categories



(c) 5 categories

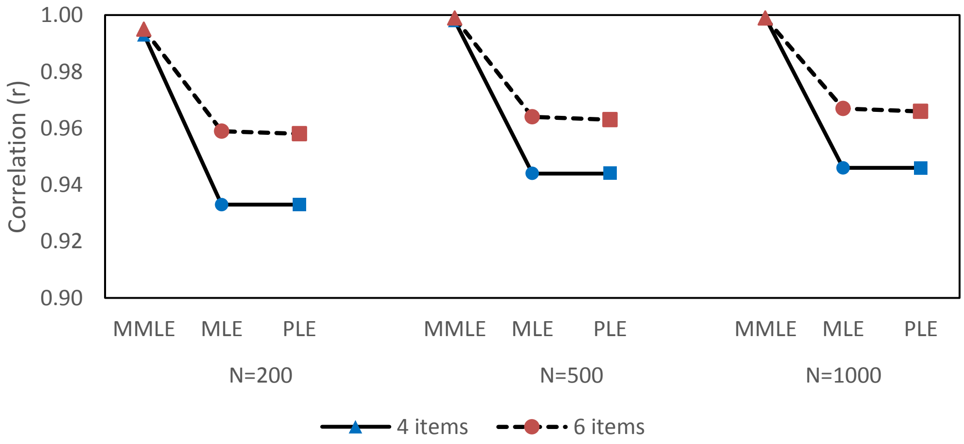


Figure 6. Correlation coefficients (*r*) for location parameters of unidimensional models with 4 and 6 items

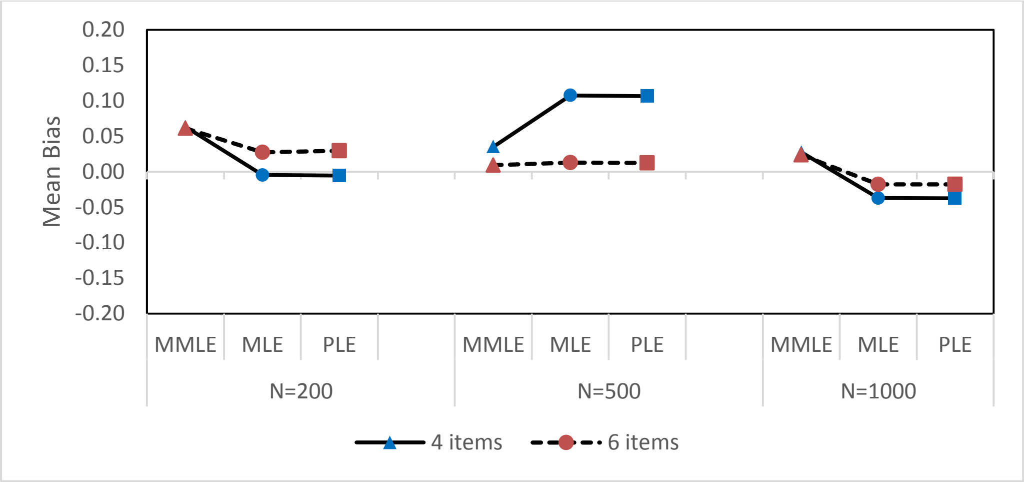
sample size of 500 with a value of 0.1067, while the 5 response category model with the same number of items and sample size yields the smallest one of 0.0024. Except for the condition with the largest mean bias of 0.1067, the absolute values of the mean bias for all 17 conditions are less than 0.04, which is quite close to 0. Also, most of the mean bias for slope parameters are smaller than those for location parameters of their corresponding models. It suggests that slope parameters of LMA models are recovered with less amount of errors than location parameters by PLE.

Figure 7 shows the mean bias results for slope parameters presented in Table 10. PLE estimates slope parameters as well as MMLE with almost the same and small mean bias values in 4 unidimensional models with 6 items, which are a binary response model with 500 persons and 5 response category models with 3 different sample sizes. The absolute values of all the mean bias for the 4 models are less than 0.02, which indicates that PLE recovered parameters used to simulate the data with small amount of errors.

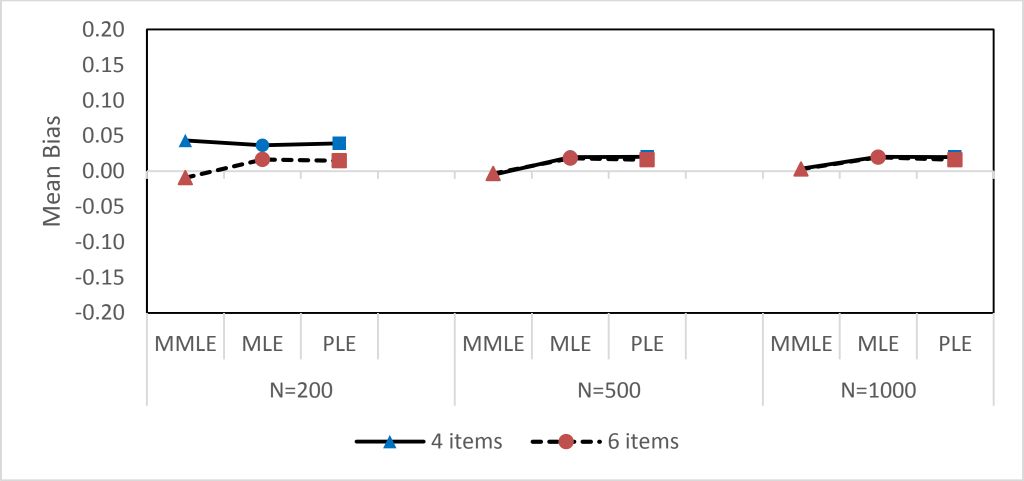
The mean RMSE for slope parameters of PLE ranges from 0.1464 to 0.5178. The largest mean RMSE values of PLE and MMLE commonly occur in the condition of 4 items, 5 response categories, and 200 persons (i.e., RMSE = 0.5118 for PLE, and RMSE = 0.5334 for MMLE). As the pattern found in the mean RMSE for location parameters, the mean RMSE values for slope parameters also increase as the number of categories increases given the same number of items and sample sizes and also larger mean RMSE values are found in the models associated with 4 items and 200 persons. Because the smaller RMSE reflects greater accuracy, this finding suggests again that the least amount of precision occurs with a short test, a small sample size, and a large number of categories in estimating slope parameters.

Figure 8 illustrates the mean RMSE results for slope parameters presented in Table 10.

(a) 2 categories



(b) 3 categories



(c) 5 categories

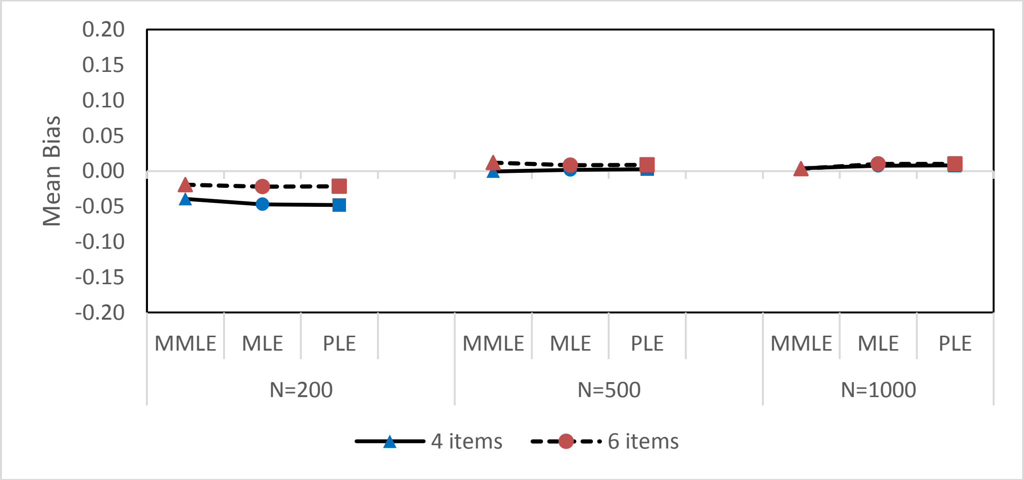
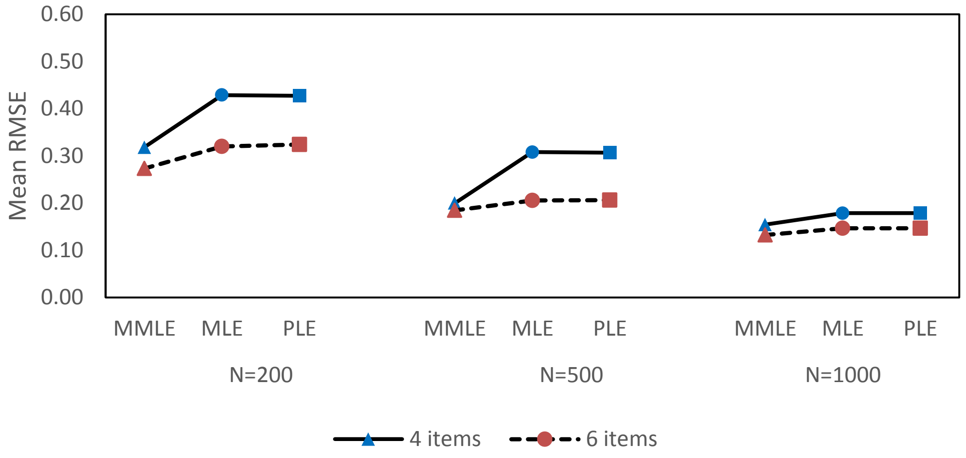
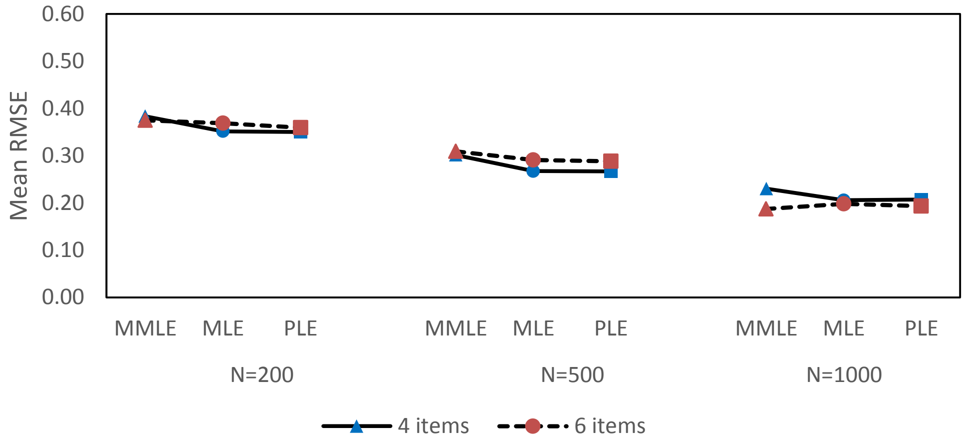


Figure 7. Mean bias for slope parameters of unidimensional models with 4 and 6 items

(a) 2 categories



(b) 3 categories



(c) 5 categories

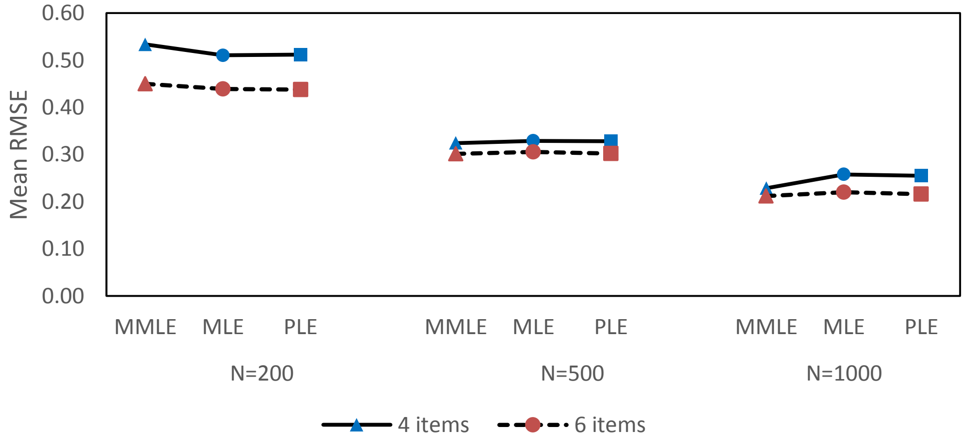


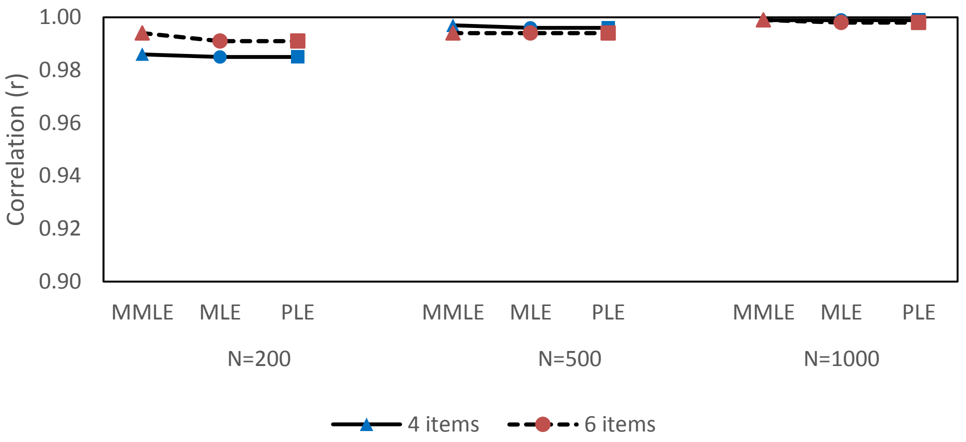
Figure 8. Mean RMSE for slope parameters of unidimensional models with 4 and 6 items

Except for binary response models, the mean RMSE values of PLE for all of 12 polytomous response models are equal to or smaller than those of MMLE. PLE outperforms MMLE in 5 models associated with 3 response categories and 2 models with 5 response categories and sample size of 200, resulting in smaller mean RMSE values than MMLE. It suggests that PLE recovered slope parameters of polytomous response models more accurately than MMLE. Overall, 6 items produced a smaller mean RMSE than 4 items and larger samples resulted in a smaller mean RMSE. This indicates that the amount of variability can be improved by longer tests and larger sample size.

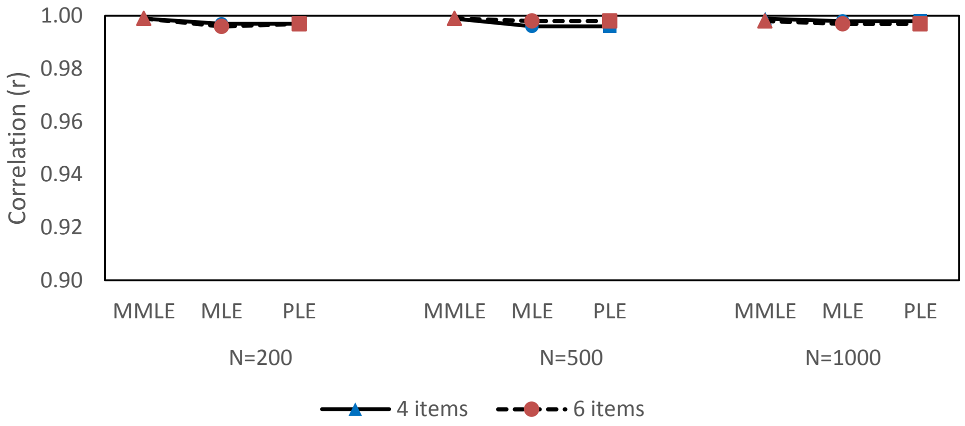
The correlation coefficients (r) between the parameters used to simulate the data and the estimated ones for slopes also presented in Table 10 and Figure 9. PLE yields consistently high correlations across all conditions, ranging from 0.983 to 0.999. Figure 9 shows that except for only one model associated with 5 response categories, 4 items, and 500 persons, the correlations of PLE are nearly identical to those of MMLE in all models with high coefficients. It is noticeable that the correlations for all 6 models with 3 category response categories and 5 models with 5 response categories are all 0.99. The correlations of PLE for slope parameters are higher than those for location parameters, indicating PLE performs better in estimating slope parameters than location parameters of LMA models.

In summary, Simulation study 1 provides three main findings for PLE of unidimensional LMA models with small numbers of items. First, PLE yields almost identical estimates to those from MLE. It is supported by small RMSDiff values and high correlations between PLE and MLE of LMA models. Second, based on mean bias and RMSE, PLE recovers parameters as well as MMLE of IRT models. Third, overall, PLE performs more excellently when it estimates slope parameters than location ones of polytomous items, which proves the success of the newly

(a) 2 categories



(b) 3 categories



(c) 5 categories

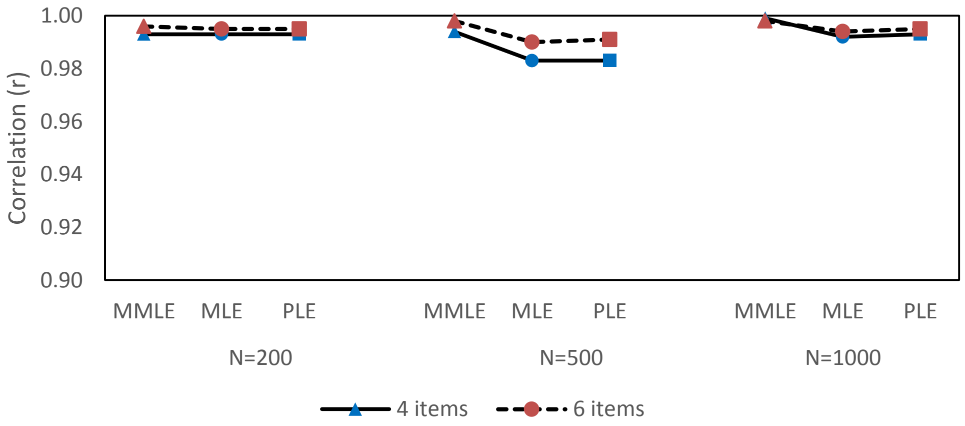


Figure 9. Correlation coefficients (*r*) for slope parameters of unidimensional models with 4 and 6 items

proposed Step 1 of the PLE algorithm in this thesis.

Unidimensional Models with Large Numbers of Items

As mentioned earlier, MLE fails to estimate parameters of LMA models when the number of items is large. Simulation study 2 was conducted to examine that PLE using Step 1 of the algorithm overcomes the limitation of MLE when fitting LMA models with large numbers of items. The item parameters of 18 unidimensional models with 20 and 50 items were estimated by PLE and MMLE. To evaluate parameter recovery, three evaluation criteria were computed for each item in all simulation conditions. As part of those computed values, Table 13 shows the bias values of parameters for unidimensional models with 20 items, 2 and 3 categories, and sample size of 1000. Once again, to simplify the interpretation of the evaluation criteria, the averaged values over items and categories are presented in this thesis.

Parameter Recovery of PLE with 20 and 50 items

Location parameters. Table 14 shows mean bias, RMSE and their standard deviations, and correlations between the true and estimated parameters for location parameters of unidimensional models with 20 and 50 items.

The mean bias values of PLE for location parameters are quite small across all polytomous response models with the minimum value of -0.0292 and the maximum value of -0.0001. Binary response models, however, have large mean bias values relative to polytomous response models, which range from -0.1347 to -0.0827. As found in the simulations with small numbers of items, polytomous response models with large numbers of items also resulted in smaller mean bias values of PLE than binary response models when the number of items are large.

Table 13. Bias of parameters for unidimensional models with 20 items, 2 and 3 categories, and sample size of 1000, by estimation method

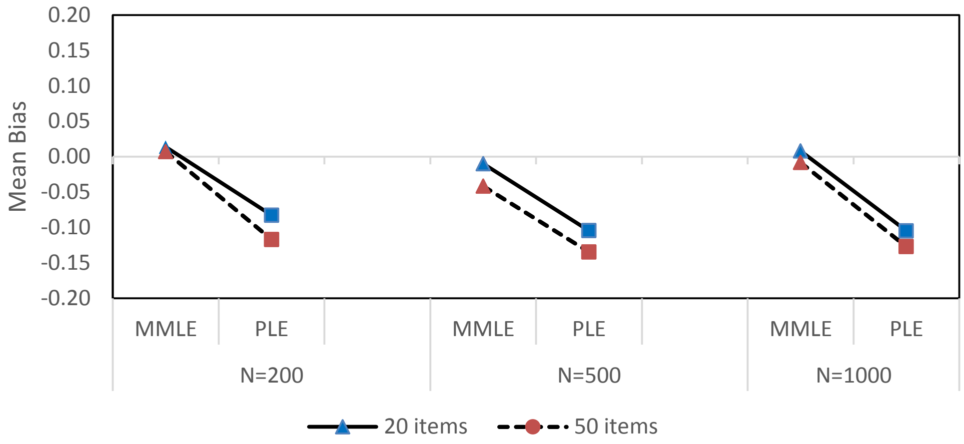
Item	Location		Slope		Item	Location		Slope	
	MMLE	PLE	MMLE	PLE		MMLE	PLE	MMLE	PLE
2 categories (N=1000)									
Item 1	-0.0031	-0.0515	0.0370	0.0153	Item 11	0.0257	-0.3329	-0.0001	-0.0398
Item 2	0.0096	-0.0049	0.0248	-0.0296	Item 12	-0.0238	-0.2742	0.0125	-0.0193
Item 3	-0.0054	-0.1718	0.0072	-0.0126	Item 13	-0.0131	-0.2852	0.0253	-0.0399
Item 4	0.0007	-0.0519	0.0245	-0.0593	Item 14	0.0241	-0.2061	-0.0091	-0.1336
Item 5	0.0179	-0.0773	-0.0497	-0.1558	Item 15	0.0124	-0.1668	-0.0128	-0.0374
Item 6	-0.0018	-0.2315	0.0245	-0.0512	Item 16	0.0070	-0.2089	0.0262	-0.0505
Item 7	0.0253	0.4900	0.0677	0.1707	Item 17	0.0230	-0.2988	-0.0120	-0.0996
Item 8	0.0249	-0.1506	0.0155	-0.1379	Item 18	-0.0094	0.0213	-0.0031	-0.0241
Item 9	0.0058	0.0191	-0.0049	-0.0646	Item 19	0.0133	-0.3007	-0.0202	-0.0741
Item 10	0.0056	0.1849	0.0084	-0.0225	Item 20	0.0231	-0.0029	0.0241	-0.0358
3 categories (N=1000)									
Item 1_2	0.0129	-0.1531	-0.0523	0.2076	Item 11_2	-0.0035	0.1400	0.0032	0.0116
Item 1_3	-0.0063	-0.1869	-0.0638	0.1942	Item 11_3	-0.0014	-0.2072	-0.0117	-0.0163
Item 2_2	0.0049	-0.0021	-0.0004	-0.1104	Item 12_2	0.0051	0.1945	0.0183	0.1621
Item 2_3	-0.0160	0.3282	0.0360	0.0627	Item 12_3	-0.0024	0.1002	-0.0144	0.1291
Item 3_2	0.0053	0.3637	0.0090	0.0349	Item 13_2	-0.0619	0.0604	0.0628	0.0094
Item 3_3	-0.0181	-0.4399	0.0079	-0.0868	Item 13_3	0.0255	0.0734	-0.0181	-0.0907
Item 4_2	-0.0078	-0.3348	-0.0057	0.1676	Item 14_2	0.0007	-0.0361	0.0022	0.0058
Item 4_3	-0.0170	0.1042	-0.0183	-0.1013	Item 14_3	0.0094	0.1269	0.0141	0.0128
Item 5_2	-0.0413	0.1905	0.0292	0.3613	Item 15_2	0.0064	-0.0932	-0.0075	-0.0017
Item 5_3	0.0260	0.0240	-0.0276	-0.2010	Item 15_3	-0.0172	0.1527	0.0063	0.0098
Item 6_2	-0.0137	-0.2957	-0.0259	-0.1726	Item 16_2	0.0601	-0.0003	0.0319	0.0350
Item 6_3	-0.0055	-0.0293	-0.0215	0.0902	Item 16_3	0.0817	-0.0183	0.0164	0.0106
Item 7_2	0.0048	-0.0637	0.0031	-0.0417	Item 17_2	0.0005	-0.0658	-0.0034	0.0778
Item 7_3	0.0082	-0.0109	-0.0040	0.0180	Item 17_3	-0.0106	-0.0746	-0.0028	0.0810
Item 8_2	0.0123	-0.0410	-0.0186	-0.1068	Item 18_2	0.0134	-0.0576	-0.0184	-0.1095
Item 8_3	-0.0335	0.2687	0.0182	0.1375	Item 18_3	0.0434	-0.0194	-0.0102	-0.0411
Item 9_2	-0.0340	-0.2533	-0.0729	0.0856	Item 19_2	-0.0197	0.0172	-0.0038	0.0007
Item 9_3	-0.0017	-0.3288	-0.0696	0.0931	Item 19_3	0.0051	-0.0482	-0.0001	0.0075
Item 10_2	-0.0579	-0.0272	0.0151	-0.0434	Item 20_2	-0.0194	-0.1025	-0.0423	0.0569
Item 10_3	0.0243	0.0266	-0.0174	0.0107	Item 20_3	0.0143	-0.0710	0.0080	0.0884

This pattern is illustrated more clearly in Figure 10. It shows that the big differences in mean bias between PLE and MMLE is found in all of the binary response models. PLE and

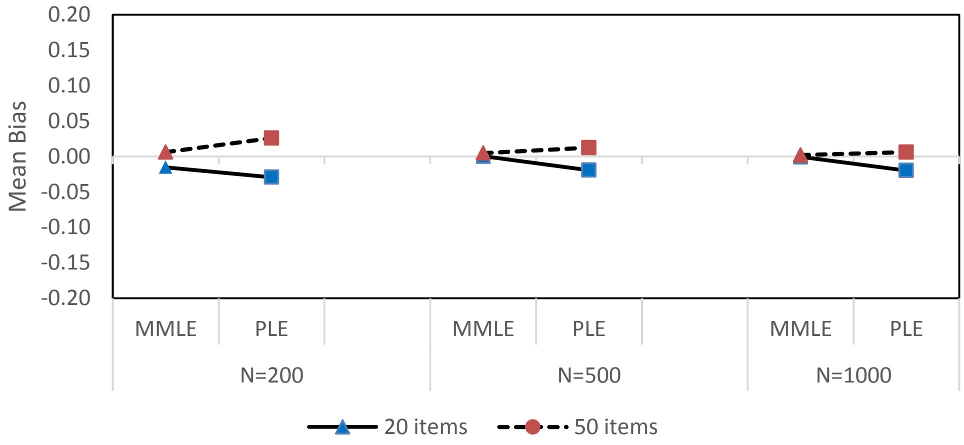
Table 14. Mean bias, RMSE and their (SDs) and correlation coefficients for location parameters of unidimensional models with 20 and 50 items

	20 items					50 items				
	Bias		RMSE		$r_{(\lambda,\ \hat{\lambda})}$	Bias		RMSE		$r_{(\lambda,\ \hat{\lambda})}$
2 categories										
					Sample size (N) =200					
MMLE	0.0135	(0.0921)	0.2925	(0.1699)	0.997	0.0072	(0.0540)	0.2611	(0.1101)	0.999
PLE	-0.0827	(0.0787)	0.2917	(0.1431)	0.996	-0.1171	(0.0582)	0.2768	(0.0983)	0.998
					Sample size (N) = 500					
MMLE	-0.0101	(0.0266)	0.1551	(0.0702)	0.999	-0.0417	(0.0548)	0.2582	(0.0853)	0.998
PLE	-0.1047	(0.1524)	0.2170	(0.0944)	0.999	-0.1347	(0.1311)	0.2148	(0.0993)	0.999
					Sample size (N) = 1000					
MMLE	0.0081	(0.0145)	0.1058	(0.0400)	0.999	-0.0084	(0.0508)	0.2532	(0.0850)	0.999
PLE	-0.1050	(0.1964)	0.2039	(0.1216)	0.999	-0.1273	(0.1441)	0.1909	(0.1025)	0.999
3 categories										
					Sample size (N) =200					
MMLE	-0.0152	(0.0737)	0.2862	(0.1307)	0.999	0.0060	(0.0862)	0.2597	(0.1198)	0.998
PLE	-0.0292	(0.1908)	0.3160	(0.1380)	0.986	0.0260	(0.1683)	0.3000	(0.1558)	0.987
					Sample size (N) = 500					
MMLE	0.0002	(0.0437)	0.1645	(0.0642)	0.999	0.0049	(0.0669)	0.1620	(0.0640)	0.998
PLE	-0.0193	(0.1846)	0.2233	(0.1143)	0.986	0.0125	(0.1410)	0.1931	(0.0929)	0.990
					Sample size (N) = 1000					
MMLE	-0.0006	(0.0271)	0.1196	(0.0488)	0.999	0.0021	(0.0524)	0.1101	(0.0446)	0.999
PLE	-0.0197	(0.1737)	0.1856	(0.1048)	0.987	0.0062	(0.1319)	0.1546	(0.0872)	0.991
5 categories										
					Sample size (N) =200					
MMLE	0.0140	(0.0863)	0.3186	(0.1349)	0.996	0.0095	(0.0984)	0.2906	(0.1062)	0.994
PLE	0.0140	(0.1245)	0.2988	(0.1352)	0.994	0.0096	(0.1215)	0.2748	(0.1076)	0.992
					Sample size (N) = 500					
MMLE	0.0032	(0.0450)	0.1869	(0.0682)	0.999	0.0013	(0.0568)	0.1819	(0.0628)	0.998
PLE	-0.0003	(0.0718)	0.1855	(0.0646)	0.997	0.0065	(0.1007)	0.1870	(0.0699)	0.993
					Sample size (N) = 1000					
MMLE	0.0041	(0.0426)	0.1278	(0.0491)	0.999	0.0028	(0.0407)	0.1218	(0.0442)	0.999
PLE	-0.0001	(0.0658)	0.1327	(0.0514)	0.998	0.0193	(0.1174)	0.1528	(0.0716)	0.990

(a) 2 categories



(b) 3 categories



(c) 5 categories

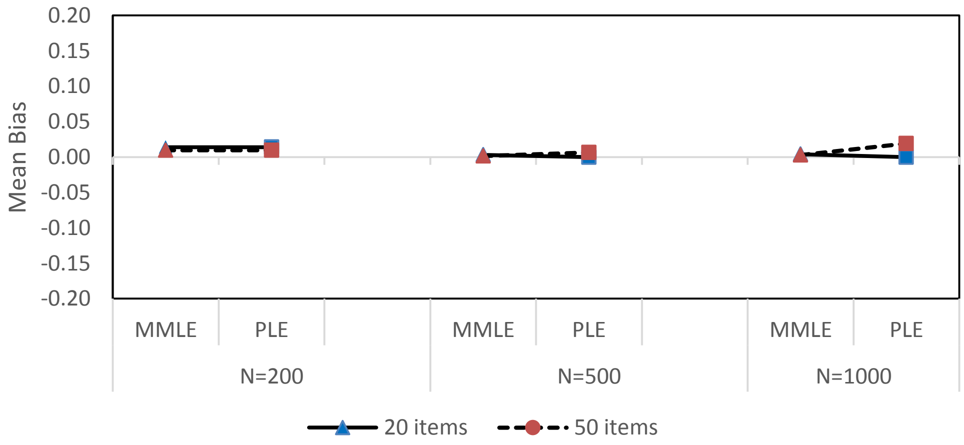


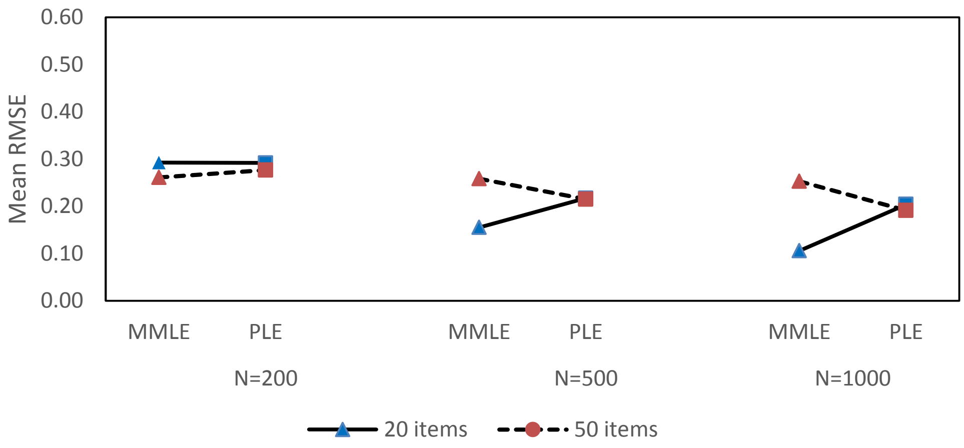
Figure 10. Mean bias for location parameters of unidimensional models with 20 and 50 items

MMLE, however, yield nearly identical mean bias for 5 category models. For 3 category models with 50 items, the mean bias of PLE gets closer to 0 as the sample size increases. The pattern shown in both short and long test lengths suggests that PLE recovered location parameters of polytomous response models better than those of binary response models regardless of test lengths.

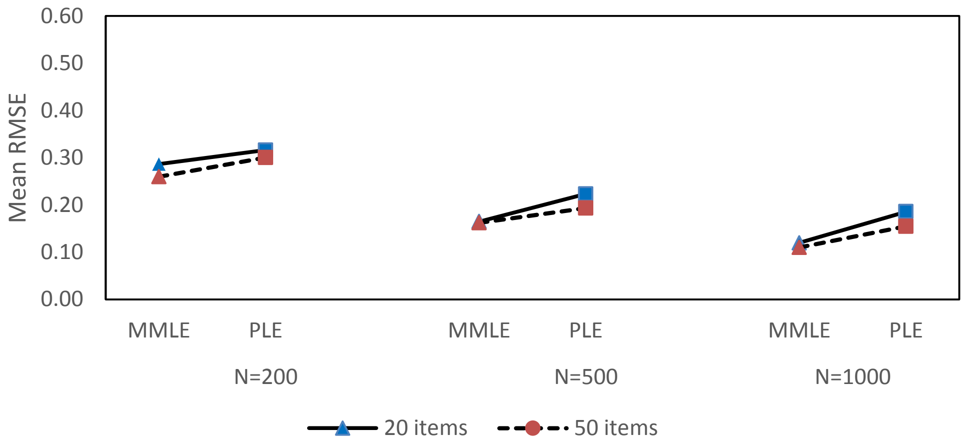
The mean RMSE for location parameters of PLE ranges from 0.1327 to 0.3160. The largest mean RMSE of PLE occur in the 3 category model with 20 items and 200 persons (i.e., RMSE = 0.3160) while the smallest does in the 5 category model with 20 items and 1000 persons (i.e., RMSE = 0.1327). Unlike the results found in the simulations with 4 and 6 items, it appears that the mean RMSE of PLE for location parameters doesn't increase as the number of categories increases given the same number of items and sample sizes. The mean RMSE values of PLE for 5 category models are smaller than those for 3 category models given the number of items and sample sizes are equal.

Figure 11 illustrates different patterns of mean RMSE for location parameters by the number of categories. For binary response models with 50 items, MMLE shows almost the same amount of estimation errors regardless of sample sizes, whereas the estimation errors of PLE decrease as sample sizes increase, yielding smaller mean RMSE than MMLE in the conditions of 50 items and sample sizes of 500 and 1000. For 3 category response models, MMLE yields smaller mean RMSE values than PLE in all conditions. For 5 category response models, PLE shows a bit smaller mean RMSE than MMLE in three conditions (i.e., 20 and 50 items with 200 persons and 20 items with 1000 persons) and almost identical ones in two conditions involving sample size of 500. Both PLE and MMLE show one clear pattern commonly in polytomous

(a) 2 categories



(b) 3 categories



(c) 5 categories

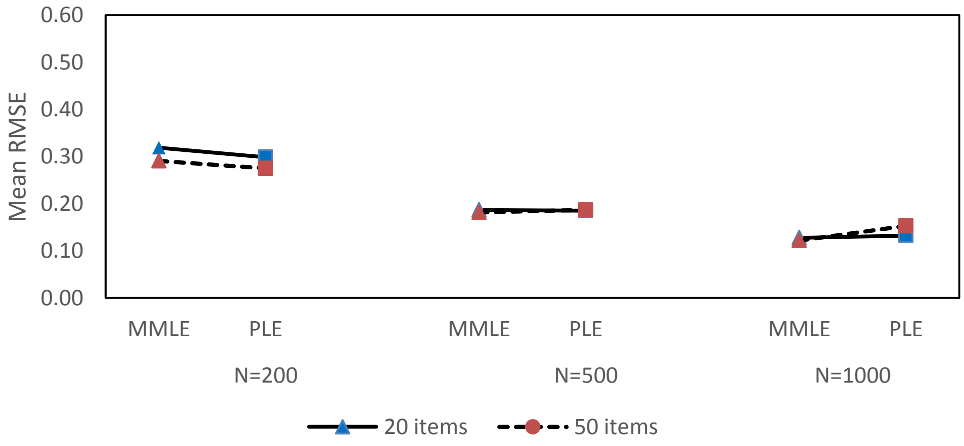


Figure 11. Mean RMSE for location parameters of unidimensional models with 20 and 50 items

response models, where larger sample sizes are associated with smaller mean RMSE when the number of items and categories are the same.

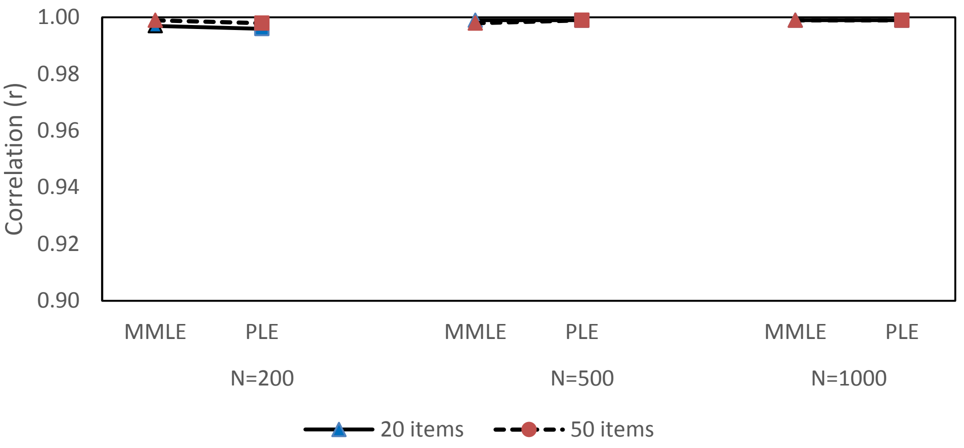
Table 14 also contains the correlation coefficients (r) between the true and estimated parameters for location parameters. MMLE yields the highest correlations across all conditions with 20 and 50 items, which is the same result as those found in simulations with 4 and 6 items. PLE also yields consistently high correlations across all conditions, ranging from 0.986 to 0.999.

The correlations are also illustrated in Figure 12. For binary response models, the correlations of PLE between the estimates and true parameters for location parameters are almost the same as those of MMLE with very high correlations of 0.99. The correlations of PLE for polytomous response models show a different pattern by test length. For 3 category response models, the test length of 50 items shows higher correlations than 20 items across all sample size while the opposite occurs in 5 category response models, but the correlations of PLE for both test lengths are still high with the coefficients of 0.99 in all 5 category response models.

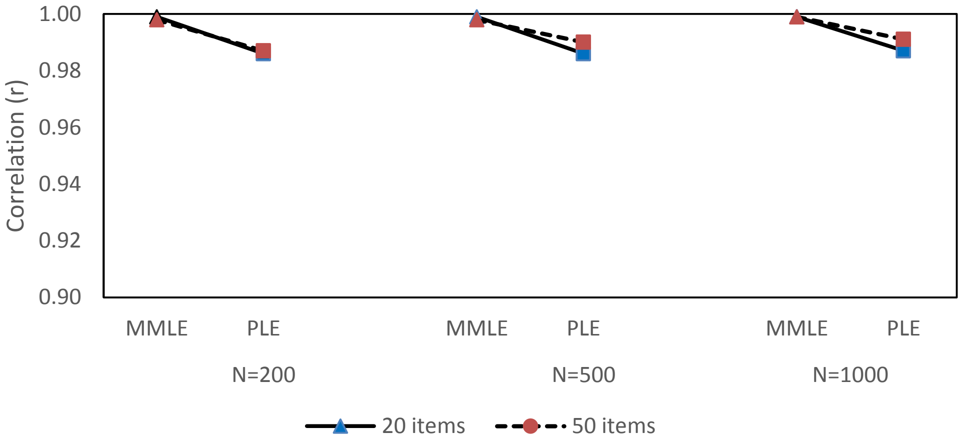
Slope parameters. Table 15 presents mean bias, mean RMSE and their standard deviations, and the correlations between the estimates $(\hat{\nu}_{ij})$ and true parameters (ν_{ij}) for slope parameters of unidimensional models with 20 and 50 items.

The mean bias of PLE for slope parameters are relatively small across all polytomous response models, ranging from -0.0057 to 0.0481 while the large mean bias is associated with binary response models, which is the same pattern found in the mean bias for location parameters. The large mean bias of PLE occur in two binary response models involving the sample size of 200 with values of -0.2177 and -0.1552 while the absolute values of mean bias are less than 0.05 for 3 category response models and 0.006 for 5 category response models.

(a) 2 categories



(b) 3 categories



(c) 5 categories

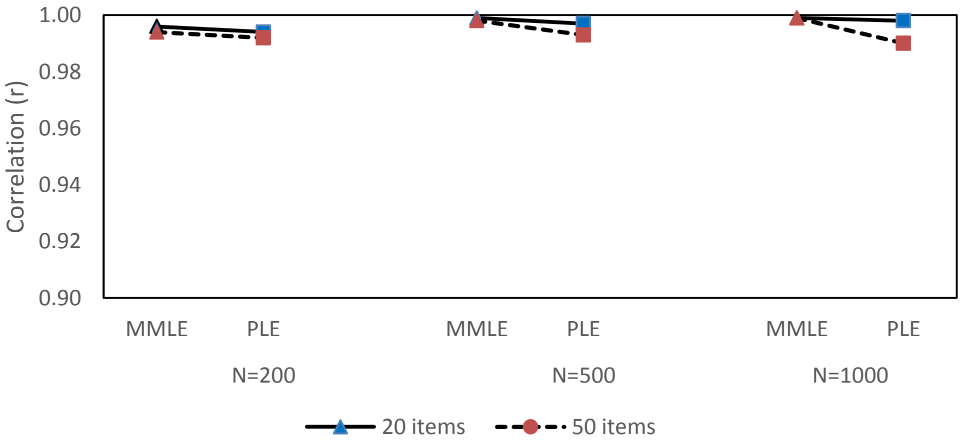


Figure 12. Correlation coefficients (*r*) for location parameters of unidimensional models with 20 and 50 items

Table 15. Mean bias, RMSE and their (SDs) and correlation coefficients for slope parameters of unidimensional models with 20 and 50 items

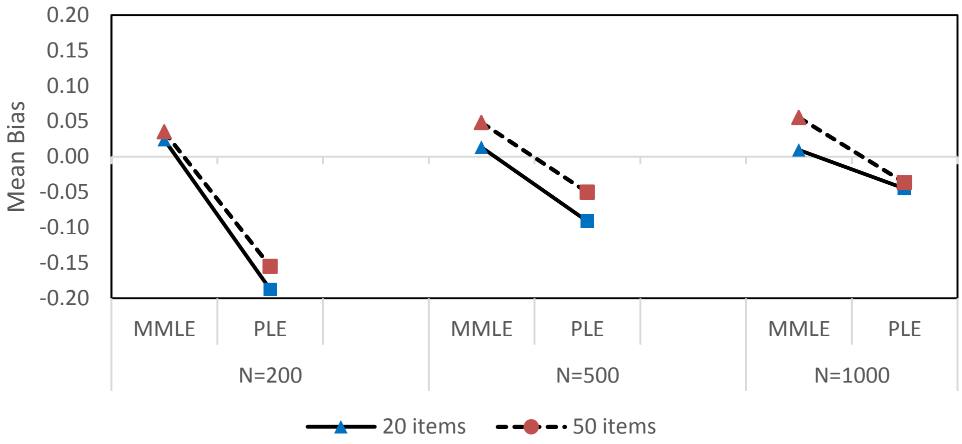
	20 items					50 items				
	Bias		RMSE		$r_{(\nu,\ \hat{\nu})}$	Bias		RMSE		$r_{(\nu,\ \hat{\nu})}$
<u>2 categories</u>										
					Sample size (N) =200					
MMLE	0.0237	(0.0778)	0.2948	(0.1256)	0.987	0.0351	(0.0520)	0.2308	(0.0737)	0.994
PLE	-0.1877	(0.0982)	0.3360	(0.0937)	0.979	-0.1552	(0.0731)	0.2494	(0.0678)	0.987
					Sample size (N) = 500					
MMLE	0.0133	(0.0322)	0.1694	(0.0638)	0.997	0.0480	(0.0413)	0.2361	(0.0785)	0.995
PLE	-0.0914	(0.0638)	0.1969	(0.0524)	0.989	-0.0503	(0.0464)	0.1558	(0.0433)	0.994
					Sample size (N) = 1000					
MMLE	0.0093	(0.0249)	0.1100	(0.0399)	0.998	0.0555	(0.0513)	0.2447	(0.0865)	0.992
PLE	-0.0451	(0.0675)	0.1394	(0.0449)	0.986	-0.0366	(0.0386)	0.1083	(0.0291)	0.995
<u>3 categories</u>										
					Sample size (N) =200					
MMLE	-0.0150	(0.0821)	0.3003	(0.1392)	0.999	0.0050	(0.0634)	0.2637	(0.1113)	0.999
PLE	0.0291	(0.1236)	0.3010	(0.1335)	0.994	0.0481	(0.2189)	0.3361	(0.2032)	0.976
					Sample size (N) = 500					
MMLE	-0.0112	(0.0425)	0.1794	(0.0689)	0.999	0.0062	(0.0372)	0.1472	(0.0528)	0.999
PLE	0.0239	(0.1093)	0.2061	(0.0807)	0.994	0.0391	(0.2002)	0.2269	(0.1504)	0.980
					Sample size (N) = 1000					
MMLE	-0.0062	(0.0278)	0.1192	(0.0438)	0.999	0.0034	(0.0335)	0.1041	(0.0367)	0.999
PLE	0.0260	(0.1094)	0.1559	(0.0734)	0.995	0.0325	(0.1899)	0.1848	(0.1436)	0.982
<u>5 categories</u>										
					Sample size (N) =200					
MMLE	0.0059	(0.0793)	0.3489	(0.1388)	0.998	0.0024	(0.0601)	0.3095	(0.1304)	0.998
PLE	-0.0057	(0.1250)	0.3407	(0.1397)	0.992	0.0032	(0.2353)	0.3840	(0.1745)	0.973
					Sample size (N) = 500					
MMLE	-0.0027	(0.0469)	0.2133	(0.0921)	0.999	-0.0026	(0.0569)	0.1908	(0.0693)	0.999
PLE	-0.0042	(0.0978)	0.2266	(0.0940)	0.994	0.0018	(0.2225)	0.2820	(0.1283)	0.975
					Sample size (N) = 1000					
MMLE	-0.0021	(0.0306)	0.1414	(0.0610)	0.999	-0.0084	(0.0676)	0.1430	(0.0578)	0.999
PLE	-0.0029	(0.0870)	0.1624	(0.0662)	0.995	0.0029	(0.1087)	0.1654	(0.0754)	0.994

As shown in the mean bias for location parameters with 20 and 50 items, Figure 13 also illustrates that the big differences in mean bias for slope parameters between PLE and MMLE exist in all binary response models. The largest difference between two estimation methods is associated with the smallest sample size, but it appears that the gap decreases as sample size gets larger. The mean bias of PLE for 5 category response models are quite small and almost equal to those of MMLE.

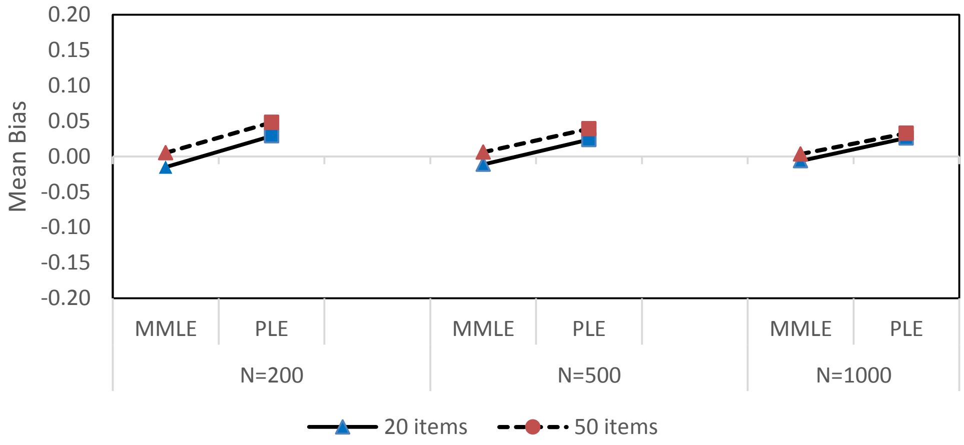
The largest mean RMSE of PLE occur in the 5 category response model with 50 items and 200 persons (i.e., RMSE = 0.3840) while the smallest does in the binary response model with 50 items and 1000 persons (i.e., RMSE = 0.1083). For polytomous response models, it appears that the mean RMSE of both estimation methods for slope parameters gets large as the number of categories increases given the same number of items and sample sizes, which is the same pattern of mean RMSE found in simulations with 4 and 6 items. The larger mean RMSE appears to be more associated with the models involving 20 items and 200 persons. It suggests again that more estimation errors occur with shorter test lengths, smaller sample size, and more categories.

Figure 14 illustrates the mean RMSE for slope parameters by the number of category. The mean RMSE for slope parameters of binary response models shows a similar pattern to what has been found in the location parameters of the models. The estimation errors of PLE decrease with larger sample sizes in all binary response models, whereas those of MMLE are similar across the binary response models involving 50, regardless of sample sizes. Different patterns by test length are also found in polytomous response models. With 20 items, the mean RMSE of PLE and MMLE are close or nearly equal to each other. PLE, however, appears to have larger mean RMSE than MMLE in polytomous response models with 50 items.

(a) 2 categories



(b) 3 categories



(c) 5 categories

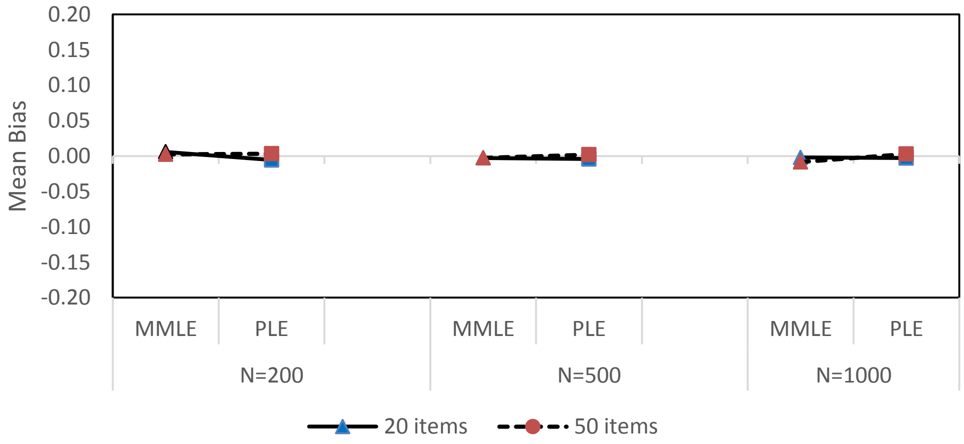
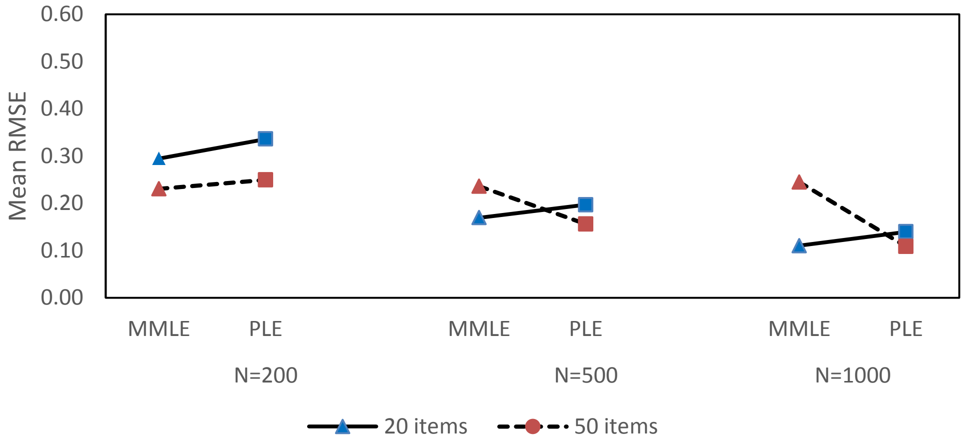
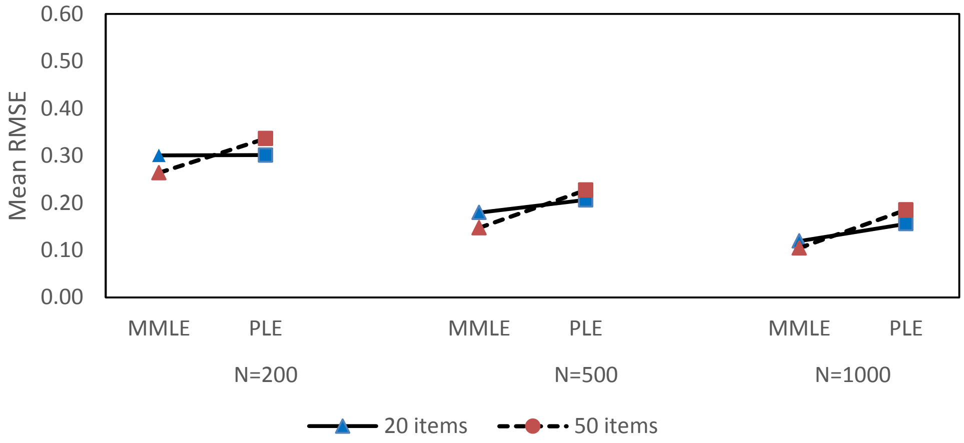


Figure 13. Mean bias for slope parameters of unidimensional models with 20 and 50 items

(a) 2 categories



(b) 3 categories



(c) 5 categories

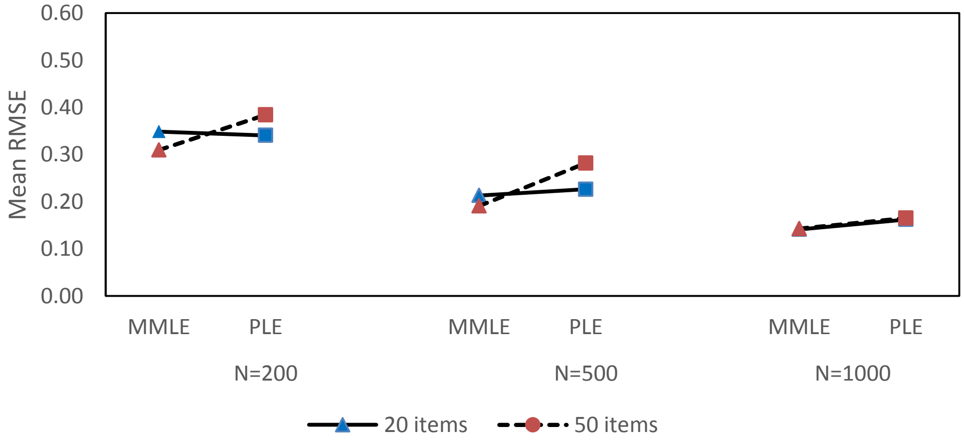


Figure 14. Mean RMSE for slope parameters of unidimensional models with 20 and 50 items

The correlation coefficients (r) between the estimated parameters and true parameters for slope parameters are also presented in Table 15. Again, MMLE yields the highest correlations across all conditions with 20 and 50 items and PLE also yields high correlations across all conditions with coefficients of 0.973 to 0.995.

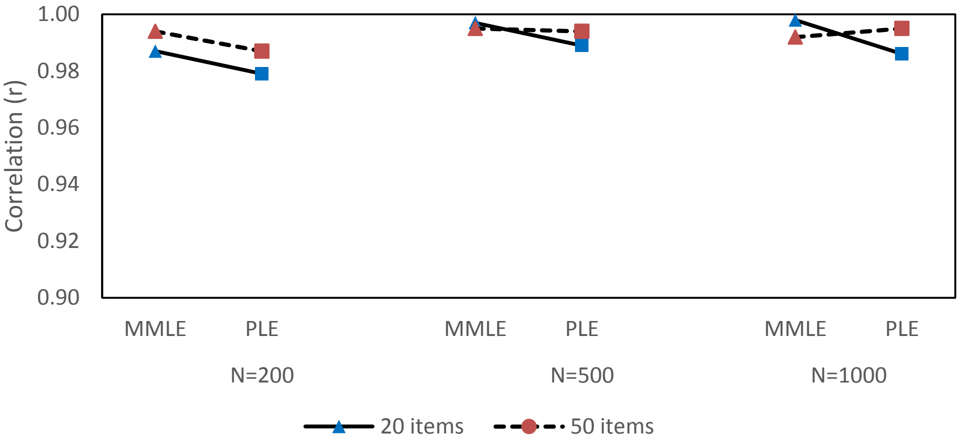
The correlations are also illustrated in Figure 15. For binary response models with 50 items and sample sizes of 500 and 100, the correlations of PLE between the estimates and true parameters for slope parameters are nearly equal to or larger than those of MMLE. For polytomous response models with 20 items, PLE yields as high correlations as MMLE with a coefficient of 0.99. On the other hand, the magnitude of correlations of PLE for polytomous models with 50 items are a bit smaller than MMLE. The correlations of PLE, however, increase as sample sizes increase in all polytomous models.

In summary, Simulation study 2 supports that the parameters of unidimensional LMA models with large numbers of items are successfully estimated by Step 1 of the PLE algorithm, overcoming the limitation of MLE of LMA models. For the simulation study, two findings can be summarized. First, PLE yields nearly unbiased item parameter estimates and very high correlations between the estimates and the parameters used to simulate the data in most of the conditions. Second, PLE performs as well as MMLE for unidimensional models with large numbers of items in terms of parameter recovery.

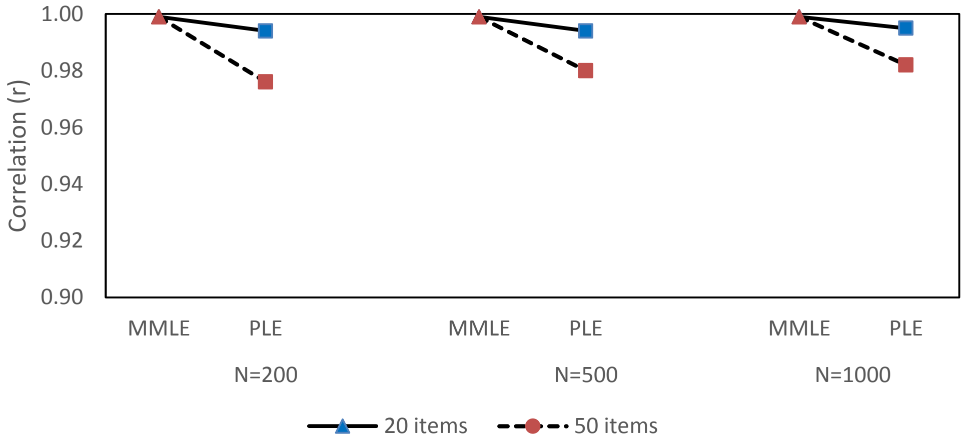
Multidimensional Models with Small Numbers of Items

This section describes how well PLE performs using the full algorithm for multidimensional models with small numbers of items. For 2 dimensional models, the number of items were varied with 4 and 6, but fixed to 6 items for 3 dimensional models. The performance of PLE are compared with MLE in terms of parameter recovery; that is, MLE of MIRT is a

(a) 2 categories



(b) 3 categories



(c) 5 categories

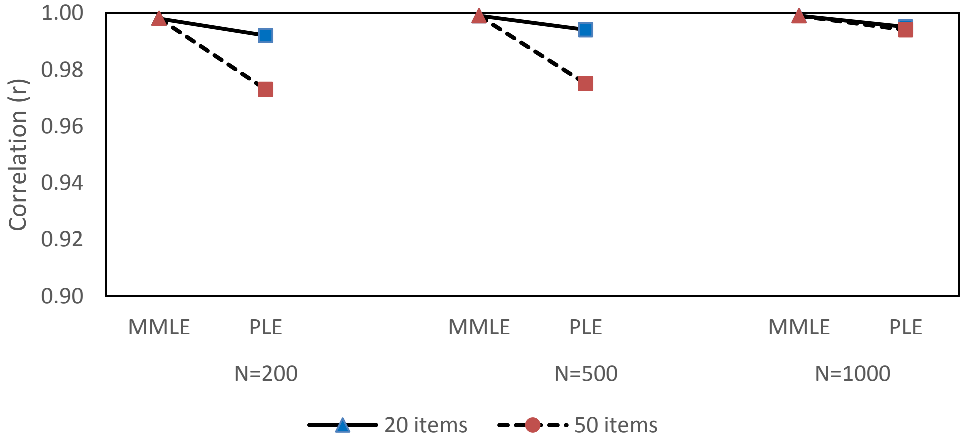


Figure 15. Correlation coefficients (*r*) for slope parameters of unidimensional models with 20 and 50 items

baseline to compare PLE of LMA. PLE of LMA should not do better than model used to simulate the data. As part of the computed bias and RMSE values for each item in all multidimensional models with small numbers of items, Table 16 contains the bias values of parameters for two-dimensional models with 6 items and sample size of 1000.

Parameter Recovery of PLE and MLE

Two-dimensional models. Tables 17 and 18 contains mean bias, RMSE and their standard deviations for location (λ_{ij}) and slope parameters (ν_{ij}) of two-dimensional models with 4 and 6 items by the number of categories. The correlations between the estimates and the parameters used to simulate the data are also shown in the tables.

Tables 17 and 18 show that mean bias, mean RMSE, and correlations of PLE are nearly identical to those of MLE across all conditions, indicating that PLE recovered true parameters with the same amount of errors as MLE in all of two-dimensional models with 4 and 6 items and performed very similarly to MLE in estimating parameters of LMA models.

For two-dimensional binary response models, both PLE and MLE show small mean bias for location parameters, ranging from -0.0264 to 0.0392. On the other hand, the mean bias for slope parameters are slightly large relative to those for location parameters, ranging from -0.1559 to 0.1851. The parameters for 6 binary items tend to be underestimated while those for 4 binary items tend to be overestimated by PLE and MLE.

For location parameters of two-dimensional polytomous response models, both PLE and MLE resulted in nearly unbiased location parameter estimates, where the absolute values of mean bias are less than 0.05 across all two-dimensional polytomous response models. This pattern also appears to be true for slope parameters of 5 response category models. However, the

Table 16. Bias of parameters for two-dimensional models with 6 items and sample size of 1000, by estimation method

Item	Location		Slope		Item	Location		Slope	
	MLE	PLE	MLE	PLE		MLE	PLE	MLE	PLE
2 categories (N=1000)									
Item 1	0.0058	0.0057	0.1615	0.1607	Item 4	-0.0582	-0.0582	0.0061	0.0049
Item 2	0.0029	0.0030	0.0537	0.0519	Item 5	-0.0660	-0.0660	-0.0190	-0.0198
Item 3	0.0100	0.0099	0.0392	0.0384	Item 6	-0.0525	-0.0525	-0.0402	-0.0406
3 categories (N=1000)									
Item 1_2	-0.2847	-0.2849	0.1729	0.1671	Item 4_2	-0.1089	-0.1094	0.1158	0.1064
Item 1_3	-0.1086	-0.1061	-0.0172	-0.0122	Item 4_3	-0.0777	-0.0794	-0.0383	-0.0337
Item 2_2	-0.1500	-0.1489	-0.1023	-0.1010	Item 5_2	0.5147	0.5159	-0.1175	-0.1189
Item 2_3	0.1856	0.1858	0.0222	0.0185	Item 5_3	-0.2458	-0.2470	0.0773	0.0774
Item 3_2	0.1114	0.1132	0.0418	0.0436	Item 6_2	-0.2393	-0.2374	-0.1821	-0.1741
Item 3_3	0.0460	0.0422	-0.1119	-0.1175	Item 6_3	-0.1496	-0.1509	0.0615	0.0612
5 categories (N=1000)									
Item 1_2	-0.2476	-0.2486	0.0468	0.0437	Item 4_2	0.3852	0.3822	-0.1214	-0.1088
Item 1_3	-0.1564	-0.1583	-0.0111	-0.0164	Item 4_3	0.1557	0.1539	-0.0453	-0.0349
Item 1_4	0.0832	0.0828	0.1144	0.1092	Item 4_4	0.3035	0.3018	-0.1253	-0.1165
Item 1_5	-0.0548	-0.0544	0.1459	0.1401	Item 4_5	-0.1742	-0.1725	0.0398	0.0397
Item 2_2	-0.0769	-0.0780	-0.0550	-0.0540	Item 5_2	-0.1068	-0.1057	0.0600	0.0562
Item 2_3	0.0295	0.0272	-0.0569	-0.0558	Item 5_3	-0.1916	-0.1905	0.0740	0.0696
Item 2_4	-0.0198	-0.0192	0.0195	0.0191	Item 5_4	-0.6859	-0.6884	-0.2046	-0.1744
Item 2_5	-0.0558	-0.0566	-0.0191	-0.0190	Item 5_5	-0.1488	-0.1500	0.0313	0.0336
Item 3_2	-0.0780	-0.0814	0.3279	0.3204	Item 6_2	0.7112	0.7132	-0.0385	-0.0466
Item 3_3	-0.1847	-0.1829	-0.2952	-0.2878	Item 6_3	0.0548	0.0578	0.0878	0.0831
Item 3_4	-0.2398	-0.2387	-0.1570	-0.1499	Item 6_4	-0.3357	-0.3358	0.0011	0.0046
Item 3_5	0.2409	0.2485	-0.1785	-0.1682	Item 6_5	-0.1307	-0.1302	0.0379	0.0383

mean bias for 3 response category models appears to be somewhat different between 4 and 6 items. With 6 items and different sample sizes, the conditions resulted in small mean bias for slope parameters, ranging from −0.0069 to 0.0109 for PLE and −0.0065 to 0.0138 for MLE, and the values of the mean bias decrease as sample size increases. On the other hand, the largest mean bias values of PLE and MLE for slope parameters of 3 response category models are

Table 17. Mean bias, RMSE and their (SDs) and correlation coefficients for location parameters of 2 dimensional models with 4 and 6 items

	4 items					6 items				
	Bias		RMSE		$r_{(\lambda,\ \hat{\lambda})}$	Bias		RMSE		$r_{(\lambda,\ \hat{\lambda})}$
2 categories										
					Sample size (N) =200					
MLE	0.0193	(0.0386)	0.1294	(0.0508)	0.995	-0.0213	(0.0646)	0.1430	(0.0219)	0.989
PLE	0.0192	(0.0386)	0.1293	(0.0508)	0.995	-0.0213	(0.0644)	0.1427	(0.0218)	0.989
					Sample size (N) = 500					
MLE	0.0378	(0.0181)	0.0887	(0.0254)	0.999	-0.0252	(0.0511)	0.1011	(0.0221)	0.993
PLE	0.0378	(0.0181)	0.0887	(0.0254)	0.999	-0.0252	(0.0512)	0.1012	(0.0221)	0.992
					Sample size (N) = 1000					
MLE	0.0392	(0.0149)	0.0700	(0.0134)	0.999	-0.0263	(0.0360)	0.0784	(0.0133)	0.996
PLE	0.0392	(0.0149)	0.0700	(0.0134)	0.999	-0.0264	(0.0360)	0.0784	(0.0133)	0.996
3 categories										
					Sample size (N) =200					
MLE	-0.0384	(0.2757)	0.2949	(0.1350)	0.955	-0.0374	(0.2439)	0.3415	(0.1096)	0.961
PLE	-0.0385	(0.2768)	0.2948	(0.1355)	0.955	-0.0373	(0.2455)	0.3420	(0.1115)	0.961
					Sample size (N) = 500					
MLE	-0.0342	(0.2988)	0.2665	(0.1571)	0.948	-0.0431	(0.2120)	0.2424	(0.1020)	0.970
PLE	-0.0340	(0.2999)	0.2673	(0.1576)	0.948	-0.0432	(0.2142)	0.2396	(0.1029)	0.969
					Sample size (N) = 1000					
MLE	-0.0398	(0.2850)	0.2445	(0.1511)	0.953	-0.0422	(0.2268)	0.2262	(0.1148)	0.965
PLE	-0.0399	(0.2853)	0.2446	(0.1513)	0.953	-0.0422	(0.2270)	0.2260	(0.1152)	0.965
5 categories										
					Sample size (N) =200					
MLE	0.0368	(0.2713)	0.3792	(0.1526)	0.948	-0.0289	(0.2718)	0.3491	(0.1534)	0.956
PLE	0.0384	(0.2698)	0.3794	(0.1520)	0.948	-0.0284	(0.2726)	0.3502	(0.1546)	0.956
					Sample size (N) = 500					
MLE	0.0270	(0.2872)	0.3076	(0.1706)	0.949	-0.0335	(0.2685)	0.2822	(0.1613)	0.958
PLE	0.0268	(0.2880)	0.3087	(0.1700)	0.948	-0.0334	(0.2685)	0.2823	(0.1611)	0.958
					Sample size (N) = 1000					
MLE	0.0257	(0.2906)	0.2623	(0.1735)	0.951	-0.0385	(0.2719)	0.2530	(0.1665)	0.959
PLE	0.0257	(0.2906)	0.2624	(0.1734)	0.951	-0.0385	(0.2724)	0.2534	(0.1669)	0.958

Table 18. Mean bias, RMSE and their (SDs) and correlation coefficients for slope parameters of 2 dimensional models with 4 and 6 items

	4 items					6 items				
	Bias		RMSE		$r_{(\nu,\ \hat{\nu})}$	Bias		RMSE		$r_{(\nu,\ \hat{\nu})}$
<u>2 categories</u>										
					Sample size (N) =200					
MLE	0.0930	(0.2965)	0.6712	(0.0744)	0.962	-0.1506	(0.1791)	0.4976	(0.0702)	0.994
PLE	0.0948	(0.2962)	0.6756	(0.0740)	0.962	-0.1559	(0.1796)	0.4944	(0.0695)	0.993
					Sample size (N) = 500					
MLE	0.1851	(0.1767)	0.6193	(0.0645)	0.989	-0.0912	(0.1457)	0.4465	(0.0818)	0.958
PLE	0.1844	(0.1763)	0.6184	(0.0645)	0.989	-0.0932	(0.1461)	0.4472	(0.0824)	0.958
					Sample size (N) = 1000					
MLE	0.1463	(0.1169)	0.6359	(0.1033)	0.995	0.0335	(0.0718)	0.2640	(0.0306)	0.991
PLE	0.1463	(0.1170)	0.6361	(0.1033)	0.995	0.0326	(0.0717)	0.2652	(0.0307)	0.991
<u>3 categories</u>										
					Sample size (N) =200					
MLE	0.1223	(0.1090)	0.5627	(0.1165)	0.998	0.0138	(0.1568)	0.4622	(0.0746)	0.993
PLE	0.1262	(0.1116)	0.5533	(0.1182)	0.998	0.0109	(0.1525)	0.4598	(0.0782)	0.993
					Sample size (N) = 500					
MLE	0.0121	(0.0391)	0.3530	(0.0866)	0.999	0.0061	(0.1092)	0.2724	(0.0477)	0.996
PLE	0.0051	(0.0424)	0.3515	(0.0844)	0.999	0.0087	(0.1097)	0.2743	(0.0498)	0.996
					Sample size (N) = 1000					
MLE	0.0683	(0.1176)	0.2914	(0.0653)	0.998	-0.0065	(0.1072)	0.2310	(0.0443)	0.996
PLE	0.0686	(0.1166)	0.2910	(0.0640)	0.998	-0.0069	(0.1046)	0.2298	(0.0437)	0.996
<u>5 categories</u>										
					Sample size (N) =200					
MLE	-0.0318	(0.1797)	0.6275	(0.2426)	0.989	-0.0373	(0.1897)	0.5015	(0.1553)	0.981
PLE	-0.0348	(0.1819)	0.6324	(0.2516)	0.989	-0.0364	(0.1885)	0.5050	(0.1564)	0.981
					Sample size (N) = 500					
MLE	-0.0172	(0.1123)	0.4574	(0.1859)	0.997	-0.0193	(0.1304)	0.3435	(0.1022)	0.990
PLE	-0.0164	(0.1198)	0.4574	(0.1906)	0.997	-0.0180	(0.1242)	0.3434	(0.1010)	0.991
					Sample size (N) = 1000					
MLE	0.0019	(0.0876)	0.3121	(0.1359)	0.997	-0.0134	(0.1298)	0.2635	(0.1088)	0.990
PLE	0.0017	(0.0891)	0.3131	(0.1375)	0.997	-0.0114	(0.1237)	0.2611	(0.1075)	0.991

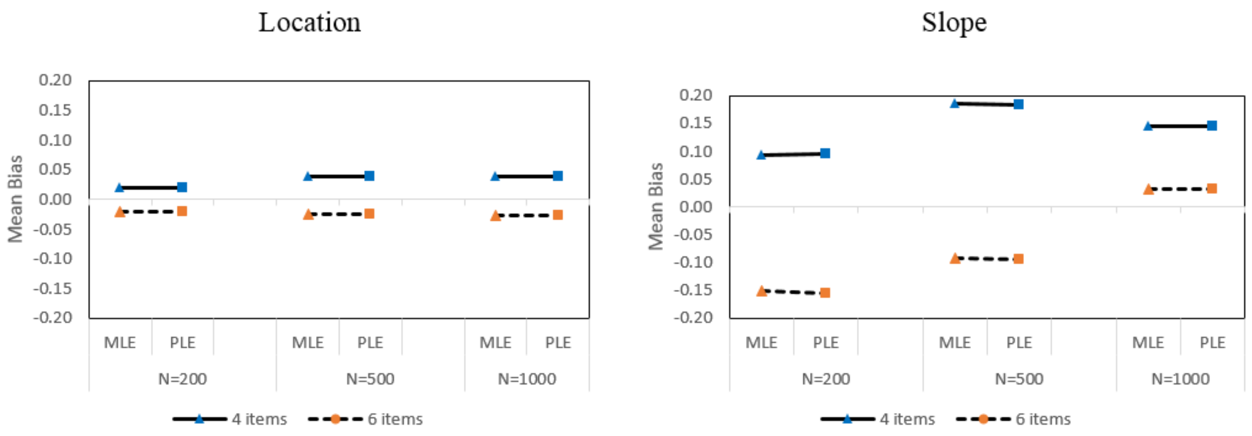
commonly found in the conditions with 4 items and sample size of 200, which are 0.1262 for PLE and 0.1223 for MLE, and sample size doesn't have an influence on getting smaller mean bias for the models with 4 items.

Figures 16 illustrates the mean bias for location and slope parameters of PLE and MLE for two-dimensional models with 4 and 6 items, and it also supports that the mean bias of PLE and MLE are nearly equal for all two-dimensional models with small numbers of items. When comparing mean bias for location and slope parameters of polytomous response models involving 6 items, slope parameter estimates of PLE and MLE are closer to true parameters than location parameter estimates of two estimation methods.

For two-dimensional binary response models, the mean RMSE of PLE and MLE for slope parameters are large relative to those for location parameters, indicating that PLE and MLE recovered location parameters better than slope parameters for two-dimensional binary response models. Overall, the mean RMSE of PLE and MLE for location and slope parameters of two-dimensional polytomous response models get smaller as the number of items and sample sizes increase, whereas they get larger as the number of categories increases. The condition of 5 response categories, 4 items, and 200 persons resulted in the largest mean RMSE for location and slope parameters with values of 0.3794 and 0.6324 for PLE and 0.3792 and 0.6275 for MLE. On the contrary, the smallest values of mean RMSE for location and slope parameters were found in the condition of 3 response categories, 6 items, and 1000 persons, which are 0.2260 and 0.2298 for PLE and 0.2262 and 0.2310 for MLE.

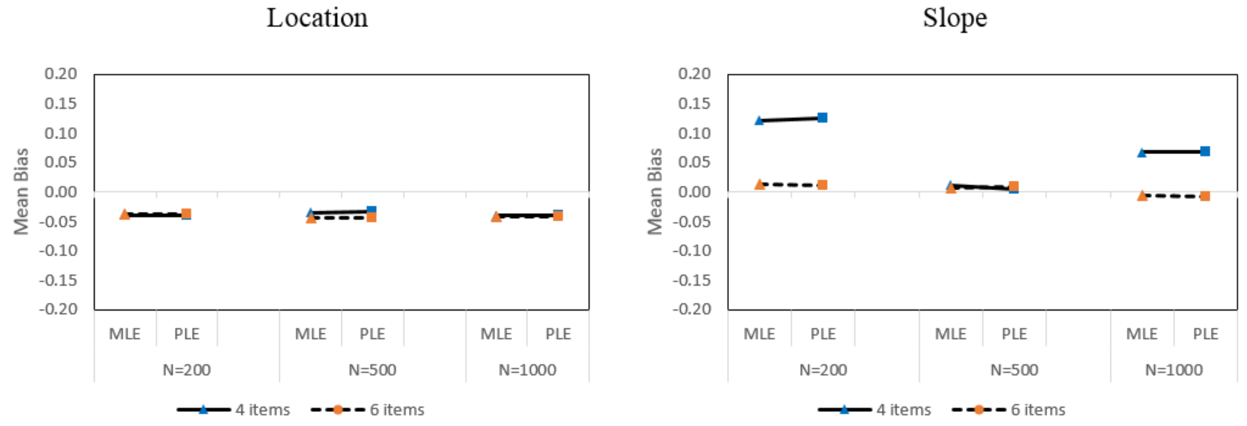
Figures 17 shows the mean RMSE for location and slope parameters of PLE and MLE for two-dimensional models with 4 and 6 items. The test lengths of 4 items and 6 items show similar variability in location parameter estimates of polytomous response models, whereas the

(a) 2 categories


Location

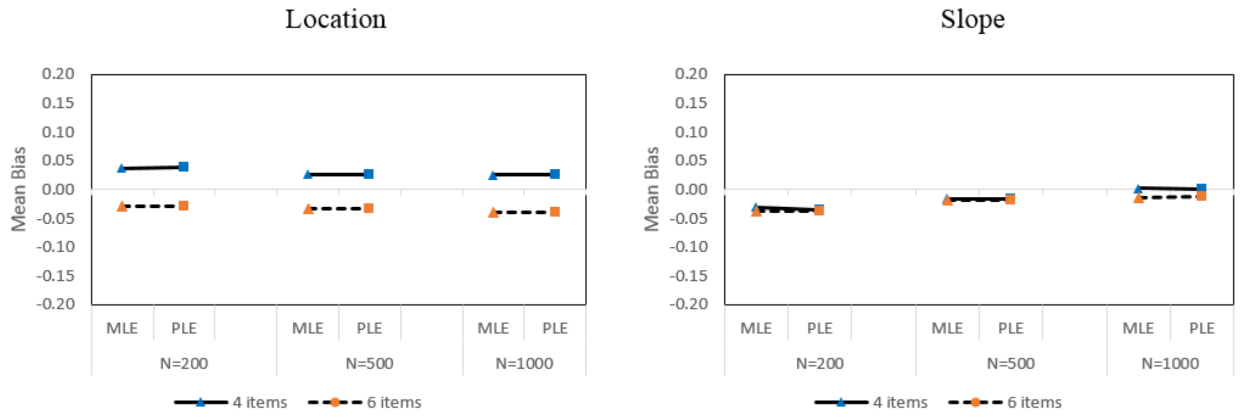
Slope

(b) 3 categories


Location

Slope

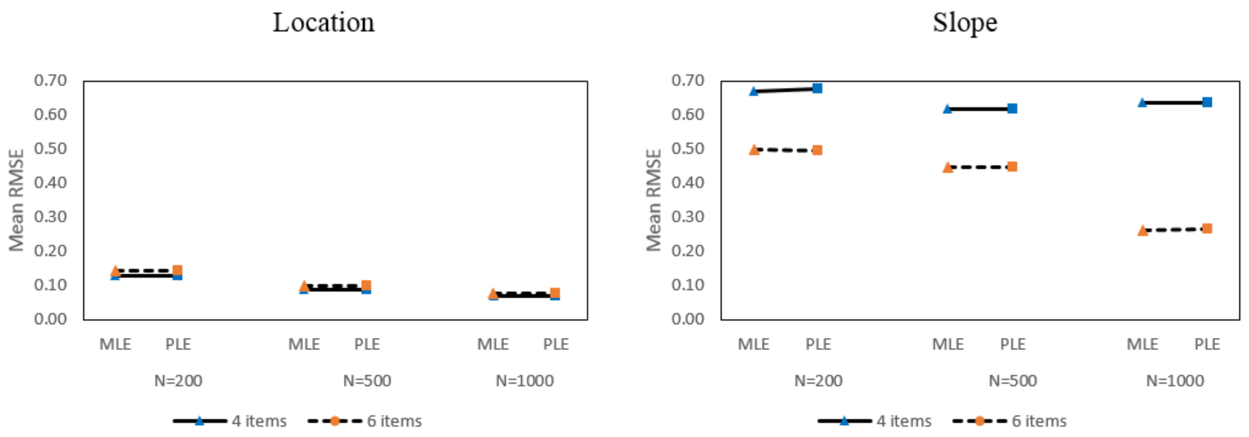
(c) 5 categories


Location

Slope

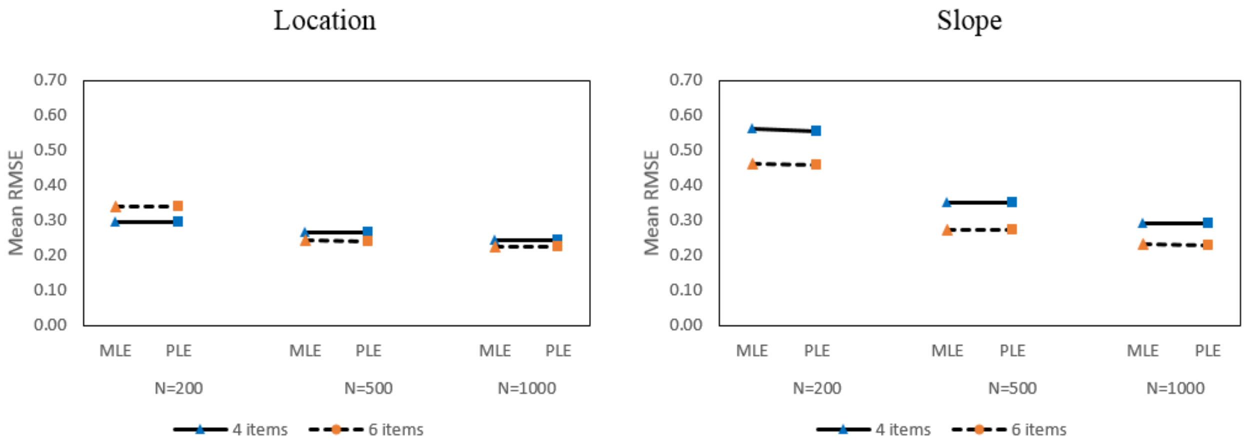
Figure 16. Mean bias for location and slope parameters of 2 dimensional models with 4 and 6 items

(a) 2 categories


Location

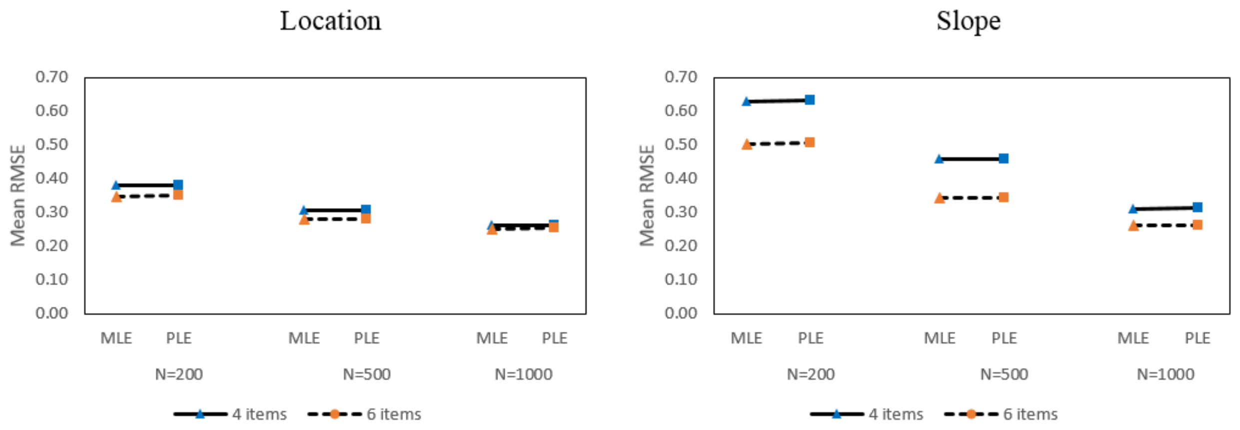
Slope

(b) 3 categories


Location

Slope

(c) 5 categories


Location

Slope

Figure 17. Mean RMSE for location and slope parameters of 2 dimensional models with 4 and 6 items

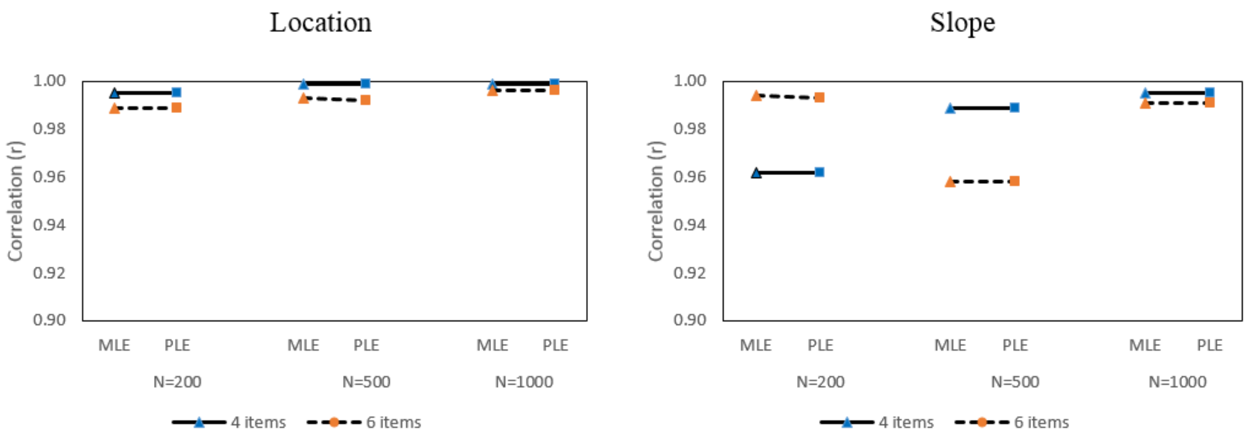
test length of 4 items has more variability than the test length of 6 items in slope parameter estimates of the models. When given sample sizes of 200 and 500, the mean RMSE for location parameters of PLE and MLE are smaller than those for slope parameters overall. However, the mean RMSE for location and slope parameters appeared to be similar in the conditions associated with sample size of 1000.

The correlation coefficients (r) between the estimated parameters and true parameters for location and slope parameters are also presented in Tables 17 and 18. Both PLE and MLE yield high correlations across all conditions, ranging from 0.948 to 0.999 for location parameters and 0.958 to 0.999 for slope parameters. In Figure 18, the location parameters of two-dimensional binary response models showed higher correlations than the slope parameters while the slope parameters of polytomous response models did higher correlations than the location parameters.

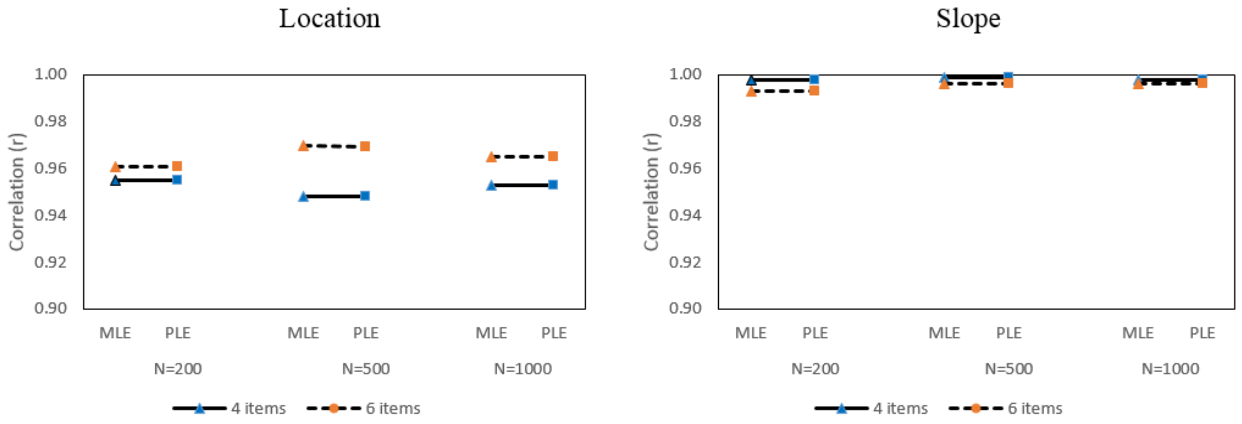
Three-dimensional models. Table 19 presents mean bias, RMSE and their standard deviations for location (λ_{ij}) and slope parameters (ν_{ij}) and the correlations between the estimated and true parameters of three-dimensional models with 6 items by the number of categories.

Again, PLE and MLE show almost the same values of mean bias, mean RMSE, and correlations across all three-dimensional models with 6 items. For binary response models, the mean bias values of PLE and MLE are around -0.03 for location parameters and range from -0.12 to 0.07 for slope parameters. For polytomous response models, the absolute values of the mean bias values are less than 0.04 for location parameters and 0.03 for slope parameters. These findings indicate that both PLE and MLE resulted in almost unbiased parameter estimates. For location parameters, the condition of 3 response categories and sample size of 1000 resulted in the largest mean bias values with values of -0.0400 for PLE and -0.0399 for MLE, whereas the

(a) 2 categories



(b) 3 categories



(c) 5 categories

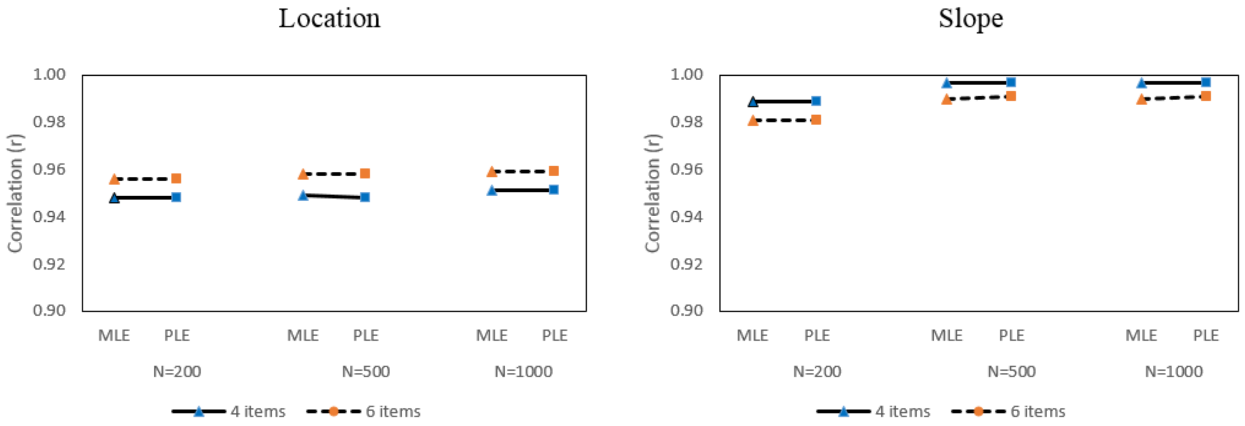


Figure 18. Correlation coefficients (*r*) for location and slope parameters of 2 dimensional models with 4 & 6 items

Table 19. Mean bias, RMSE and their (SDs) and correlation coefficients for location and slope parameters of 3 dimensional models with 6 items

	6 items_Location					6 items_Slope				
	Bias		RMSE		$r_{(\lambda,\ \hat{\lambda})}$	Bias		RMSE		$r_{(\nu,\ \hat{\nu})}$
2 categories										
					Sample size (N) =200					
MLE	-0.0394	(0.0594)	0.1516	(0.0360)	0.991	-0.0774	(0.2019)	0.4114	(0.0802)	0.921
PLE	-0.0393	(0.0594)	0.1515	(0.0360)	0.991	-0.1234	(0.2045)	0.4439	(0.1014)	0.924
					Sample size (N) = 500					
MLE	-0.0300	(0.0225)	0.0892	(0.0187)	0.999	-0.0437	(0.1013)	0.3469	(0.0664)	0.990
PLE	-0.0299	(0.0223)	0.0893	(0.0187)	0.999	-0.0469	(0.1007)	0.3443	(0.0678)	0.991
					Sample size (N) = 1000					
MLE	-0.0392	(0.0293)	0.0758	(0.0224)	0.998	0.0736	(0.0981)	0.2876	(0.0695)	0.979
PLE	-0.0392	(0.0293)	0.0759	(0.0225)	0.998	0.0732	(0.0979)	0.2853	(0.0691)	0.979
3 categories										
					Sample size (N) =200					
MLE	-0.0332	(0.3272)	0.3360	(0.1690)	0.924	0.0228	(0.1091)	0.4550	(0.0988)	0.996
PLE	-0.0324	(0.3272)	0.3388	(0.1718)	0.924	0.0218	(0.1115)	0.4647	(0.0885)	0.996
					Sample size (N) = 500					
MLE	-0.0362	(0.2979)	0.2877	(0.1389)	0.938	-0.0093	(0.0990)	0.3128	(0.0781)	0.996
PLE	-0.0363	(0.2980)	0.2874	(0.1392)	0.937	-0.0074	(0.0973)	0.3195	(0.0823)	0.997
					Sample size (N) = 1000					
MLE	-0.0399	(0.2820)	0.2619	(0.1367)	0.944	-0.0105	(0.1113)	0.2596	(0.0782)	0.996
PLE	-0.0400	(0.2820)	0.2621	(0.1368)	0.944	-0.0055	(0.1072)	0.2625	(0.0798)	0.996
5 categories										
					Sample size (N) =200					
MLE	0.0098	(0.2593)	0.3692	(0.1353)	0.960	-0.0299	(0.1586)	0.5661	(0.2148)	0.991
PLE	0.0076	(0.2601)	0.3678	(0.1346)	0.960	-0.0313	(0.1635)	0.5689	(0.2183)	0.991
					Sample size (N) = 500					
MLE	0.0099	(0.2435)	0.2730	(0.1335)	0.968	0.0022	(0.1276)	0.3961	(0.1544)	0.993
PLE	0.0096	(0.2438)	0.2728	(0.1333)	0.968	0.0020	(0.1279)	0.3972	(0.1538)	0.993
					Sample size (N) = 1000					
MLE	0.0143	(0.2534)	0.2413	(0.1484)	0.965	0.0123	(0.1020)	0.2723	(0.1056)	0.996
PLE	0.0141	(0.2539)	0.2415	(0.1485)	0.965	0.0119	(0.1041)	0.2728	(0.1059)	0.995

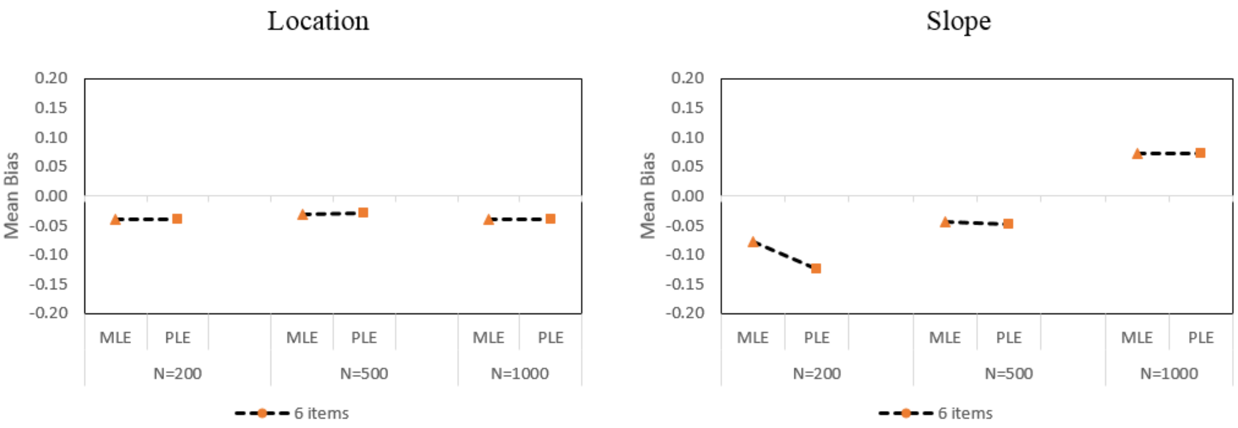
condition of 5 response categories and sample size of 200 showed the smallest ones of 0.0076 for PLE and 0.0098 for MLE. For slope parameters, the 2 response category model with 200 persons had the largest mean bias with values of −0.1234 for PLE and −0.0774 for MLE while the 5 response category models with 500 persons showed the smallest mean bias with values of 0.0020 for PLE and 0.0022 for MLE.

Figure 19 provides the mean bias for location and slope parameters of PLE and MLE for three-dimensional models with 6 items. All of 3 category response models yielded less mean bias values for slope parameter estimates than for location parameter estimates, whereas binary and 5 category response models showed less mean bias values for location parameter estimates than for slope parameter estimates, regardless of sample sizes.

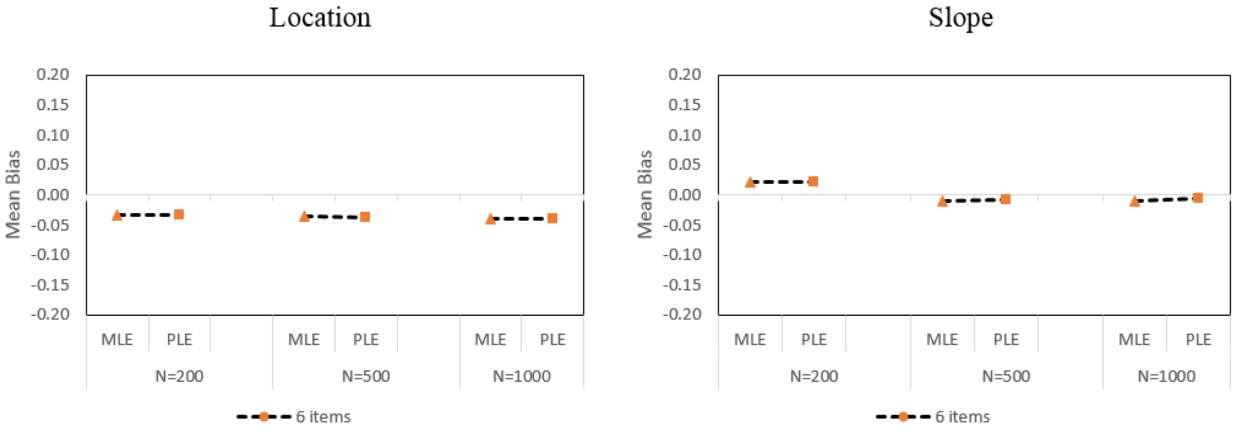
For location parameters of three-dimensional response models, the mean RMSE values ranged from 0.0758 to 0.3692. The binary response models with 1000 persons had the smallest mean RMSE values of 0.0759 for PLE and 0.0758 for MLE. The largest mean RMSE values are 0.3678 for PLE and 0.3692 for MLE resulted from the condition of 5 response categories and 200 persons. Overall, the mean RMSE values for slope parameters are larger than those for location parameters, ranging from 0.2596 to 0.5689 for two estimation methods. Again, the 5 response categories and sample size of 200 showed the largest mean RMSE for slope parameters with values of 0.5689 for PLE and 0.5661 for MLE. The smallest mean RMSE values for slope parameters are found in 3 response categories with sample size of 1000, which are 0.2625 for PLE and 0.2596 for MLE.

Figures 20 shows the mean RMSE for location and slope parameters of PLE and MLE for three-dimensional models with 6 items. Across all conditions, larger samples showed smaller mean RMSE values for both location and slope parameters. As mentioned earlier, the mean

(a) 2 categories



(b) 3 categories



(c) 5 categories

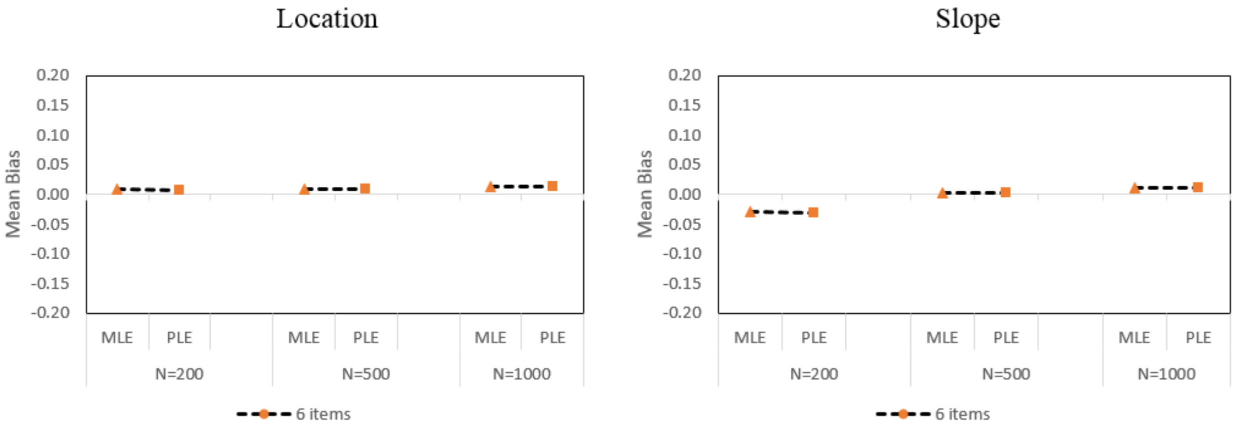
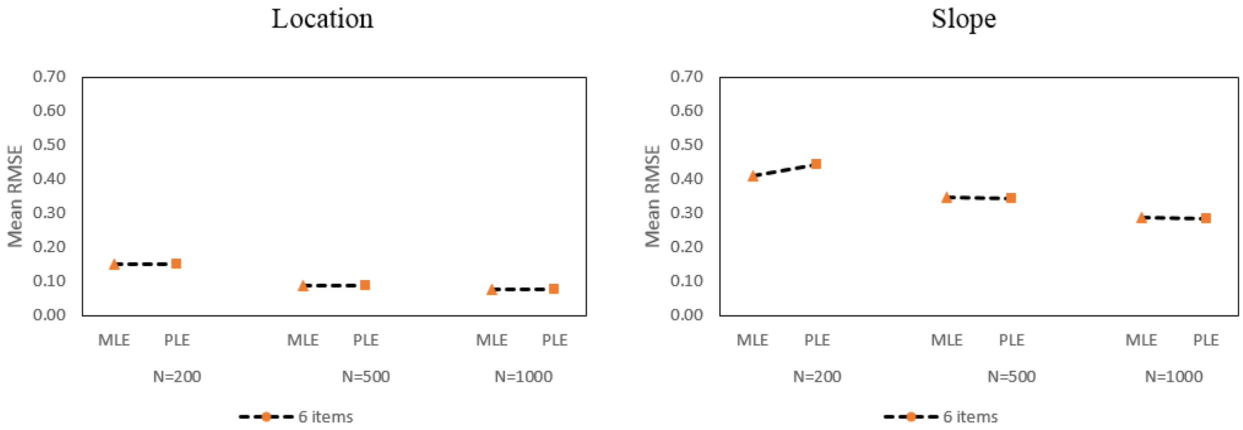
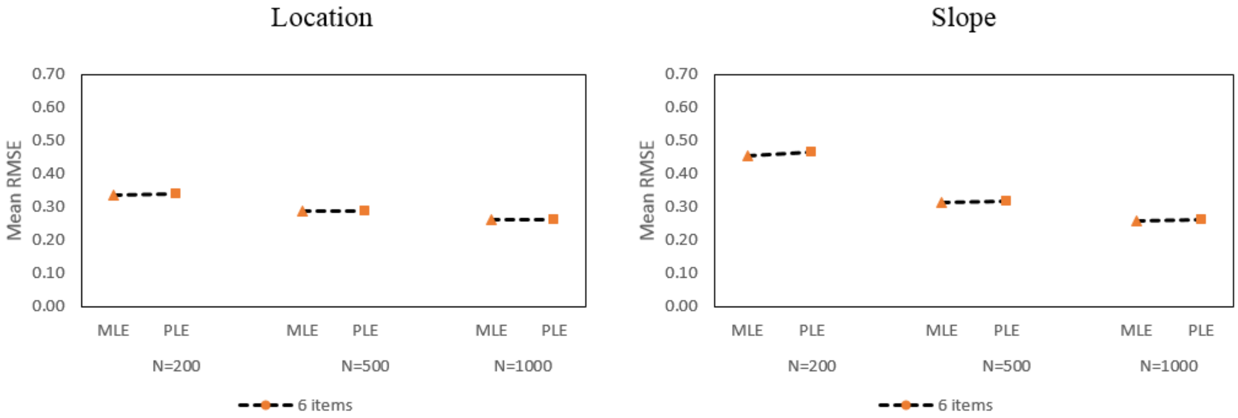


Figure 19. Mean bias for location and slope parameters of 3 dimensional models with 6 items

(a) 2 categories



(b) 3 categories



(c) 5 categories

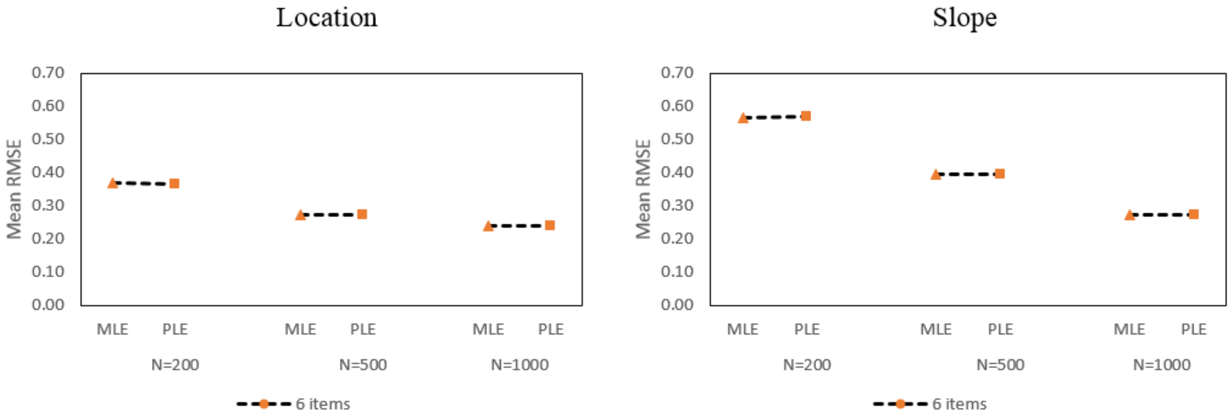


Figure 20. Mean RMSE for location and slope parameters of 3 dimensional models with 6 items

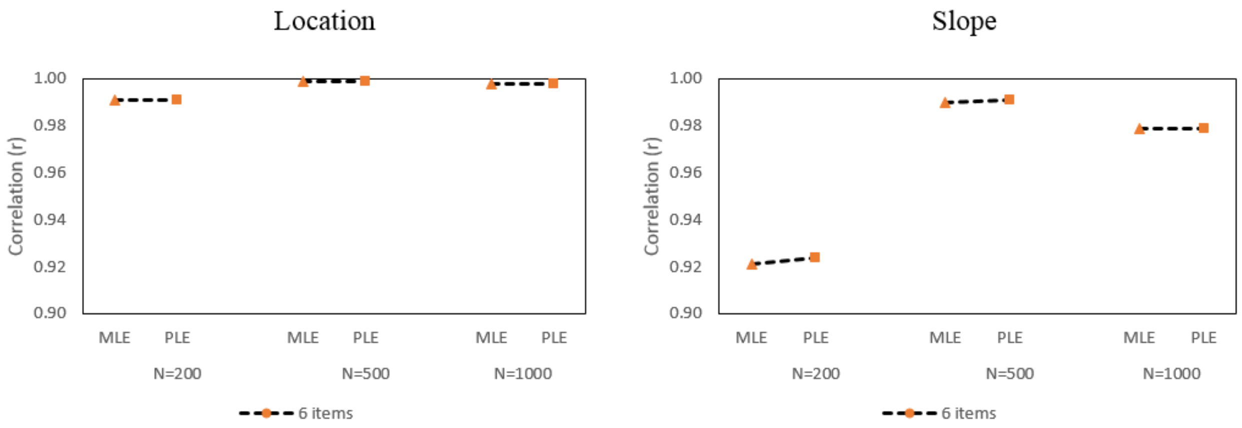
RMSE values for slope parameters appear to be larger than those for location parameters, but the polytomous models associated with sample size of 1000 showed similar mean RMSE values between location and slope parameters.

The correlation coefficients (r) between the estimated parameters and true parameters for location and slope parameters are also shown in Table 16 and illustrated in Figure 21. All of the estimates from PLE and MLE are highly related to true parameters. For three-dimensional binary response models, PLE and MLE yielded high correlations between the estimates and true parameters for both location and slope parameters. The correlation for location parameters are as high as $\mathrm{r} = 0.99$ and those for slope parameters range from $\mathrm{r} = 0.92$ to $\mathrm{r} = 0.99$.

For polytomous response category models, slope parameters had slightly higher correlations than location parameters. The correlations between the estimates and true parameters for slope parameters are as high as 0.996 under all 3 response category models while the correlations for location parameters range from 0.924 to 0.944. For 5 response category models, the correlations between parameter estimates and true parameters are above 0.96 for location parameters and above 0.99 for slope parameters in all models with different sample sizes.

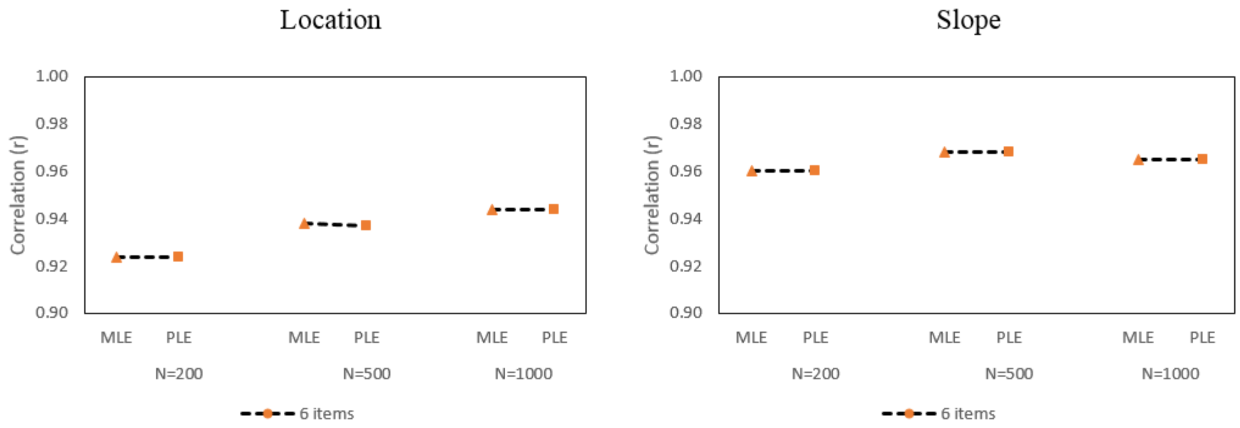
As shown in Tables 17, 18, and 19, the mean bias, mean RMSE, and correlations of PLE are almost identical to those of MLE across all multidimensional models with small numbers of items. Figures 16 through 21 also illustrate that PLE yield almost equal parameter estimates to MLE. To compare the performance of PLE for multidimensional models with MLE, the correlations of the estimates between two estimation methods were computed and they are shown in Table 20. The estimates from PLE and MLE are correlated as highly as 0.999 under all conditions, indicating that the parameter estimates between MLE and PLE are all equivalent.

(a) 2 categories


Location

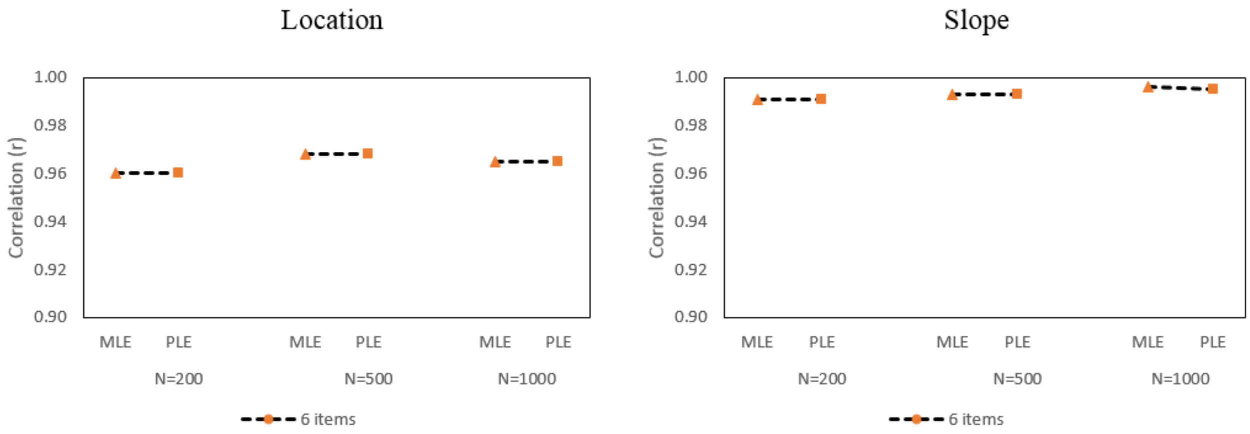
Slope

(b) 3 categories


Location

Slope

(c) 5 categories


Location

Slope

Figure 21. Correlation coefficients (*r*) for location and slope parameters of 3 dimensional models with 6 items

Table 20. Correlation coefficients (*r*) between the parameter estimates obtained from MLE and PLE for multidimensional models with 4 and 6 items

	2 categories			3 categories			5 categories		
	Sample size			Sample size			Sample size		
	N=200	N=500	N=1000	N=200	N=500	N=1000	N=200	N=500	N=1000
2 dimensional models									
Location parameters									
4 items	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
6 items	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Slope parameters									
4 items	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
6 items	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3 dimensional models									
Location parameters									
6 items	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Slope parameters									
6 items	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 21. Mean RMSDiff of parameter estimates between MLE and PLE for multidimensional models with 4 and 6 items.

	2 categories			3 categories			5 categories		
	Sample size			Sample size			Sample size		
	N=200	N=500	N=1000	N=200	N=500	N=1000	N=200	N=500	N=1000
2 dimensional models									
Location parameters									
4 items	0.000	0.000	0.000	0.012	0.004	0.003	0.028	0.008	0.003
6 items	0.001	0.000	0.000	0.042	0.037	0.010	0.014	0.006	0.004
Slope parameters									
4 items	0.003	0.001	0.000	0.038	0.018	0.013	0.058	0.027	0.010
6 items	0.011	0.006	0.001	0.035	0.024	0.009	0.021	0.011	0.006
3 dimensional models									
Location parameters									
6 items	0.001	0.001	0.000	0.029	0.010	0.008	0.024	0.006	0.003
Slope parameters									
6 items	0.019	0.006	0.003	0.057	0.019	0.018	0.067	0.014	0.009

In addition to correlations, the mean RMSDiff values of the parameter estimates between MLE and PLE of LMA models were also computed for multidimensional models with 4 and 6 items to measure how close the parameter estimates from PLE are to those from MLE. They are shown in Table 21. The mean RMSDiff of the parameter estimates between MLE and PLE range from 0.000 to 0.042 for location parameters and 0.000 to 0.067 for slope parameters. The findings indicate that the parameter estimates obtained from MLE and PLE are not only linearly related but also close to each other. Overall, the largest sample size resulted in the smallest mean RMSDiff value, which is the same tendency found in unidimensional models.

In summary, the results of Simulation study 3 shows that PLE behaves similarly to MLE when it estimates parameters of multidimensional LMA models with small numbers of items using both Steps 1 and 2 of the algorithm. The correlations of the estimates between PLE and MLE are as high as $\mathrm{r} = 1.000$ for all conditions in Simulation study 3 and the mean RMSDiff of the parameter estimates between PLE and MLE values are also small.

Multidimensional Models with Large Numbers of Items

This section describes how well PLE overcomes the limitation of MLE using the full algorithm for multidimensional models when the number of items is large. The item parameters of 54 multidimensional models with 20 and 50 items were estimated by PLE, varying the number of latent variables from 2 to 4. Table 22 shows the bias values of parameters for two- and three-dimensional models with 20 items, 2 and 3 categories and sample size of 1000. Including the values shown in the table, the computed bias and RMSE values for each item in all multidimensional models with large numbers of items were averaged over items and categories for interpretation and the averaged values are reported in Tables 23 and 24.

Table 22. Bias of parameters for two- and three-dimensional models with 20 items, 2 and 3 categories and sample size of 1000

Item	Location		Slope		Item	Location		Slope	
	2D	3D	2D	3D		2D	3D	2D	3D
2 categories (N=1000)									
Item 1	0.0537	0.0199	0.0259	-0.0176	Item 11	-0.2631	-0.1490	0.0061	0.0162
Item 2	0.1084	0.0814	0.0046	-0.0107	Item 12	-0.2152	-0.1476	-0.0003	0.0182
Item 3	0.0580	0.0407	0.0233	0.0129	Item 13	-0.3844	-0.1851	-0.0168	0.0027
Item 4	0.1369	0.0803	0.0037	-0.0243	Item 14	-0.5914	-0.3154	-0.0760	0.0073
Item 5	0.1575	0.0928	-0.0202	-0.0517	Item 15	-0.1990	-0.1093	-0.0116	0.0848
Item 6	0.1139	0.0782	0.0361	0.0229	Item 16	-0.4103	-0.2331	-0.0062	0.0605
Item 7	0.1619	0.0725	0.0070	-0.0353	Item 17	-0.4819	-0.2554	-0.0421	0.0190
Item 8	0.1748	-0.3557	-0.0379	-0.0436	Item 18	-0.2229	-0.1638	0.0008	0.0736
Item 9	0.1239	-0.2116	0.0221	0.0050	Item 19	-0.3443	-0.2049	-0.0138	0.0078
Item 10	0.0650	-0.1542	0.0113	0.0161	Item 20	-0.4469	-0.2844	0.0197	0.0557
3 categories (N=1000)									
Item 1_2	-0.1592	-0.2271	0.1573	0.1914	Item 11_2	0.0816	0.0664	0.0015	0.0163
Item 1_3	-0.0354	-0.1531	0.0967	0.1216	Item 11_3	-0.0901	-0.0670	0.0284	-0.0489
Item 2_2	0.0063	-0.0764	-0.1092	-0.0454	Item 12_2	0.0801	0.0375	0.1392	0.1021
Item 2_3	-0.0012	0.2218	0.1025	-0.0684	Item 12_3	0.3700	0.0269	0.1410	0.1004
Item 3_2	0.0735	0.2009	0.0386	0.0818	Item 13_2	0.1546	0.0632	-0.0781	0.0505
Item 3_3	0.0185	-0.1650	-0.0879	-0.1931	Item 13_3	0.0556	-0.0638	-0.0632	0.0377
Item 4_2	-0.0467	-0.0996	0.1532	0.0990	Item 14_2	-0.0986	-0.0379	-0.0170	-0.0243
Item 4_3	0.0285	0.0329	-0.0970	-0.0367	Item 14_3	0.0854	0.1045	0.0114	0.0414
Item 5_2	-0.0076	0.1377	0.2880	0.1892	Item 15_2	-0.1002	-0.0199	0.0081	0.0031
Item 5_3	0.0332	-0.0366	-0.1329	-0.1103	Item 15_3	0.1453	0.0549	-0.0127	-0.0206
Item 6_2	-0.0839	-0.1406	-0.1766	-0.1500	Item 16_2	-0.0456	0.0030	0.0105	0.0423
Item 6_3	0.0053	-0.0331	0.1013	0.0907	Item 16_3	0.0745	0.1109	0.0101	0.0753
Item 7_2	0.0238	-0.0452	-0.0608	-0.0707	Item 17_2	-0.2159	-0.1423	0.0337	0.0469
Item 7_3	-0.0717	-0.0167	0.0425	0.0441	Item 17_3	-0.1920	-0.1884	0.0430	0.0254
Item 8_2	-0.0545	-0.1415	-0.0983	-0.0454	Item 18_2	-0.1821	-0.1614	0.0065	0.0004
Item 8_3	0.0253	0.1634	0.1464	0.1044	Item 18_3	-0.2063	-0.0625	0.0319	-0.0334
Item 9_2	-0.0506	-0.1207	0.1179	0.1838	Item 19_2	-0.0708	-0.0604	-0.0129	0.0084
Item 9_3	-0.0997	-0.2310	0.1385	0.1247	Item 19_3	-0.0283	-0.0030	0.0067	-0.0095
Item 10_2	0.0838	0.0487	-0.0715	-0.0254	Item 20_2	-0.2510	-0.1719	0.0312	-0.0271
Item 10_3	-0.0189	-0.0132	0.0313	0.0021	Item 20_3	-0.1703	-0.1887	0.0743	0.0809

Parameter Recovery of PLE with 20 and 50 items

Location parameters. Table 23 shows mean bias, RMSE and their standard deviations, and correlations between the estimated and true parameters for location parameters (λ_{ij}) of multidimensional models with 20 and 50 items.

The mean bias values of PLE for location parameters are small across all polytomous response models with the minimum value of -0.0381 and the maximum value of 0.0179, whereas those are large in binary response models relative to polytomous response models, which range from -0.1388 to -0.0986. Overall, location parameters of binary and 3 response category models were underestimated and the estimates of 5 response category models yielded less mean bias values than those of binary and 3 response category models. Figure 22 illustrates these findings more clearly. The number of latent variables didn't show any pattern across all conditions in estimating location parameters of LMA models.

The mean RMSE of PLE for location parameters ranges from 0.1459 to 0.3344. The largest mean RMSE of PLE occurs in the 3 dimensional 3 response category model with 50 items and 200 persons (i.e., RMSE = 0.3344) while the smallest does in the 2 dimensional binary response model with 50 items and 1000 persons (i.e., RMSE = 0.1327). The larger sample sizes generally resulted in smaller mean RMSE of PLE for location parameters.

Figure 23 shows the mean RMSE of PLE for location parameters of multidimensional models with large numbers of items. The dimensions of models didn't show any particular pattern in estimating location parameters of LMA models. Given the sample size, the mean RMSE values of PLE for binary and 5 response category models with 50 items tend to increase over time with an increase of dimensions. Regardless of the dimensions, however, irregular

Table 23. Mean bias, RMSE and their (SDs) and correlation coefficients of PLE for location parameters of multidimensional models with 20 and 50 items

(a) 2 categories

	20 items					50 items				
	Bias		RMSE		$r_{(\lambda,\ \hat{\lambda})}$	Bias		RMSE		$r_{(\lambda,\ \hat{\lambda})}$
					Sample size (N) =200					
2D_PLE	-0.1206	(0.2621)	0.2991	(0.1302)	0.957	-0.1202	(0.0598)	0.2174	(0.0541)	0.998
3D_PLE	-0.0986	(0.1083)	0.1977	(0.0815)	0.992	-0.1249	(0.0998)	0.2226	(0.0723)	0.995
4D_PLE	-0.1255	(0.2021)	0.2612	(0.1208)	0.978	-0.1179	(0.1752)	0.2476	(0.1120)	0.983
					Sample size (N) = 500					
2D_PLE	-0.1195	(0.2593)	0.2667	(0.1395)	0.957	-0.1217	(0.0576)	0.1670	(0.0505)	0.998
3D_PLE	-0.1154	(0.1455)	0.1919	(0.0797)	0.986	-0.1238	(0.0891)	0.1737	(0.0706)	0.996
4D_PLE	-0.1321	(0.2097)	0.2362	(0.1239)	0.975	-0.1246	(0.1641)	0.1965	(0.1248)	0.985
					Sample size (N) = 1000					
2D_PLE	-0.1203	(0.2598)	0.2499	(0.1491)	0.957	-0.1205	(0.0534)	0.1459	(0.0492)	0.998
3D_PLE	-0.1152	(0.1493)	0.1785	(0.0856)	0.985	-0.1262	(0.0925)	0.1541	(0.0817)	0.995
4D_PLE	-0.1388	(0.2174)	0.2258	(0.1397)	0.972	-0.1273	(0.1663)	0.1804	(0.1301)	0.985

(b) 3 categories

	20 items					50 items				
	Bias		RMSE		$r_{(\lambda,\ \hat{\lambda})}$	Bias		RMSE		$r_{(\lambda,\ \hat{\lambda})}$
					Sample size (N) =200					
2D_PLE	-0.0211	(0.1495)	0.2896	(0.1031)	0.992	0.0080	(0.2347)	0.3081	(0.1497)	0.972
3D_PLE	-0.0228	(0.1127)	0.2733	(0.0816)	0.995	-0.0178	(0.2648)	0.3344	(0.1620)	0.964
4D_PLE	0.0037	(0.2596)	0.3282	(0.1311)	0.973	-0.0181	(0.2364)	0.3148	(0.1589)	0.971
					Sample size (N) = 500					
2D_PLE	-0.0235	(0.1215)	0.1946	(0.0593)	0.992	0.0008	(0.2299)	0.2421	(0.1407)	0.973
3D_PLE	-0.0381	(0.1121)	0.1906	(0.0560)	0.995	-0.0259	(0.2668)	0.2655	(0.1660)	0.964
4D_PLE	-0.0070	(0.2477)	0.2525	(0.1385)	0.976	-0.0239	(0.2320)	0.2322	(0.1557)	0.972
					Sample size (N) = 1000					
2D_PLE	-0.0234	(0.1186)	0.1530	(0.0618)	0.992	-0.0045	(0.2253)	0.2068	(0.1411)	0.974
3D_PLE	-0.0349	(0.1157)	0.1561	(0.0531)	0.995	-0.0244	(0.2662)	0.2304	(0.1733)	0.964
4D_PLE	-0.0040	(0.2578)	0.2372	(0.1447)	0.974	-0.0179	(0.2292)	0.1945	(0.1598)	0.973

Table 23. (cont.)

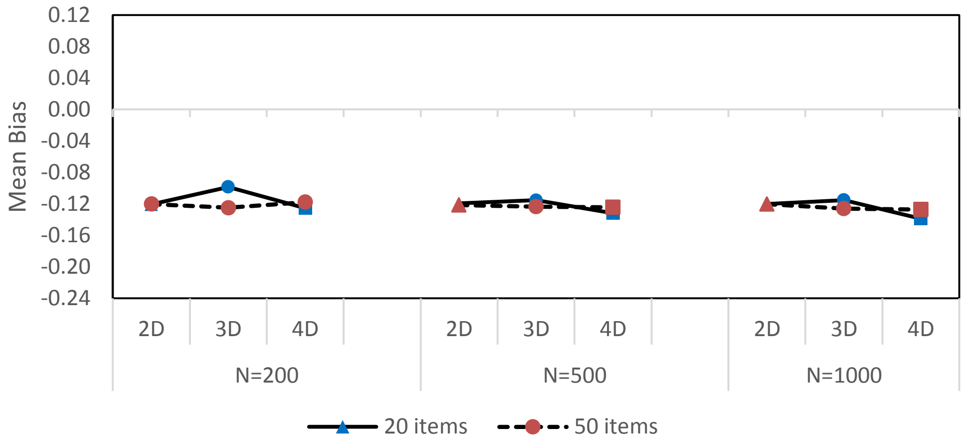
(c) 5 categories

	20 items					50 items				
	Bias		RMSE		$r_{(\lambda,\ \hat{\lambda})}$	Bias		RMSE		$r_{(\lambda,\ \hat{\lambda})}$
					Sample size (N) =200					
2D_PLE	0.0081	(0.1891)	0.3337	(0.1179)	0.975	0.0160	(0.1440)	0.2849	(0.1105)	0.983
3D_PLE	0.0130	(0.1889)	0.3275	(0.1190)	0.977	0.0179	(0.1467)	0.2897	(0.1059)	0.987
4D_PLE	0.0143	(0.1923)	0.3199	(0.1205)	0.976	0.0102	(0.1682)	0.3035	(0.1086)	0.980
					Sample size (N) = 500					
2D_PLE	0.0009	(0.1821)	0.2284	(0.0960)	0.976	0.0170	(0.1243)	0.1919	(0.0762)	0.985
3D_PLE	0.0116	(0.1808)	0.2277	(0.1093)	0.978	0.0142	(0.1288)	0.1921	(0.0767)	0.988
4D_PLE	0.0066	(0.1834)	0.2272	(0.1115)	0.978	0.0093	(0.1526)	0.2107	(0.0868)	0.982
					Sample size (N) = 1000					
2D_PLE	-0.0010	(0.1810)	0.1964	(0.0952)	0.974	0.0127	(0.1166)	0.1507	(0.0670)	0.989
3D_PLE	0.0068	(0.1821)	0.1873	(0.1106)	0.978	0.0112	(0.1168)	0.1509	(0.0681)	0.990
4D_PLE	0.0035	(0.1843)	0.1865	(0.1154)	0.978	0.0054	(0.1506)	0.1697	(0.0877)	0.982

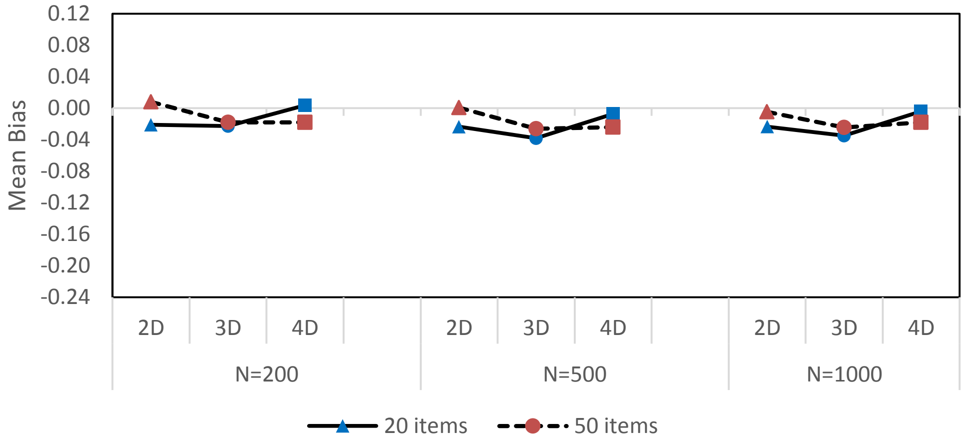
patterns of the mean RMSE were found in most of other models. It suggests that increasing dimensionality may not affect estimation accuracy of PLE.

The correlation coefficients (r) between the estimated and true parameters for location parameters are also presented in Table 23. All of the location parameter estimates obtained from PLE were highly correlated with their corresponding true parameters, ranging from 0.957 to 0.998. Of all multidimensional models with large numbers of items, the two-dimensional binary response models with different sample sizes resulted in the lowest and the highest coefficients for location parameters, which were 0.957 and 0.998, respectively. For 3 response category models with 20 items, most of the coefficients were as high as 0.99 for 2- and 3 dimensional models and around 0.97 for 4 dimensional models. For 5 response category models, the correlation coefficients were 0.974 or above.

(a) 2 categories



(b) 3 categories



(c) 5 categories

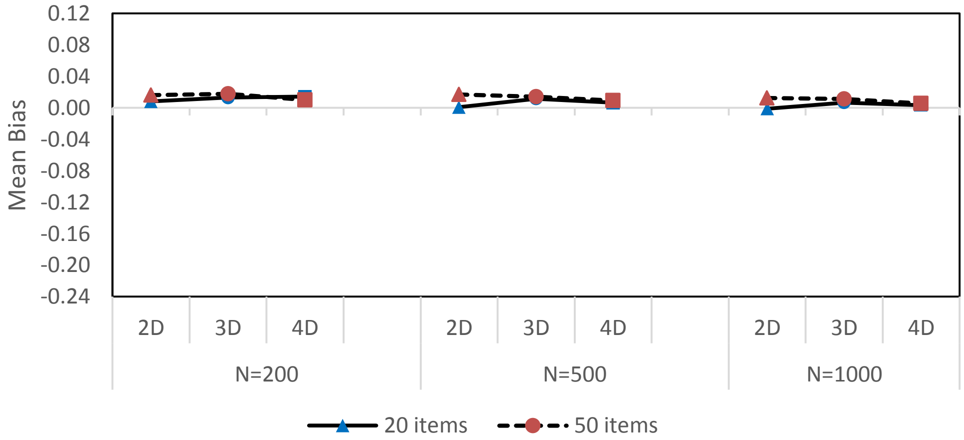
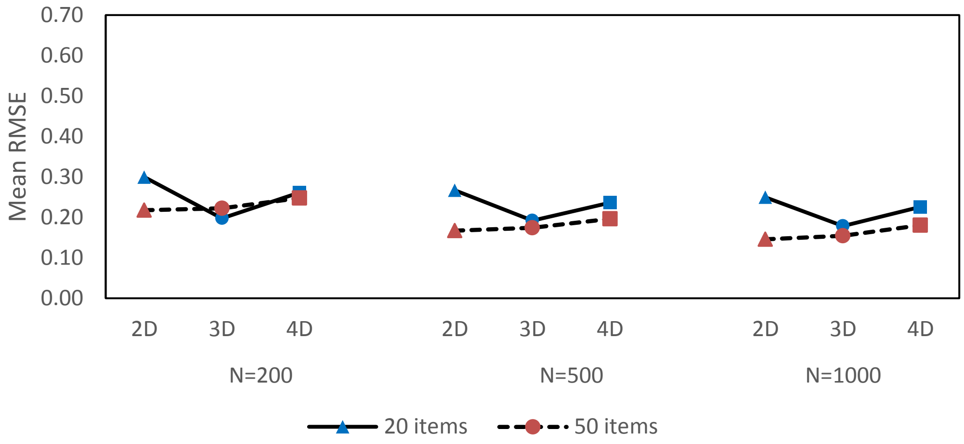
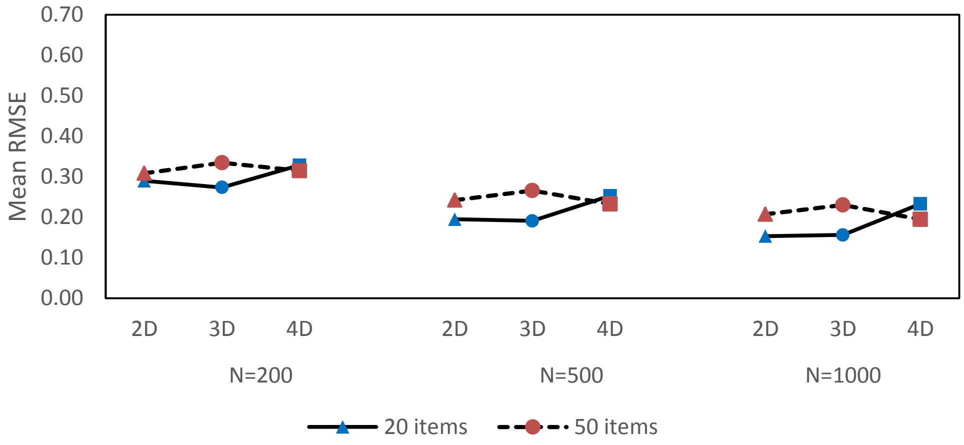


Figure 22. Mean bias of PLE for location parameters of multidimensional models with 20 and 50 items

(a) 2 categories



(b) 3 categories



(c) 5 categories

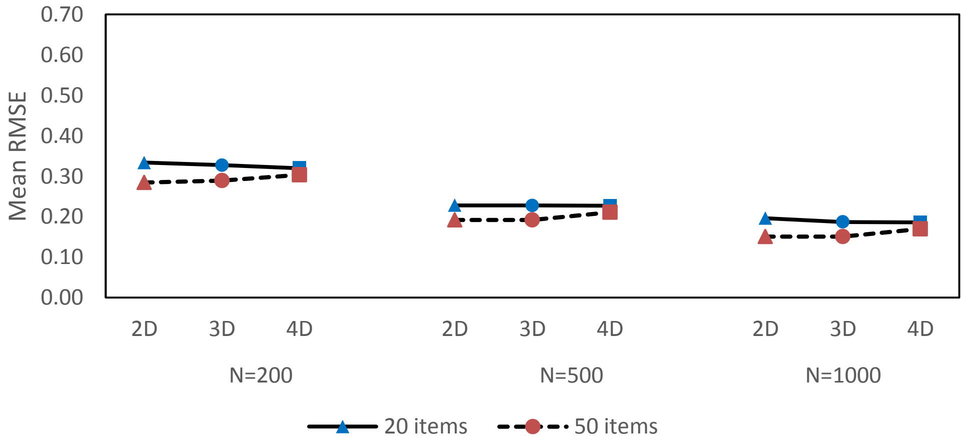


Figure 23. Mean RMSE of PLE for location parameters of multidimensional models with 20 and 50 items

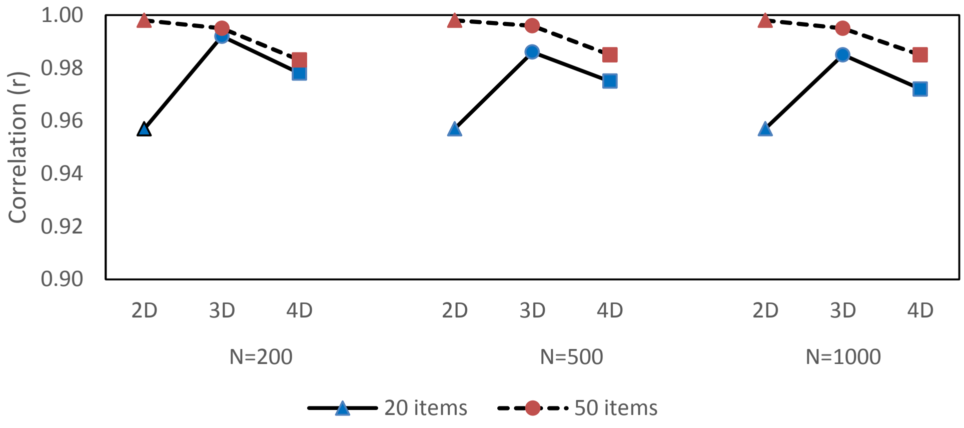
The correlations between location parameter estimates of PLE and true parameters are shown in Figure 24. Given the number of dimensions and sample size, longer test length of 50 items yielded higher correlations under all binary and 5 response category models while shorter test length of 20 items resulted in higher correlations in most of 3 response category models.

Slope parameters. Table 24 presents mean bias, mean RMSE and their standard deviations, and the correlations between the estimates ($\hat{\nu}_{ij}$) and true parameters (ν_{ij}) for slope parameters of multidimensional models with 20 and 50 items.

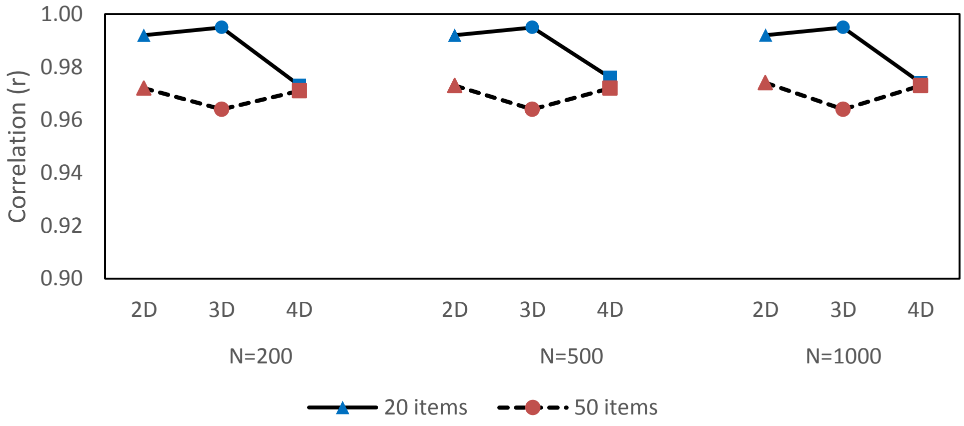
The mean bias of PLE for slope parameters of multidimensional models were relatively small across all polytomous response models, ranging from -0.0132 to 0.0355 while those values associated with binary response models were large relative to ploytomous models. This is the same pattern found in the mean bias for location parameters. Regardless of the dimensions, slope parameters of binary response models were generally underestimated, whereas those of 3 response category models were overestimated. The large mean bias values were found in all three multidimensional binary response models with 50 items and 200 persons, which are 0.1491 for 2 dimensions, -0.1390 for 3 dimensions, and -0.1251 for 4 dimension. The absolute values of the mean bias for slope parameters were less than 0.01 for most of multidimensional 5 response category models, indicating that slope parameter estimates of the models were recovered with the least amount of errors by PLE.

Figure 25 illustrates the mean bias of PLE for slope parameters of multidimensional models with large numbers of items. For polytomous response models, slope parameters under different number of dimensions and test lengths were estimated by PLE, showing a similar and small amount of mean bias values. The slope parameters of multidimensional binary response

(a) 2 categories



(b) 3 categories



(c) 5 categories

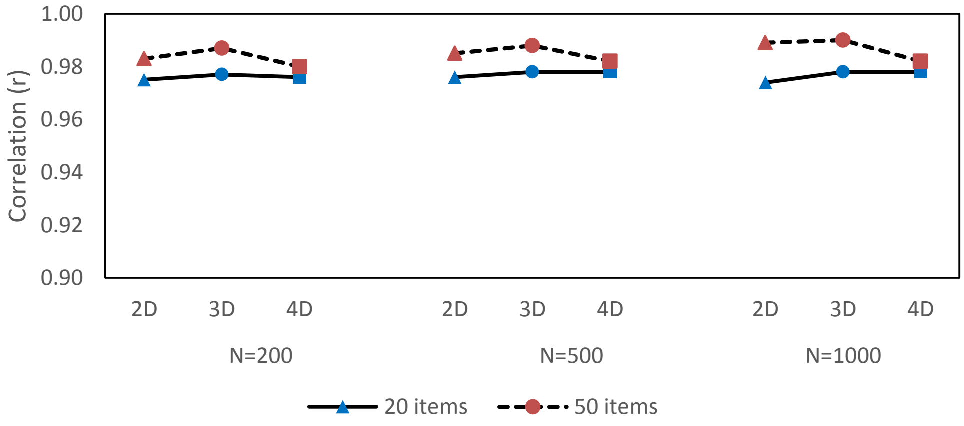


Figure 24. Correlation coefficients (*r*) of PLE for location parameters of multidimensional models with 20 and 50 items

Table 24. Mean bias, RMSE and their (SDs) and correlation coefficients of PLE for slope parameters of multidimensional models with 20 and 50 items

(a) 2 categories

	20 items					50 items				
	Bias		RMSE		$r_{(\nu,\ \hat{\nu})}$	Bias		RMSE		$r_{(\nu,\ \hat{\nu})}$
					Sample size (N) =200					
2D_PLE	-0.0786	(0.0890)	0.2899	(0.0481)	0.989	-0.1491	(0.0736)	0.2652	(0.0653)	0.989
3D_PLE	-0.1266	(0.0803)	0.3230	(0.0791)	0.987	-0.1390	(0.0924)	0.2690	(0.0642)	0.974
4D_PLE	-0.0661	(0.1194)	0.4736	(0.1496)	0.973	-0.1251	(0.0772)	0.2835	(0.0626)	0.989
					Sample size (N) = 500					
2D_PLE	-0.0560	(0.0514)	0.2057	(0.0356)	0.994	-0.0498	(0.0380)	0.1608	(0.0362)	0.997
3D_PLE	-0.0571	(0.0451)	0.2265	(0.0409)	0.996	-0.0644	(0.0484)	0.1747	(0.0376)	0.993
4D_PLE	-0.0142	(0.0676)	0.3237	(0.0872)	0.991	-0.0289	(0.0498)	0.1840	(0.0351)	0.992
					Sample size (N) = 1000					
2D_PLE	-0.0032	(0.0266)	0.1419	(0.0223)	0.998	-0.0116	(0.0265)	0.1189	(0.0202)	0.998
3D_PLE	0.0110	(0.0369)	0.1951	(0.0530)	0.997	-0.0303	(0.0285)	0.1280	(0.0245)	0.998
4D_PLE	-0.0112	(0.0323)	0.2210	(0.0937)	0.997	-0.0051	(0.0300)	0.1288	(0.0221)	0.997

(b) 3 categories

	20 items					50 items				
	Bias		RMSE		$r_{(\nu,\ \hat{\nu})}$	Bias		RMSE		$r_{(\nu,\ \hat{\nu})}$
					Sample size (N) =200					
2D_PLE	0.0355	(0.1152)	0.3168	(0.1079)	0.994	0.0246	(0.1448)	0.2976	(0.1215)	0.988
3D_PLE	0.0296	(0.1222)	0.3416	(0.1246)	0.993	0.0286	(0.1389)	0.3040	(0.1261)	0.991
4D_PLE	0.0195	(0.1191)	0.3145	(0.1179)	0.992	0.0163	(0.1155)	0.2885	(0.1192)	0.993
					Sample size (N) = 500					
2D_PLE	0.0274	(0.0885)	0.1960	(0.0675)	0.996	0.0278	(0.1249)	0.2013	(0.0922)	0.992
3D_PLE	0.0128	(0.0864)	0.2093	(0.0829)	0.996	0.0245	(0.1237)	0.2025	(0.0891)	0.992
4D_PLE	0.0127	(0.0909)	0.2041	(0.0806)	0.996	0.0136	(0.0921)	0.1865	(0.0707)	0.996
					Sample size (N) = 1000					
2D_PLE	0.0243	(0.0953)	0.1536	(0.0641)	0.996	0.0211	(0.1183)	0.1535	(0.0778)	0.992
3D_PLE	0.0239	(0.0859)	0.1610	(0.0617)	0.997	0.0283	(0.1143)	0.1551	(0.0780)	0.993
4D_PLE	0.0095	(0.0928)	0.1584	(0.0697)	0.996	0.0144	(0.0938)	0.1373	(0.0613)	0.996

Table 24. (cont.)

(c) 5 categories

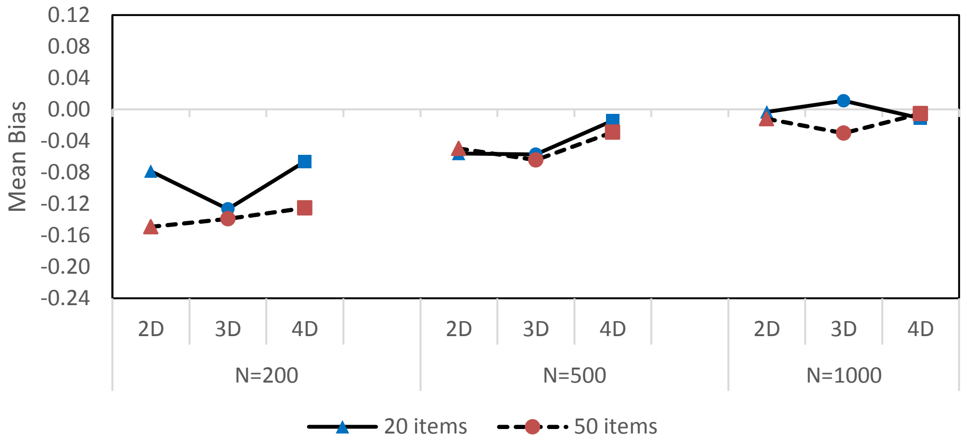
	20 items					50 items				
	Bias		RMSE		$r_{(\nu,\ \hat{\nu})}$	Bias		RMSE		$r_{(\nu,\ \hat{\nu})}$
					Sample size (N) =200					
2D_PLE	-0.0072	(0.1233)	0.3644	(0.1477)	0.992	0.0021	(0.1374)	0.3391	(0.1304)	0.989
3D_PLE	-0.0132	(0.1332)	0.3810	(0.1687)	0.992	0.0019	(0.1302)	0.3366	(0.1314)	0.994
4D_PLE	-0.0026	(0.1319)	0.3779	(0.1427)	0.991	0.0008	(0.1238)	0.3339	(0.1259)	0.994
					Sample size (N) = 500					
2D_PLE	-0.0062	(0.0949)	0.2283	(0.0871)	0.994	0.0084	(0.1059)	0.2126	(0.0834)	0.992
3D_PLE	0.0002	(0.1013)	0.2334	(0.0918)	0.994	0.0079	(0.1209)	0.2186	(0.0909)	0.993
4D_PLE	0.0020	(0.0938)	0.2410	(0.0999)	0.995	0.0024	(0.0956)	0.2184	(0.0794)	0.996
					Sample size (N) = 1000					
2D_PLE	-0.0063	(0.0908)	0.1676	(0.0652)	0.994	0.0043	(0.1037)	0.1605	(0.0696)	0.992
3D_PLE	0.0022	(0.0874)	0.1680	(0.0669)	0.995	0.0062	(0.1035)	0.1663	(0.0728)	0.995
4D_PLE	0.0002	(0.0887)	0.1792	(0.0799)	0.995	0.0024	(0.0897)	0.1584	(0.0643)	0.996

models, however, were more underestimated in the test length of 50 items than 20 items, showing fluctuation in mean bias values with an increase of dimensions.

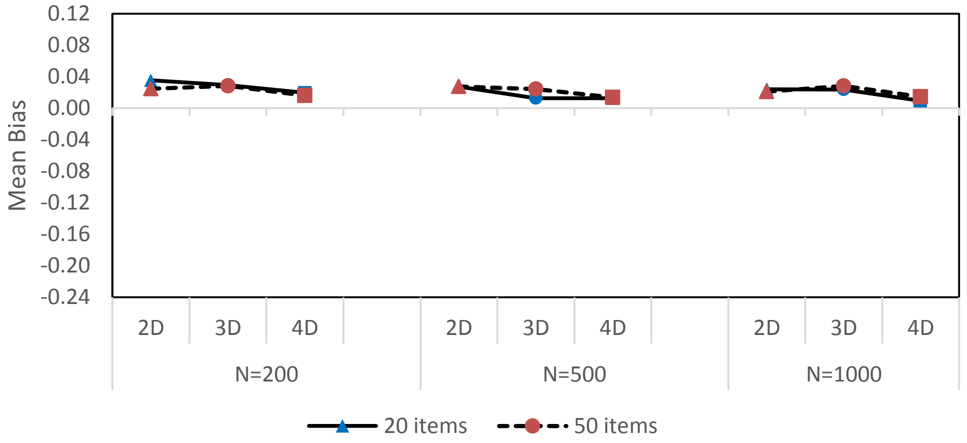
The four-dimensional binary response model with 20 items and 200 persons resulted in the largest mean RMSE of PLE for slope parameters with a value of 0.4736 while the two-dimensional binary response model associated with 50 items and 1000 persons had the smallest mean RMSE value of 0.1189. Except the models involving 20 items and 2 categories, given the test length and the number of categories, all of the models yielded similar mean RMSE values across all dimensions.

Figure 26 illustrates the mean RMSE for slope parameters by the number of category. As shown in the mean RMSE for location parameters, the larger sample sizes generally yielded smaller mean RMSE of PLE for slope parameters in all multidimensional models. For binary

(a) 2 categories



(b) 3 categories



(c) 5 categories

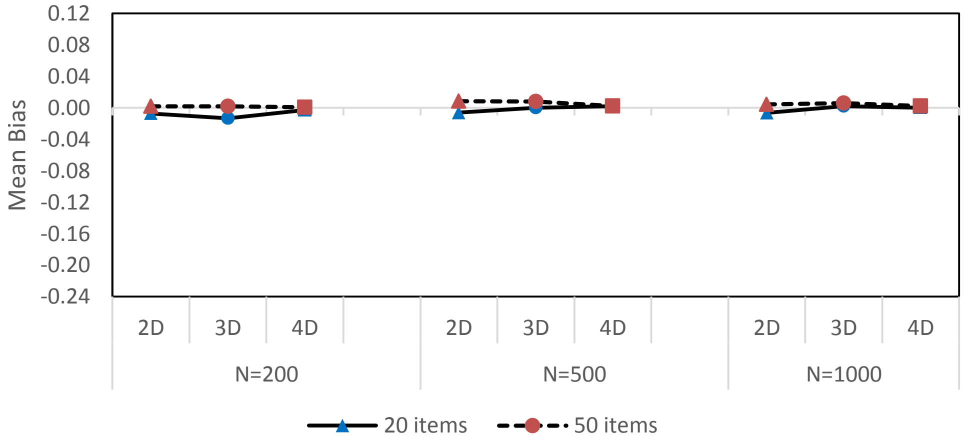
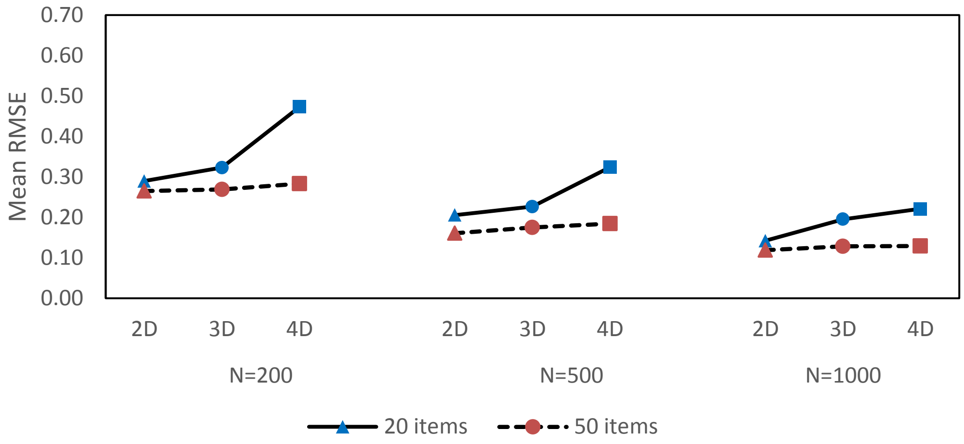
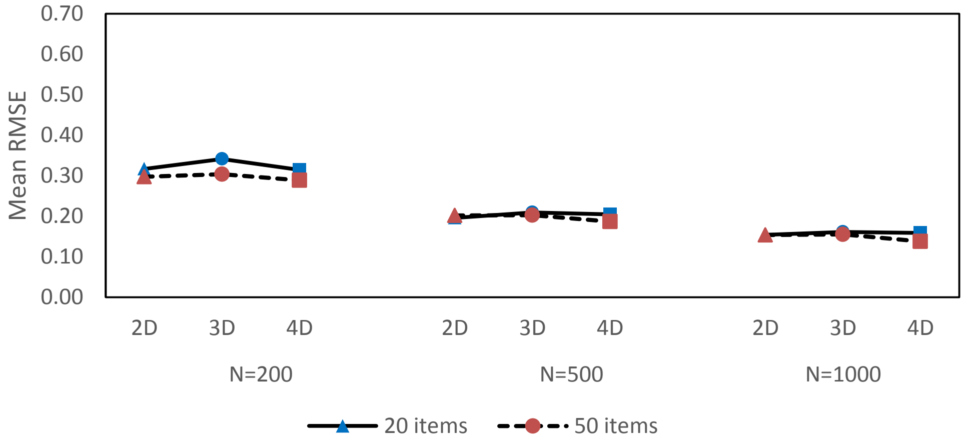


Figure 25. Mean bias of PLE for slope parameters of multidimensional models with 20 and 50 items

(a) 2 categories



(b) 3 categories



(c) 5 categories

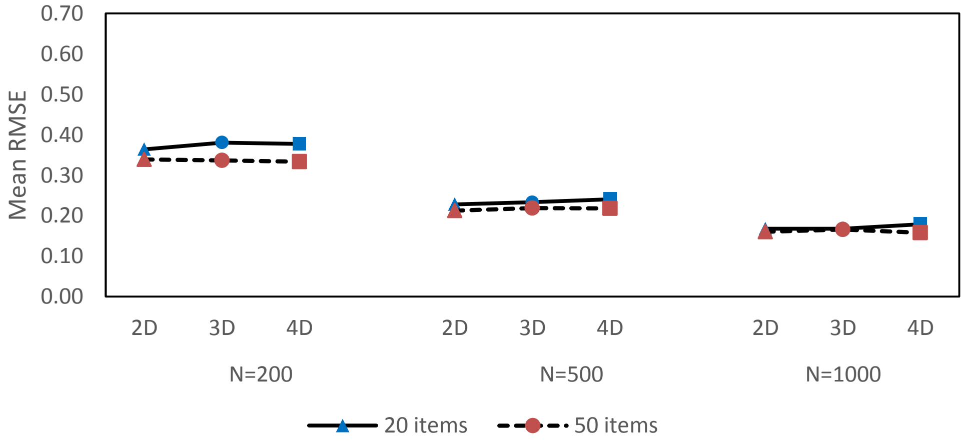


Figure 26. Mean RMSE of PLE for slope parameters of multidimensional models with 20 and 50 items

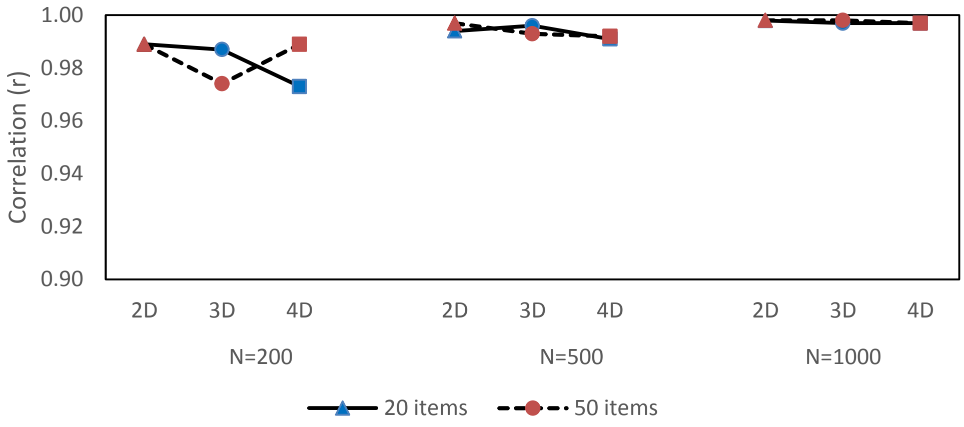
response models, the test length of 20 items were less accurate than 50 items in estimating slope parameters. For polytomous response models, both test lengths showed similar mean RSME across all dimensions, indicating that the test length didn't affect estimation accuracy of PLE for slope parameters.

The correlation coefficients (r) between the estimated and true parameters for slope parameters are also presented in Table 24. Again, PLE yielded high correlations for slope parameters across all conditions. Given sample size of 200, the correlation coefficients for slope parameters of binary response models ranged from 0.973 to 0.989, but the correlations were as high as 0.99 in all conditions associated with larger sample sizes of 500 and 1000 persons. For polytomous models, all of the correlations of PLE for slope parameters were very high, ranging from 0.988 to 0.997.

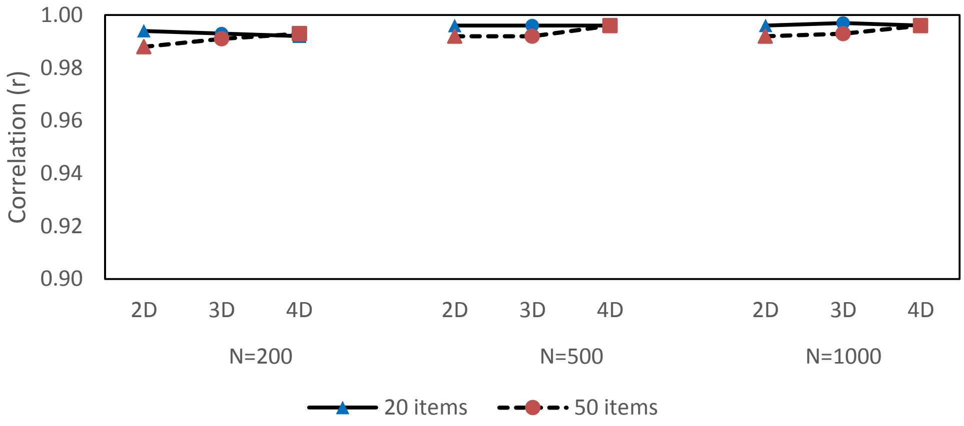
Figure 27 illustrates the correlations between slope parameter estimates of PLE and true parameters of multidimensional models with large numbers of items. Except for binary response models with sample size of 200, the correlation coefficients between different test lengths were very close or equal in all conditions, and it appears that the number of dimensions didn't affect the correlations for slope parameters of LMA models.

In summary, Simulation study 4 supports that the parameters of multidimensional LMA models with large numbers of items are successfully estimated by the full algorithm of PLE, showing again that PLE overcome the limitation of MLE of LMA models. The findings of the simulation study are similar to those found in Simulation study 2 for unidimensional models with large numbers of items; that is, PLE yields almost unbiased item parameter estimates and high correlations between the estimates and the parameters used to simulate the data in most of the conditions.

(a) 2 categories



(b) 3 categories



(c) 5 categories

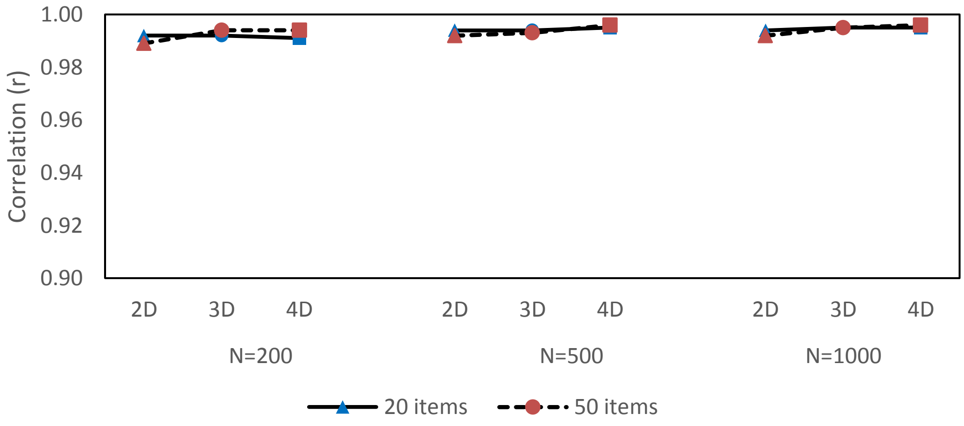


Figure 27. Correlation coefficients (*r*) of PLE for slope parameters of multidimensional models with 20 and 50 items

Estimation of Standard Errors of Pseudo-Likelihood Estimates

This section describes the results of the jackknife standard errors which were obtained to correct the standard errors given by PLE. To estimate the standard errors for the estimates from PLE, the jackknife procedure was performed for 18 unidimensional models with 4 and 6 items. The correlation coefficients (*r*) were computed to evaluate the jackknife standard errors with the standard errors given by MLE and MMLE. The results for the models involving 4 items are basically the same as those for 6 items; therefore, only the results for the 6 items are presented in this section.

Comparisons of Jackknife Standard Errors with MLE Standard Errors

In Figures 28, 29, and 30, the jackknife standard errors for the estimates from PLE are plotted against the MLE standard errors for 6 items with 2, 3, and 5 response categories. In the plots, the x-axis represents the MLE standard errors, y-axis represents the standard error estimates from jackknife for each parameter estimated by PLE, and the lines are trend lines. The PLE standard errors are also plotted against the MLE standard errors to examine the standard errors from PLE and MLE before the jackknife procedure is performed.

The jackknife standard errors for parameter estimates from PLE are strongly correlated with MLE standard errors over all conditions. For the binary response models, the correlations between the jackknife standard errors and MLE standard errors are above $\mathrm{r} = 0.99$ under all three conditions with 6 items and different sample sizes, and the standard errors from jackknife are slightly smaller than those from MLE. For the polytomous response models, the correlations between the standard errors from jackknife and MLE range from $\mathrm{r} = 0.968$ to $\mathrm{r} = 0.998$ for location parameters and $\mathrm{r} = 0.978$ to $\mathrm{r} = 0.993$ for slope parameters. For a given number of response categories, the lowest correlations between two standard errors resulted in the

(a) 2 categories

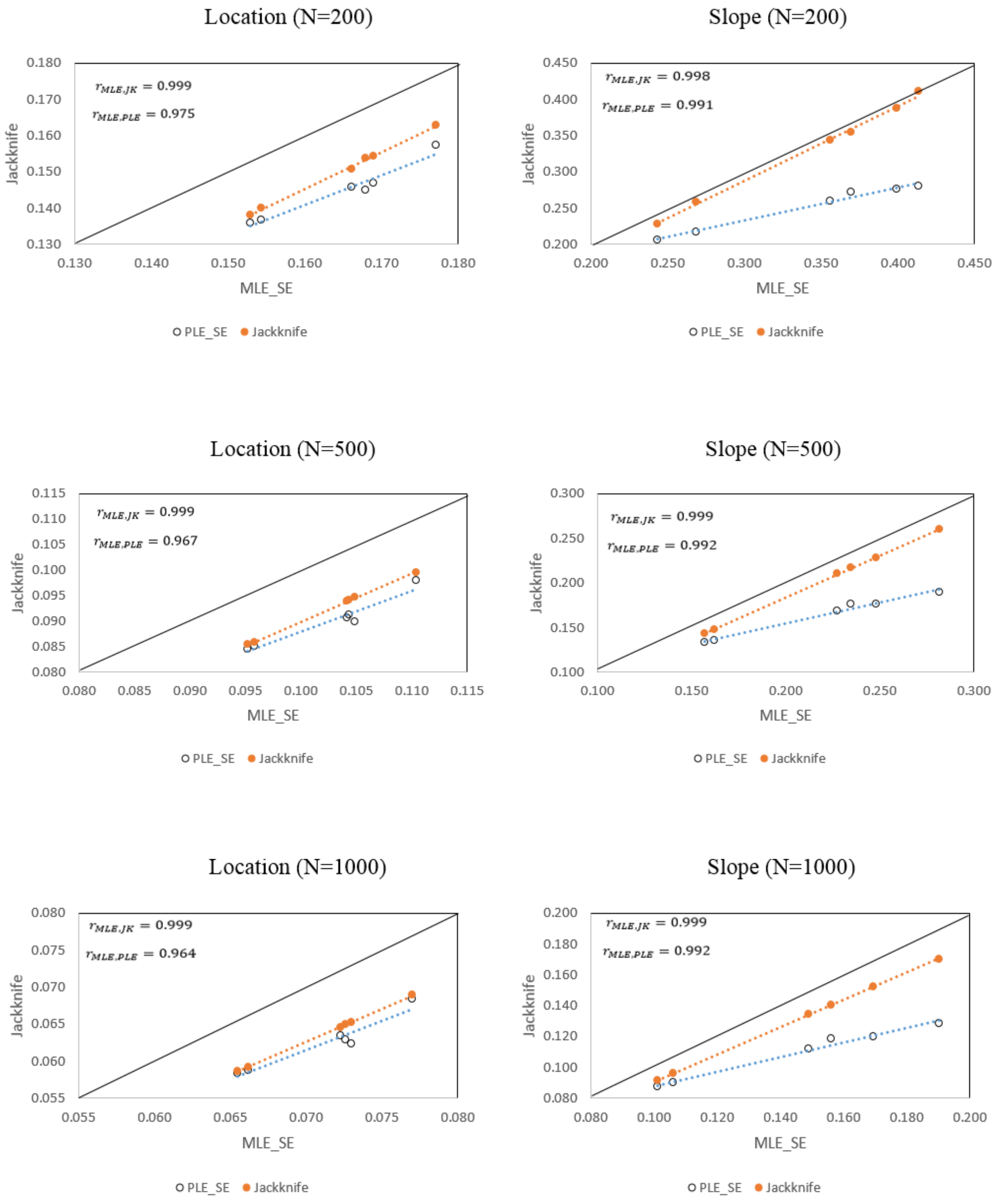


Figure 28. Jackknife standard errors from PLE vs. MLE standard errors for 6 items and 2 response categories

(b) 3 categories

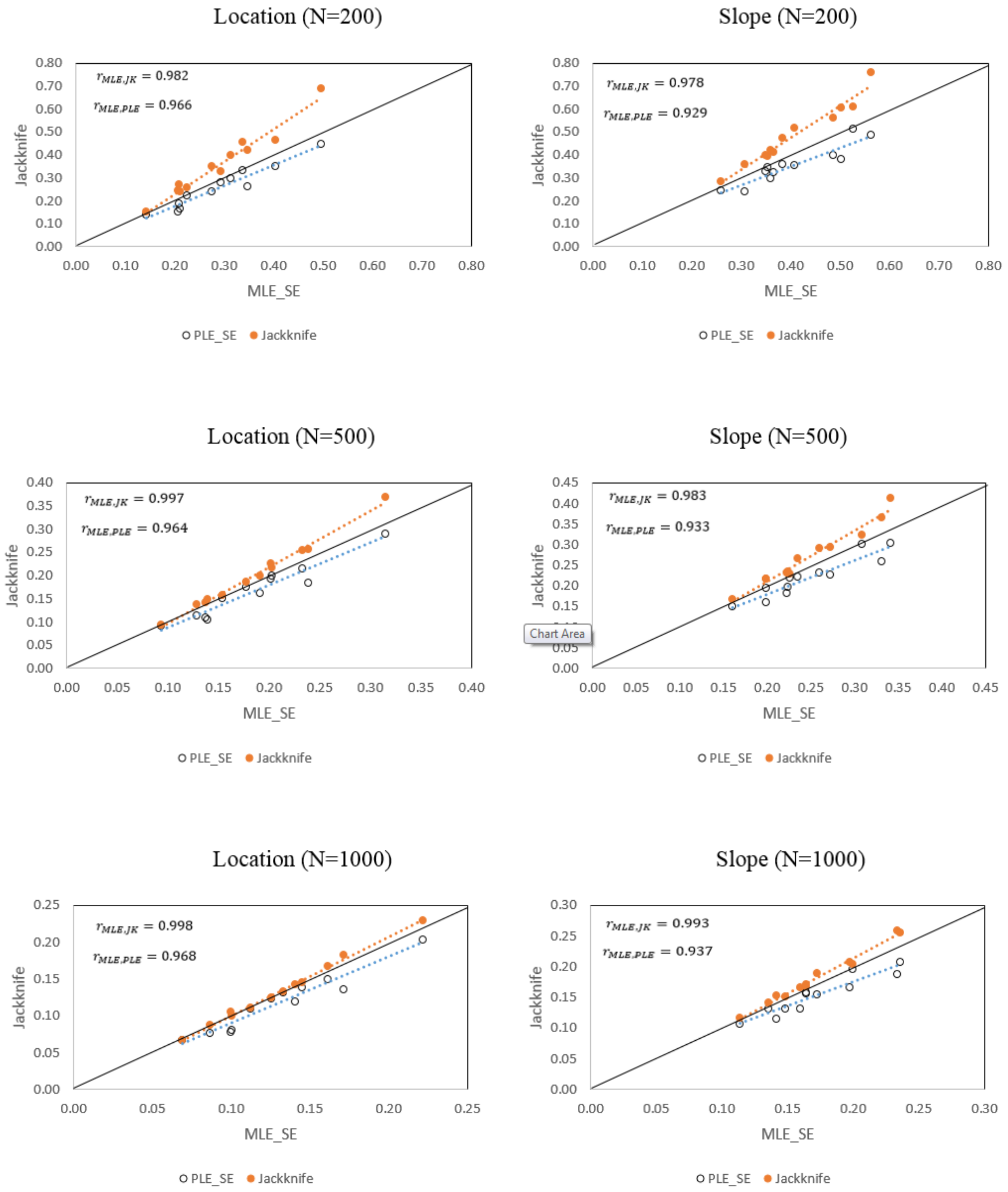


Figure 29. Jackknife standard errors from PLE vs. MLE standard errors for 6 items and 3 response categories

(c) 5 categories

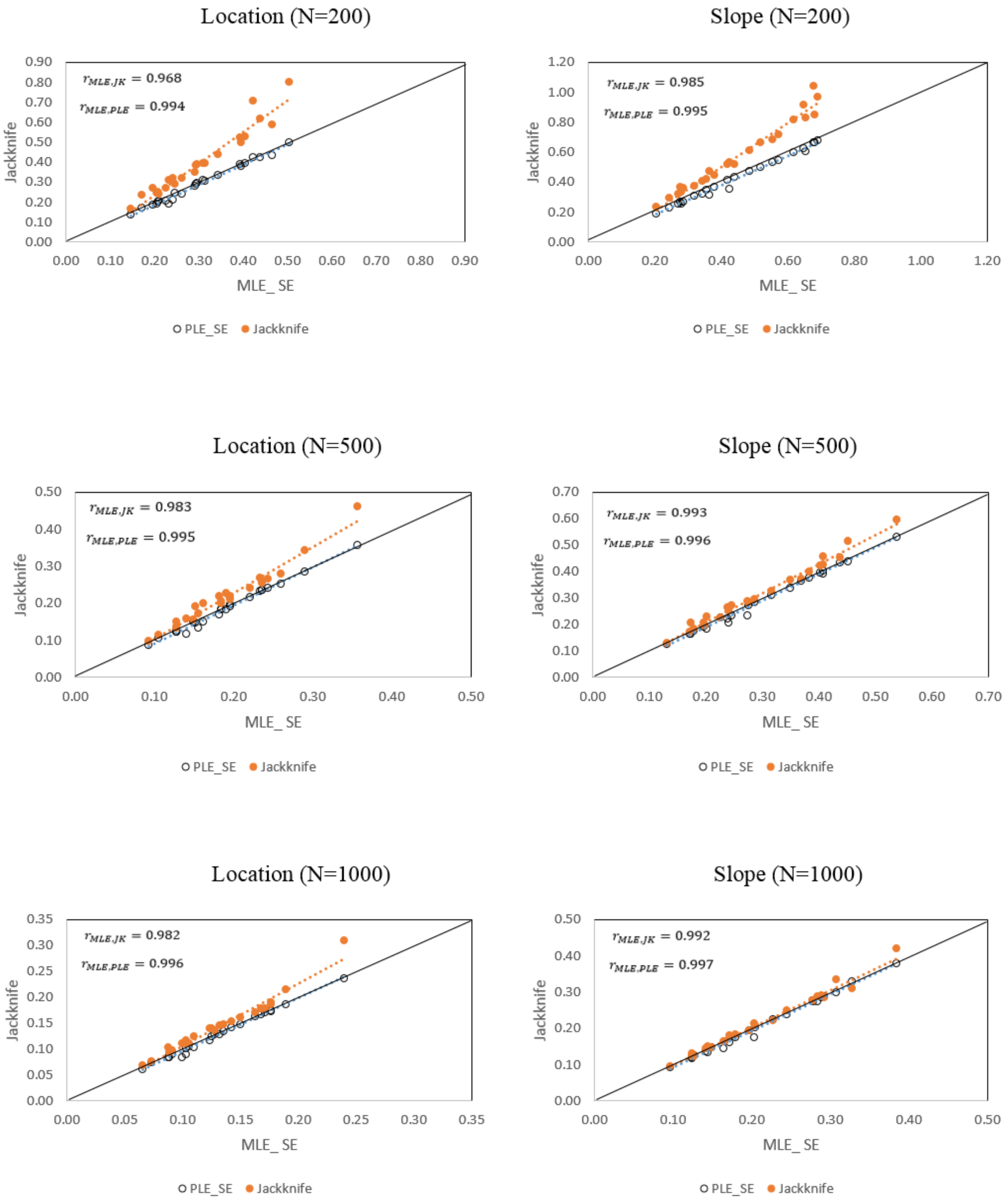


Figure 30. Jackknife standard errors from PLE vs. MLE standard errors for 6 items and 5 response categories

conditions associated with the smallest sample size of 200. Larger sample sizes are associated with smaller standard errors from both procedures and also higher correlations between the jackknife standard errors and MLE standard errors. When comparing with correlations between PLE standard errors and MLE standard errors, except for the 5 response category models, the correlations between the jackknife standard errors and MLE standard errors are higher than those between the standard errors from PLE and MLE.

In all plots, PLE standard errors are small relative to MLE standard errors, indicating that the standard errors for the parameter estimates obtained from PLE are underestimated. One noticeable result is, however, that PLE standard errors are strongly correlated with MLE standard errors even though they are underestimated. For the binary response models with 6 items and different sample sizes, the correlation coefficients between the standard errors from PLE and MLE are as high as $r = 0.97$ for location parameters and $r = 0.99$ for slope parameters. High correlations between two standard errors are also found in the polytomous response models, ranging from $r = 0.964$ to $r = 0.996$ for location parameters and $r = 0.929$ to $r = 0.997$. The correlations resulted from the 5 response category models are as very high as $r = 0.99$ for both location and slope parameters and, surprisingly, PLE standard errors are more strongly correlated with MLE standard errors than the corrected standard errors by jackknife procedure. This finding indicates that even though PLE underestimates standard errors for the estimates, they may not be problematic.

Comparisons of Jackknife Standard Errors with MMLE Standard Errors

Figures 31, 32, and 33 show the plots of the jackknife standard errors for the parameter estimates from PLE against the MMLE standard errors for 6 items with 2, 3, and 5 response categories. In the plots, the x-axis represents the MMLE standard errors, y-axis represents the

(a) 2 categories

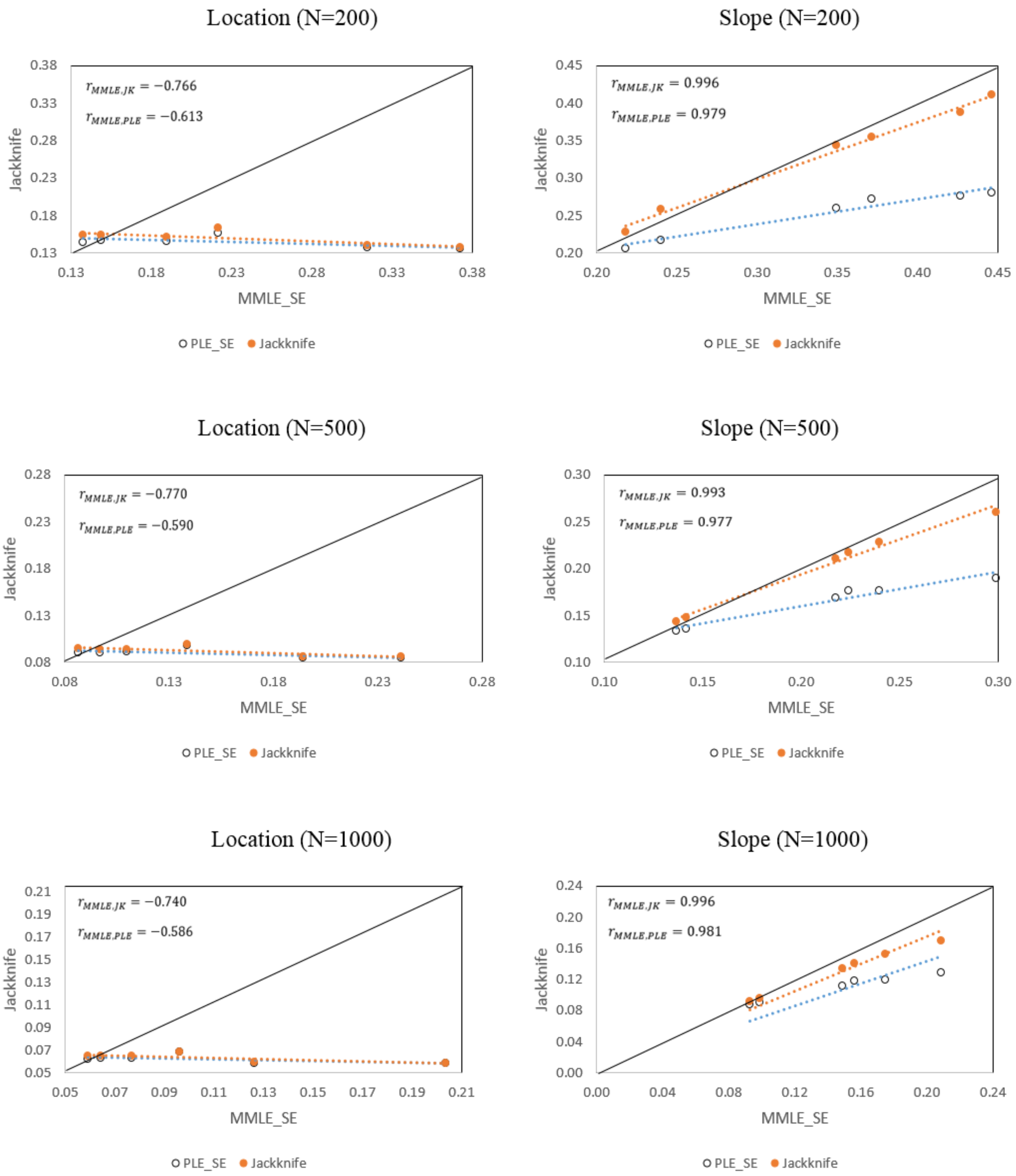


Figure 31. Jackknife standard errors from PLE vs. MMLE standard errors for 6 items and 2 response categories

(b) 3 categories

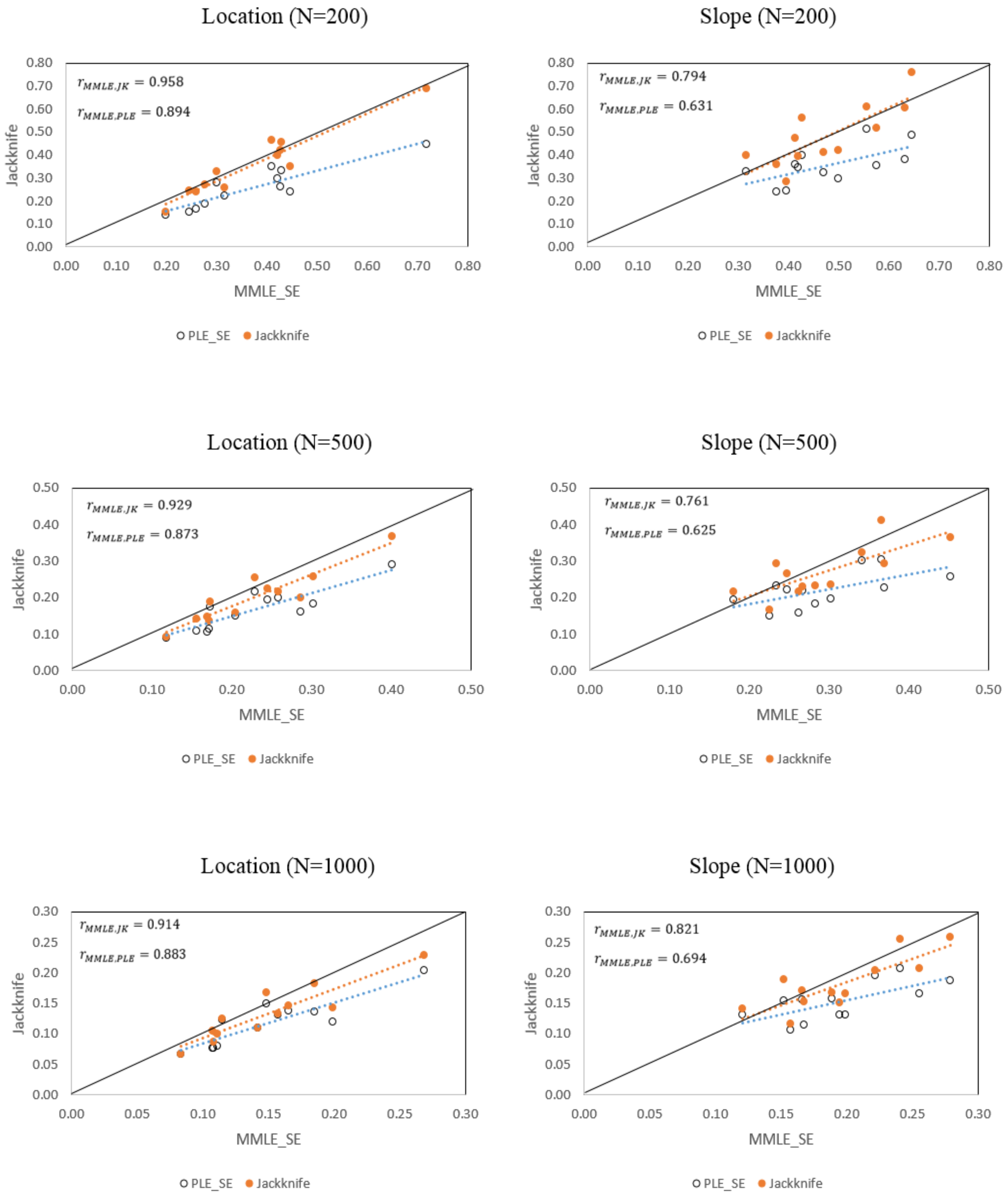


Figure 32. Jackknife standard errors from PLE vs. MMLE standard errors for 6 items and 3 response categories

(c) 5 categories

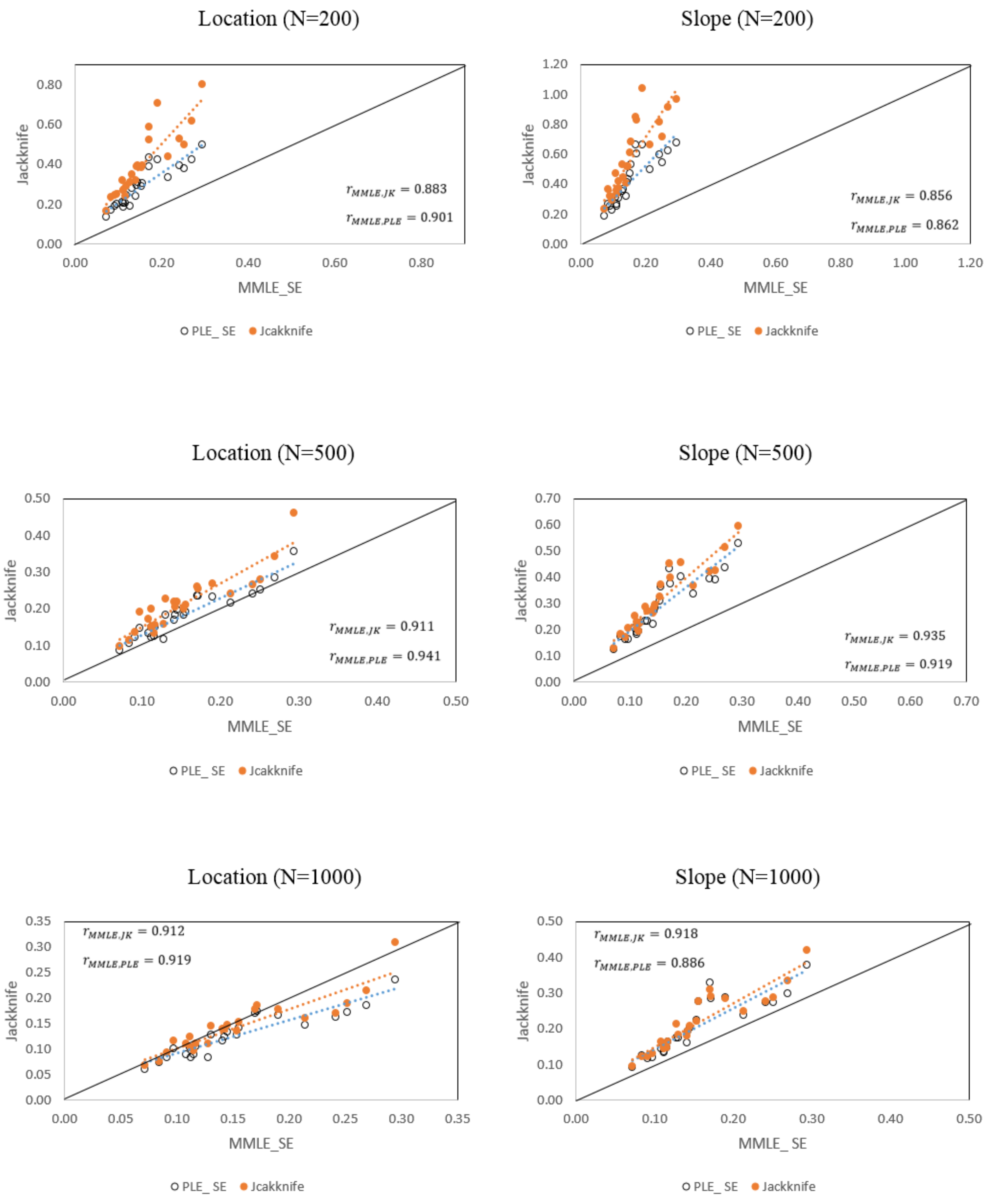


Figure 33. Jackknife standard errors from PLE vs. MMLE standard errors for 6 items and 5 response categories

jackknife standard errors for parameter estimates from PLE, and the lines are trend lines. The standard errors given by PLE are also plotted against the MMLE of IRT standard errors.

Overall, the correlations between the standard errors from the jackknife and MMLE are not as high as those between those from the jackknife and MLE. While the correlations between the jackknife standard errors and MLE standard errors were consistently high across all unidimensional models with 6 items, the correlations between those given by the jackknife and MMLE tend to vary with the models considered.

For the binary response models, the jackknife standard errors are moderately negatively correlated with MMLE standard errors for location parameters (i.e., $\mathrm{r} = -0.766, -0.770,$ and -0.740 for the sample size of 200, 500, and 1000, respectively). According to the plots of the jackknife standard errors against MMLE standard errors for location parameters of the binary response model with 6 items in Figure 31, PLE yielded very small standard errors over all 6 items, and some of the PLE standard errors are saliently small relative to those from MMLE. It also appears that only slight differences exist between the corrected standard errors by the jackknife and the original PLE standard errors. These findings may suggest that only small amount of improvement would be expected in correcting PLE standard errors by the jackknife procedure when they are extremely small. On the other hand, the standard errors for slope parameters estimated by the jackknife and MMLE are strongly positively correlated, yielding $\mathrm{r} = 0.99$ in all three binary response models with 6 items and 3 different sample sizes.

For the polytomous response models, the correlations between the standard errors from jackknife and MMLE range from $\mathrm{r} = 0.883$ to $\mathrm{r} = 0.958$ for location parameters and $\mathrm{r} = 0.761$ to $\mathrm{r} = 0.935$ for slope parameters. For the 3 response category models, the standard errors for location parameters between the jackknife and MMLE show as high correlations as $\mathrm{r} = 0.91$

above whereas those for slope parameters between two methods yield moderately high correlations. For the 5 response category models, it is illustrated in Figure 33 that MMLE yielded smaller standard errors than the jackknife overall. With the sample size of 1000, the correlations between the jackknife standard errors and MMLE standard errors are as high as $\mathrm{r} = 0.91$ for both location and slope parameters.

In sum, except for the location parameters of the binary response models, the jackknife standard errors showed strong correlations with MMLE standard errors under the models associated with the sample size of 1000, ranging from $\mathrm{r} = 0.821$ to $\mathrm{r} = 0.996$. It can be inferred that, with much larger sample sizes, the PLE standard errors corrected by the jackknife procedure would be close to those from MMLE. Further research on this is needed.

Computational Time of PLE for Large Numbers of Items

The simulation studies for large numbers of items were conducted on desktop computers with a 3.5 GHz CPU and 8 GB of RAM and a 3.6 GHz CPU and 16 GB of RAM. One problem occurred is that PLE wasn’t successful in fitting models with 50 items with a 3.5 GHz CPU and 8 GB of RAM. Therefore, the desktop computer with a 3.5 GHz CPU and 8 GB of RAM was used for 20 items and the one with a 3.6 GHz CPU and 16 GB of RAM for 50 items.

Table 25 summarizes the computational time of PLE for 20 and 50 items. Based on the convergence criteria described in Chapter 4, the number of iterations for convergence and computational time were recorded for each of 30 replications over all 72 uni- and multidimensional models with 20 and 50 items. The table contains the average number of iterations and computational time for convergence. Generally, different replication data sets showed different number of iterations for convergence. To simplify the information, the average number of iterations for convergence was computed over 30 replications.

Table 25. Computational time of PLE for large numbers of items

	20 items (3.5 GHz CPU & 8 GB of RAM)				50 items (3.6 GHz CPU & 16 GB of RAM)			
	No. of iterations for convergence			Average computational time for convergence	No. of iterations for convergence			Average computational time for convergence
Dim.	Min.	Max.	Average		Min.	Max.	Average	
2 categories								
Sample size (N) =200								
1D	12	14	13	23 secs	10	16	13	1 min 45 secs
2D	11	17	14	55 secs	14	20	17	3 mins 30 secs
3D	31	93	57	3 mins 10 secs	19	32	23	4 mins 30 secs
4D	28	90	48	2 mins 20 secs	26	50	37	8 mins
Sample size (N) = 500								
1D	11	13	13	39 secs	11	14	13	3 mins 12 secs
2D	12	16	14	1 min	14	20	16	6 mins
3D	11	18	16	1 min	17	25	21	7 mins 30 secs
4D	26	55	36	2 mins 30 secs	25	44	33	13 mins 45 secs
Sample size (N) = 1000								
1D	11	14	13	1 min 5 secs	11	13	12	6 mins 30 secs
2D	14	16	14	1 min 15 secs	11	16	14	10 mins 30 secs
3D	14	22	18	1 min 35 secs	17	22	20	12 mins
4D	23	49	32	4 mins	24	32	28	21 mins
3 categories								
Sample size (N) =200								
1D	11	14	12	42 secs	10	13	12	4 min 50 secs
2D	11	20	15	1 min 10 secs	11	15	14	5 mins
3D	15	27	21	1 mins 30 secs	12	23	16	7 mins
4D	28	60	45	3 mins 15 secs	18	38	27	11 mins 15 secs
Sample size (N) = 500								
1D	11	14	13	1 min 15 secs	10	12	12	12 mins
2D	12	16	14	1 min 30 secs	11	15	13	9 mins
3D	14	22	17	2 mins	7	16	13	9 mins 30 secs
4D	24	58	39	5 mins	18	37	26	24 mins
Sample size (N) = 1000								
1D	11	14	13	2 mins 10 secs	10	13	12	23 mins 15 secs
2D	12	15	14	2 min 30 secs	11	14	12	20 mins
3D	14	20	17	3 min 30 secs	11	16	14	24 mins 15 secs
4D	22	46	33	8 mins	21	30	25	50 mins

Table 25. (cont.)

	20 items (3.5 GHz CPU & 8 GB of RAM)				50 items (3.6 GHz CPU & 16 GB of RAM)			
	No. of iterations for convergence			Average computational time for convergence	No. of iterations for convergence			Average computational time for convergence
Dim.	Min.	Max.	Average		Min.	Max.	Average	
				5 categories				
				Sample size (N) =200				
1D	10	13	12	58 secs	10	12	11	6 min 25 secs
2D	10	17	13	1 min 20 secs	10	17	13	10 mins
3D	7	28	15	1 mins 40 secs	9	15	11	11 mins
4D	14	27	21	2 mins 10 secs	10	17	14	12 mins
				Sample size (N) = 500				
1D	11	13	13	2 min 50 secs	9	11	11	17 mins 35 secs
2D	10	14	13	3 mins	11	13	13	30 mins
3D	10	15	13	3 mins	8	12	11	26 mins
4D	13	27	19	5 mins	10	15	12	28 mins
				Sample size (N) = 1000				
1D	11	14	12	6 mins	10	13	12	48 mins 50 secs
2D	11	14	13	6 min 30 secs	11	13	13	2 hours 45 mins
3D	10	14	13	6 min 45 secs	10	12	11	2 hours 20 mins
4D	13	28	19	12 mins	10	14	12	2 hours 32 mins

Each time the analysis was done by PLE with a different replication data set, time per iteration with the data set was also measured. The average computational time for each condition was computed by multiplying the average number of iterations for convergence and time (in seconds or minutes) per iteration together. For example, the average computational time, 6 minutes 30 seconds for two-dimensional models with 20 items, 5 categories, and 1000 people was obtained by 30 seconds (time per iteration) × 13 (average number of iterations for convergence). In addition to these two pieces of information, the minimum and maximum number of iterations for convergence are also presented in the table, which were obtained from 30 replication data sets for each condition.

For unidimensional models with 20 items and the largest sample size of 1000 people, the convergence occurred in 1 minute, 2 minutes, and 6 minutes for 2 categories, 3 categories, and 5 categories, respectively. Unidimensional models with 20 and 50 items were also fit with flexMIRT for binary items and MULTILOG for polytomous items to compare PLE with MMLE in terms of parameter recovery, and it took about 1 to 5 seconds to get estimates of unidimensional models with those software. Even though PLE takes more time to fit unidimensional models with 20 items than flexMIRT or MULTILOG does, the computational time appears to be reasonable, which ranges from 23 seconds to 6 minutes.

For multidimensional models with 20 items, 4-dimensional models with 1000 people resulted in the longest computational time, which are 4 minutes, 8 minutes, and 12 minutes for 2 categories, 3 categories, and 5 categories, respectively. It implies that parameters of all other models with a smaller number of latent variables than 4 or smaller sample sizes than n = 1000 would converge in less than 12 minutes by PLE using a computer with a 3.5 GHz CPU and 8 GB of RAM.

The models with 50 items need much more time and better computers for PLE. With 4-dimensions and the largest sample size of 1000, it required 21 minutes and 50 minutes for 2 categories and 3 categories. For 5 categories with 1000 people, the convergence occurred in less than 3 hours with the average number of iterations of 11 to 13. When considering the complexity of the latent structures and the performance of SAS, the computational time of PLE for multidimensional models with large numbers of items is reasonable.

As anticipated, the models associated with longer test lengths, more latent variables, and larger sample sizes generally require more computational time. Given the number of items, categories, and latent variables, however, the number of iterations for convergence is similar across the sample sizes. Given the number of categories and sample size, the models with more complicated latent structures would be expected to need more time until parameter estimates converge. The simulation studies show, however, that this anticipation may not occur in all models. For 20 items given the number of categories and sample sizes, similar amount of computational time is found among uni-, 2-, and 3-dimensional models while much more time is required in 4-dimensional models. It is also found in the uni- and multidimensional models with the sample size of 500 and 1000 given 50 items and 3 categories. For 50 items and 5 categories given the sample size, there is a big gab in computational time between uni- and 2-dimensional models, but small amount of increase in time among multidimensional models. It suggest that item parameters of LMA models could be estimated by PLE with reasonable amount of increase in time as the number of latent variables increases given the same number of items, categories, and sample sizes.

Chapter 7

Empirical Studies

Two sets of empirical studies were conducted to demonstrate the practical use of PLE for uni- and multidimensional LMA models. Again, the focus is on showing that PLE behaves similarly to MLE and estimates parameters successfully for long tests where MLE is infeasible. To achieve these goals, both small and large numbers of items were considered in each set of empirical studies. For the empirical studies, the dataset from a study on aggression during early adolescence (Espelage, Holt, & Henkel, 2003) was used. The dataset consists of polytomous responses to 18 items assessing bullying, victimization, and fighting and they were obtained from 384 middle school students (196 boys and 188 girls). The response categories to these items are “Never”, “1 or 2 times”, “3 or 4 times”, “5 or 6 times”, and “7 or more times”. All analyses were conducted on a desktop computer with a 3.5 GHz CPU and 8 GB of RAM.

Unidimensional Models with Bullying Items

For the analysis of unidimensional models, the 9 items were chosen from bullying sub-scale (Espelage, Holt, & Henkel, 2003):

- *I upset other students for the fun of it.*
- *I teased other students.*
- *In a group I teased students.*
- *I helped harass other students.*
- *I was mean to someone when angry.*
- *I spread rumors about others.*
- *I started arguments or conflicts.*
- *I encouraged people to fight.*
- *I excluded others.*

Unidimensional Model with Small Numbers of Items

The first portion of the empirical study for unidimensional models focuses on showing how well Step 1 of the algorithm performs for unidimensional models using existing data as well as how similarly PLE behaves to MLE. To achieve this goal, the responses by 319 students to the first 4 items of 9 items on bullying sub-scale were analyzed using Step 1 of PLE. Since the model involves 4 items and 5 response categories, the number of possible item response patterns equals $5^4 = 625$, which MLE is still feasible. The number of response patterns that were actually responded by the students is 86. For identification in this study, λ_{ij} and ν_{ij} are centered as the location constraint (i.e., $\sum_j \lambda_{ij} = 0$ and $\sum_j \nu_{ij} = 0$) and $\sigma_{11} = 1$ as the scaling constraint.

In Figure 34, the location and slope parameter estimates from PLE are plotted against the corresponding estimates from MLE for 4 bullying items. In the plots, the x- and y-axis represent the parameter estimates from MLE and PLE, respectively, and the lines are identity lines. The estimated parameters by PLE are very close to the corresponding estimated parameters by MLE with r = .997 for location parameters and r = .998 for slope parameters.

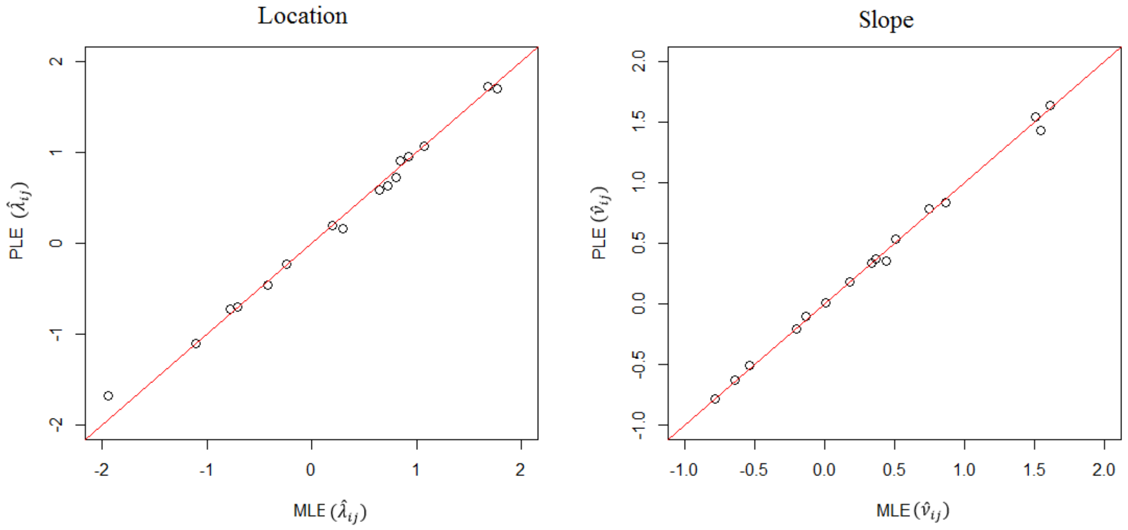


Figure 34. Parameter estimates of a unidimensional model from PLE vs. MLE for 4 bullying items

The jackknife procedure was performed to correct the standard errors for the estimates from PLE and to compare the corrected standard errors with those from MLE. Since the sample size for the unidimensional model with 4 bullying items was *n* = 319, PLE was performed 319 times, leaving one subject out of the sample. To ensure full convergence, the number of iterations was set to 40 and it required 1 hour 40 minutes of run time to complete the jackknife procedure. The item parameter estimates with 319 different data sets converged in 20 ~ 25 iterations. With the actual number of iterations for convergence, the expected run time to complete jackknife procedure would be 50 ~ 60 minutes.

Figure 35 shows the plots of the jackknife standard errors for the parameter estimates from PLE against the MLE standard errors for 4 bullying items. In the plots, the x-axis represents the MLE standard errors, y-axis represents the jackknife standard errors for parameter estimates from PLE, the symbols represent specific parameters, and the lines are identity lines. The PLE standard errors are also plotted against the MLE standard errors to examine the standard errors from PLE and MLE before the jackknife procedure is performed.

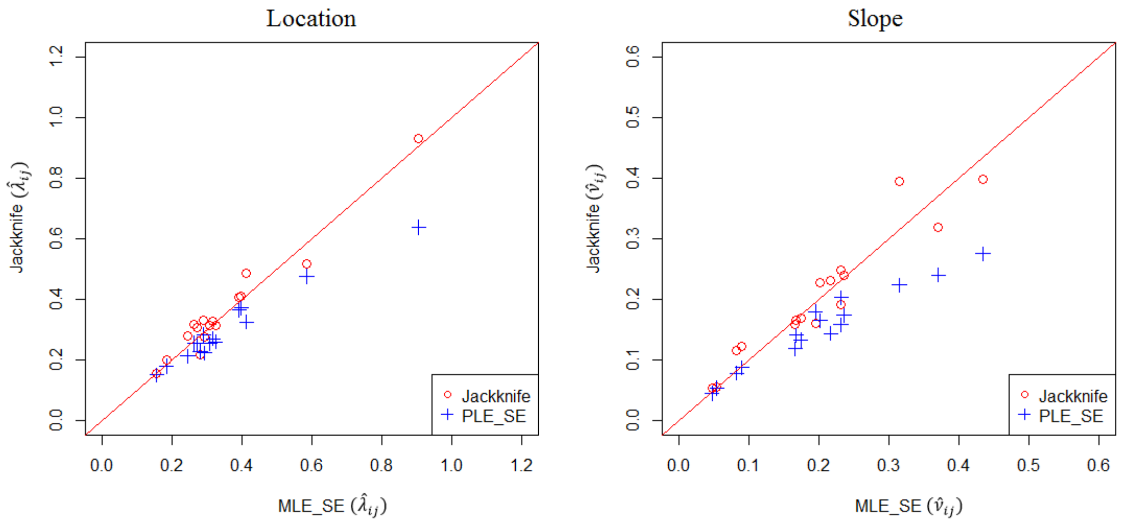


Figure 35. Jackknife standard errors from PLE vs. MLE standard errors for 4 bullying items

The jackknife standard error estimates from PLE are very close to those from MLE for location and slope parameters, which was also found in the simulation studies. The correlation coefficients between the corrected standard errors by jackknife and the corresponding MLE standard errors are $\mathrm{r} = .978$ for location parameters and $\mathrm{r} = .952$ for slope parameters, showing that the jackknife standard error estimates are strongly correlated with MLE standard errors across 4 bullying items. In the Figure, the small PLE standard errors relative to MLE standard errors indicate that the standard errors are underestimated by PLE, but they still tend to be close to the MLE standard errors.

Unidimensional Model with Large Numbers of Items

For large numbers of items, the MLE fails because the number of possible response patterns increases exponentially as the number of items and response options per item increase. For example, the number of possible response patterns of the model used in the first empirical study is $5^4 = 625$ and it grows to $5^5 = 3{,}125, \ldots$; $5^8 = 390{,}625$; and to $5^9 = 1{,}953{,}125$ each time one more items is added to the model, which makes the MLE of LMA models infeasible.

The second portion of the empirical study for unidimensional models was performed to illustrate that PLE overcomes the limitation of MLE when fitting unidimensional LMA models for polytomous items with large numbers of items. For this study, the responses by 315 students to 9 bullying items were analyzed, repeating the same procedure in the first portion of the empirical study (i.e., only Step 1). The number of response patterns that were actually responded by the students is 186.

With the results of the simulation studies in previous chapter, the first analysis for a unidimensional model with 4 bullying items has also shown that the parameter estimates from PLE using only Step 1 and their standard errors corrected by jackknife are very close to those from MLE. Therefore, the parameter estimates obtained by PLE and the jackknife standard error estimates for 9 bullying items can be treated as those that are close to those from MLE.

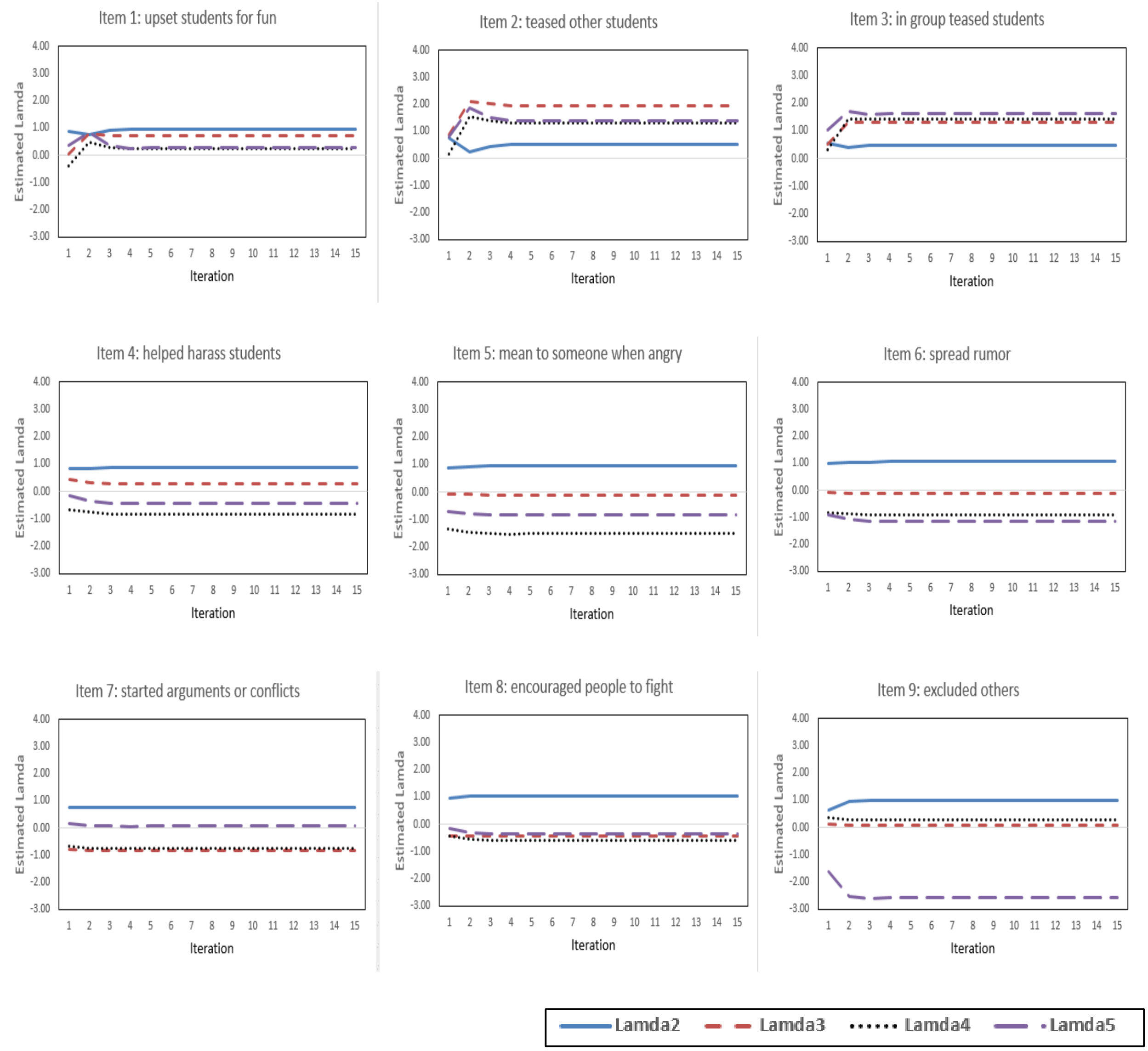


Figure 36. Estimated location parameters of a unidimensional model by PLE for 9 bullying items

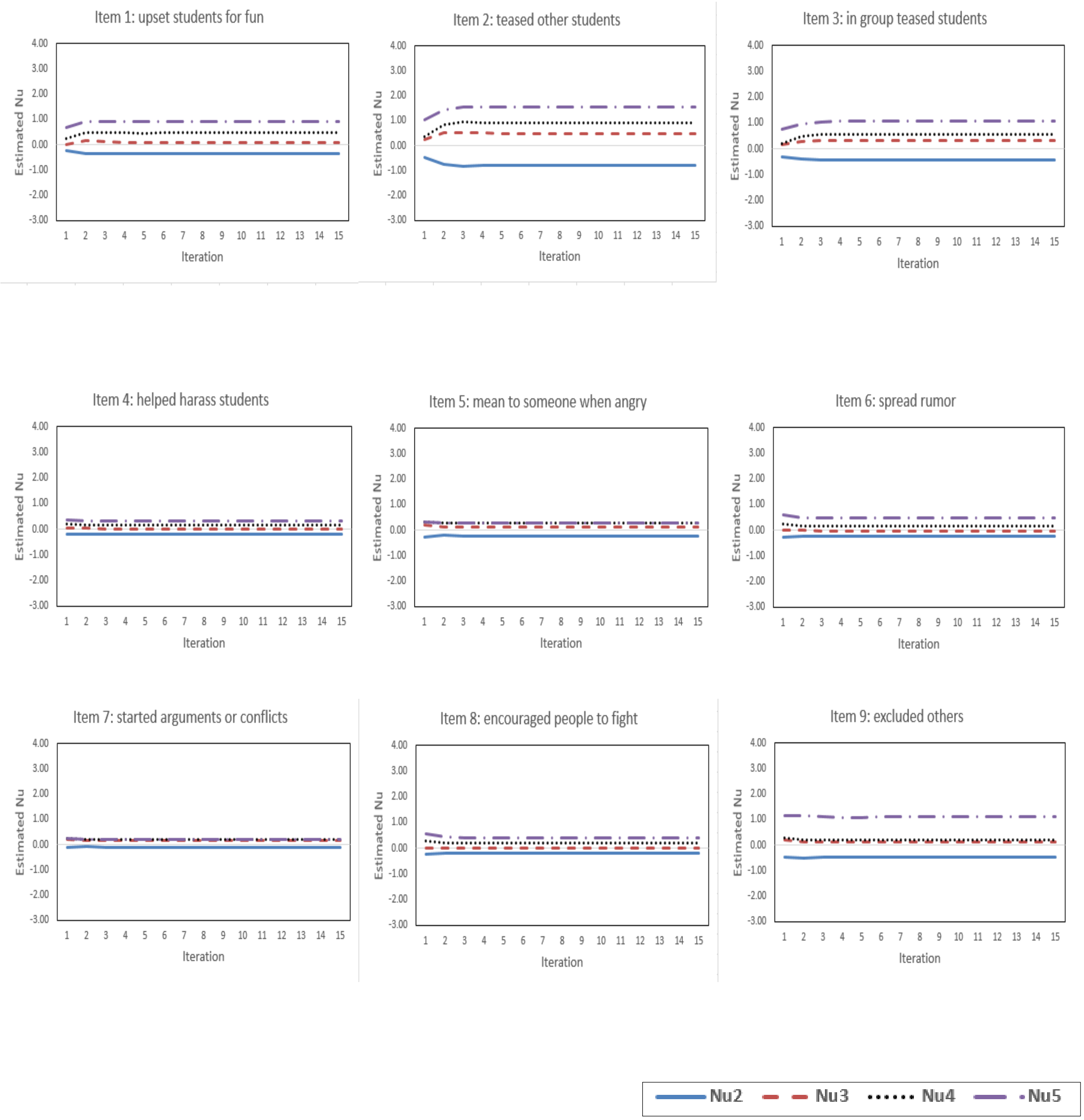


Figure 37. Estimated slope parameters of a unidimensional model by PLE for 9 bullying items

The estimated location and slope parameters for each item by PLE are illustrated in Figures 36 and 37, respectively. The x-axis represents the number of iterations and the y-axis represents the estimated values for the parameters by iteration. The four lines in the legend on each figure correspond to response categories, “1 or 2 times”, “3 or 4 times”, “5 or 6 times”, and “7 or more times”, respectively. Most of the location and slope parameter estimates for 9 bullying items converged in 15 iterations, and it took about 20 seconds for convergence on a desktop computer with a 3.5 GHz CPU and 8 GB memory. It indicates that PLE can estimate parameters very fast for unidimensional models with large numbers of items.

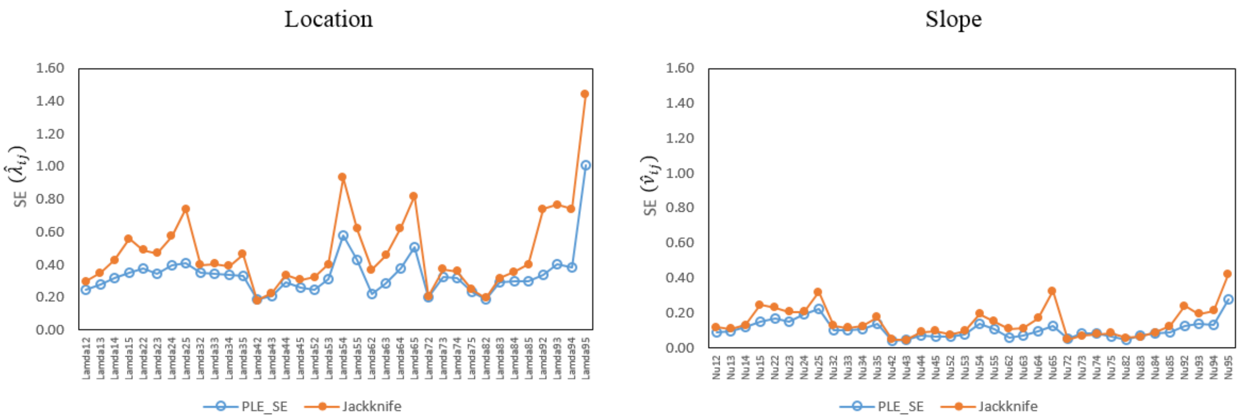


Figure 38. Jackknife standard errors from PLE vs. PLE standard errors for 9 bullying items

Figure 38 shows the plots of the jackknife standard errors from PLE and the original PLE standard errors for 9 bullying items. In the plots, x-axis presents the location and slope parameters to be estimated and y-axis represents the standard error estimates from the jackknife and PLE. With the sample size of 315, 9 items, and the number of iterations of 35 for full convergence, it required 4 hours of computation time on a desktop computer with a 3.5 GHz and 8 GB of RAM to finish the jackknife procedure.

As expected, the PLE standard errors tend to be smaller than the corresponding jackknife standard error estimates. The PLE standard errors ranged from 0.1856 to 1.0073 for location parameters and 0.0400 to 0.2763 for slope parameters. The ranges, however, has changed by jackknife procedure from 0.1791 to 1.4453 for location parameters and 0.0447 to 0.4215 for slope parameters. The wider ranges of the jackknife standard error estimates imply that the underestimated standard errors from PLE were corrected by jackknife, which is also supported by simulation studies.

Multidimensional Models with Bullying and Victimization Items

The second set of the empirical studies involves multidimensional models and was intended to illustrate how well full algorithm of PLE (i.e., Steps 1 and 2) performs for multidimensional models using existing data. For the analysis of multidimensional models, four items from victimization sub-scale were added to the 9 bullying items used for unidimensional models and two correlated latent variables, bullying and victimization were assumed. The added victimization items are:

- *Other students picked on me.*
- *Other students made fun of me.*
- *Other students called me names.*
- *I got hit and pushed by other students.*

Multidimensional Model with Small Numbers of Items

The first portion of the empirical study for multidimensional models examines the similarity of the parameter estimates between PLE and MLE. For this analysis, the responses by 322 students to 3 items from bullying sub-scale and 3 items from victimization sub-scale were selected, and those two latent variables were assumed to be correlated. Since it is a simple multidimensional model with two items and two correlated latent variables, not only MLE was feasible but also the performance of the full algorithm of PLE was examined by comparing similarity of the parameter estimates obtained from PLE and MLE. For identification constraints, zero-sum constraint was imposed and σ_{11} and σ_{22} were set equal to 1 for scaling constraints, which means only σ_{12} is estimated. The number of response patterns that were actually responded by students is 165.

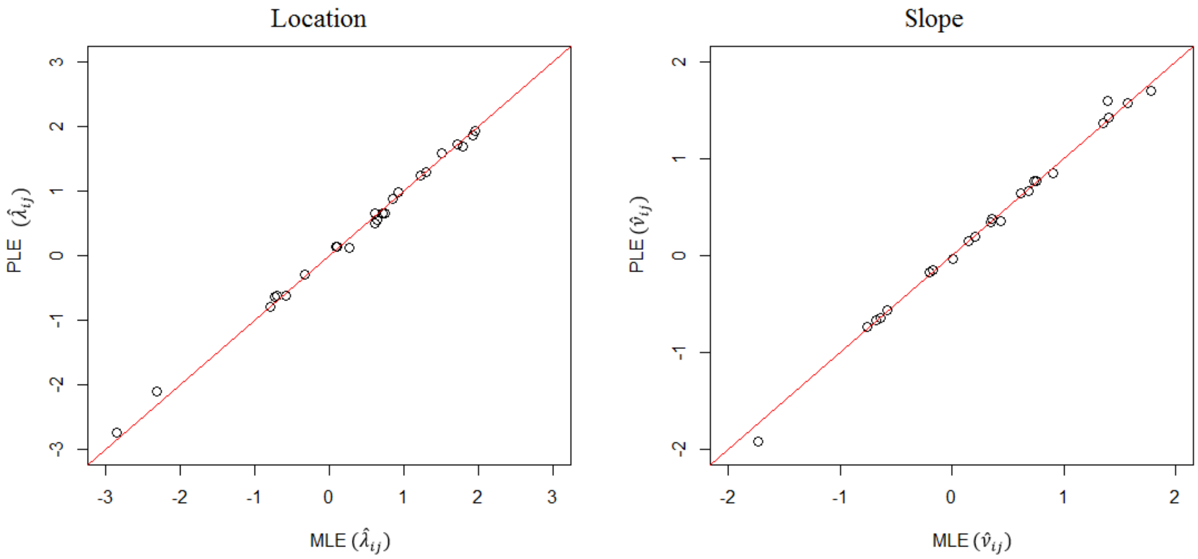


Figure 39. Parameter estimates of a multidimensional model from PLE vs. MLE for 6 bullying and victimization items

In Figure 39, the location and slope parameter estimates from PLE are plotted against the corresponding estimates from MLE for a multidimensional model with 6 items and two correlated latent variables. In the plots, the x- and y-axis represent the parameter estimates from MLE and PLE, respectively, and the lines are identity lines. The estimates for location and slope parameters from PLE are almost identical to the corresponding estimates from MLE. The correlations between PLE and MLE are $\mathrm{r} = 0.998$ for location parameters and $\mathrm{r} = 0.999$ for slope parameters. The estimated association parameter (σ_{12}) is $\sigma_{12} = 0.006$ from PLE and $\sigma_{12} = 0.003$ from MLE. The standard errors of the estimated association parameters from PLE and MLE are S.E.=0.1717 and S.E.=0.0054, respectively. Based on the standard errors, it appears that two latent variables, bullying and victimization are not correlated.

Multidimensional Model with Large Numbers of Items

The second portion of the empirical study for multidimensional models involves large numbers of items to demonstrate that PLE overcomes the limitation of MLE that it fails to estimate parameters of LMA models with large numbers of items. As in the first portion of the empirical study for multidimensional models, parameter estimates were obtained using the full algorithm of PLE for a multidimensional model with a 9 bullying items, 4 victimization items, and two correlated latent variables. The number of response patterns that were actually responded by 322 students is 133.

Based on the results of the simulation studies and the empirical study with 6 bullying and victimization items, the parameter estimates obtained from PLE for the 13 items in the model can be treated as those that are close to those from MLE.

The estimated location and slope parameters for each item by PLE are illustrated in Figures 40 and 41, respectively. The x-axis represents the number of iterations and the y-axis represents the estimated values for the parameters by iteration. The four lines in the legend on each figure correspond to response categories, “1 or 2 times”, “3 or 4 times”, “5 or 6 times”, and “7 or more times”, respectively. The rapid ascent (or descent) in parameter estimates are shown at the early stage of the estimation, but most of the parameter estimates for 13 bullying and victimization items converged in 25 iterations. One noticeable thing is that the order of response categories switch after first few iterations.

It shows that even if the initial order is not correct, PLE fixes it (i.e., PLE is not sensitive to starting values). For computational time, it took about 1 minute 10 seconds for convergence with a desktop computer equipped with 3.5 GHz CPU and 8 GB of RAM, indicating that PLE can also fit multidimensional models with large numbers of items very fast.

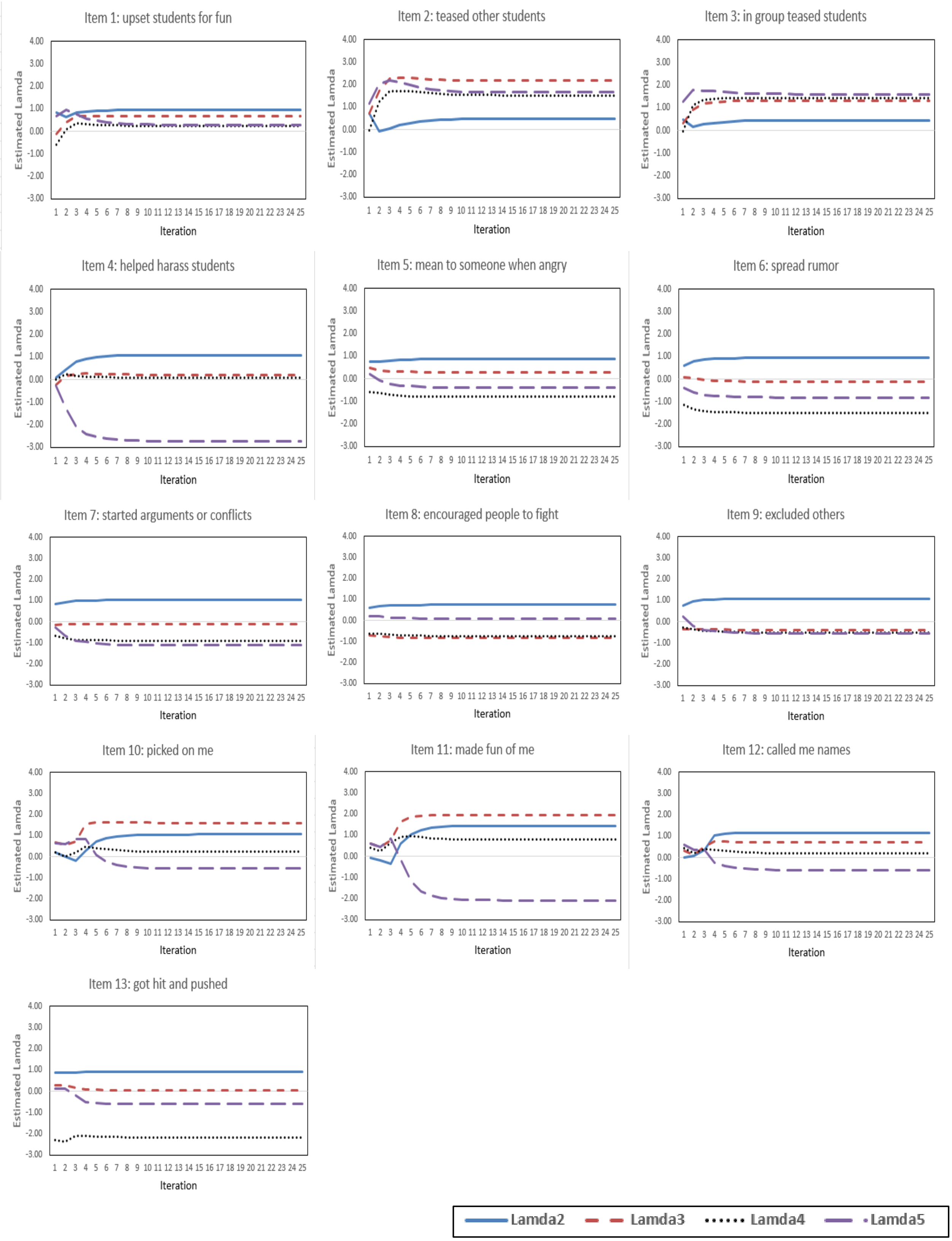

Item 1: upset students for fun
Item 2: teased other students
Item 3: in group teased students
Item 4: helped harass students
Item 5: mean to someone when angry
Item 6: spread rumor
Item 7: started arguments or conflicts
Item 8: encouraged people to fight
Item 9: excluded others
Item 10: picked on me
Item 11: made fun of me
Item 12: called me names
Item 13: got hit and pushed

Figure 40. Estimated location parameters of 2-dimensional model by PLE for bullying and victimization items

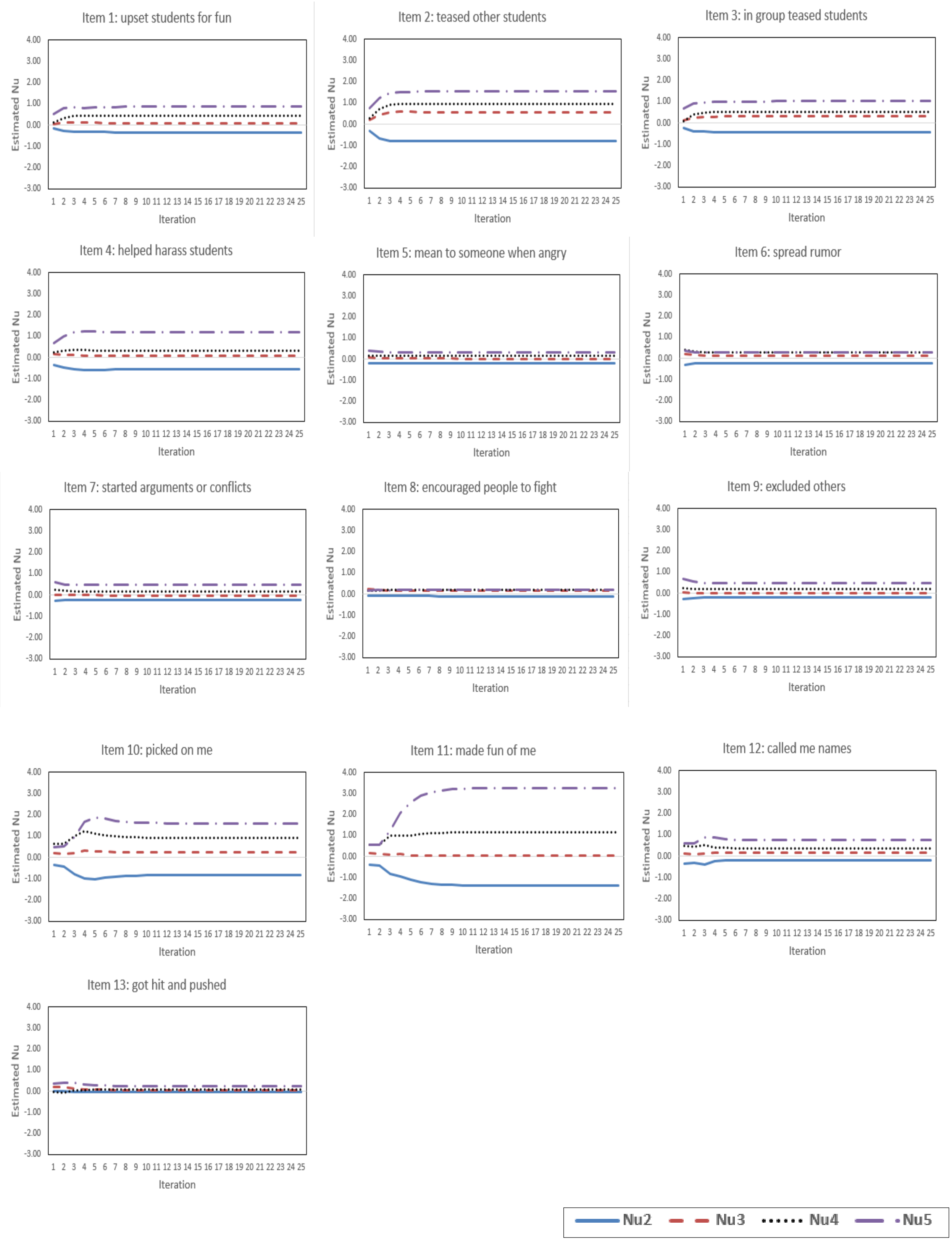


Figure 41. Estimated slope parameters of 2-dimensional model by PLE for bullying and victimization items

Chapter 8

Discussion and Conclusions

Throughout this study, the performance of the proposed PLE algorithm for more general LMA models as MIRT models was examined. Three main goals were set for the study: (1) how well does the newly proposed step for estimating slope parameters perform?; (2) how well does PLE of LMA models using the new two-step algorithm perform relative to MLE of LMA models?; and lastly, (3) how well and fast does the algorithm of PLE perform for LMA models as MIRT models with large numbers of items? In this chapter, the findings and their implications about these goals are summarized. Additionally, the possible further research is provided.

Estimation of Slope Parameters by PLE

The first goal of the study was to determine that how well slope parameters are estimated by the newly proposed step (i.e., Step 1). The parameter estimation requires only Step 1 of the full PLE algorithm for unidimensional LMA models while it does full steps (i.e., Step 1 and Step 2) of PLE for multidimensional LMA models. In this section, the findings of the simulation studies for uni- and multidimensional models with small numbers of items are provided, focusing on the estimation of slope parameters of PLE.

Unidimensional Models for Small Numbers of Items

Parameters of unidimensional LMA models were estimated by using only Step 1 of PLE. The mean bias of PLE for slope parameters were relatively small, where the absolute values of the mean bias were less than 0.04 for most of unidimensional models. It implies that the estimated slope parameters of unidimensional LMA models by PLE were almost close to the corresponding parameters used to simulate the data and that the newly proposed step of PLE (i.e., Step 1) performs successfully in estimating slope parameters of unidimensional LMA models.

The excellence of Step 1 of PLE in slope parameter estimation is supported again when the mean bias values and correlations for slope parameters are compared with those for location parameters of unidimensional models with small numbers of items. Overall, the simulation conditions with the models resulted in smaller mean bias values and higher correlations for slope parameters than those for location parameters. It suggests that PLE using Step 1 yielded more unbiased estimates for slope parameters than for location parameters of unidimensional LMA models.

Simulation studies also showed that PLE performed as well as MMLE or outperformed MMLE in slope parameter estimation. For several unidimensional models with 6 items, the mean bias values of PLE for slope parameters were as small as MMLE. Another successful example was found in the comparison of the mean RMSE values of PLE with those of MMLE. For polytomous items, the mean RMSE values of PLE for slopes were equal to or smaller than those of MMLE, suggesting that not only the new step of PLE estimated slope parameters successfully but also its performance was as good as MMLE for unidimensional polytomous models.

Multidimensional Models for Small Numbers of Items

The simulation studies involving multidimensional LMA models with small numbers of items have shown that the slope parameters of the models were successfully estimated by PLE using full steps (i.e., Steps 1 and 2). Especially, PLE was more excellent in polytomous items than in binary items when it estimated slope parameters of the models. More of smaller mean bias values and higher correlations for slope parameters were found in the conditions associated with polytomous items. Based on these findings, it can be inferred that the estimation of slope parameters of PLE would be better for polytomous items than binary items when considered multiple latent variables.

In summary, the previous PLE implemented for LMA models in the Rasch family has been successfully extended to more general models (i.e., 2PL). The slope parameters of LMA models have been estimated well by PLE using the newly proposed step (i.e., Step 1) which performs conditional (multinomial) logistic regression for each item. In addition, the performance of the extended PLE was as good as MMLE for unidimensional polytomous models.

Comparison of PLE with MLE

The second goal of the study was to demonstrate that how similarly PLE behaves relative to MLE (the gold standard) in estimating parameters of LMA models. Since MLE is feasible to estimate item parameters of LMA models with small numbers of items, both simulation and empirical studies for small numbers of items were conducted and the performance of PLE was compared with MLE.

In terms of similarity of the estimates between PLE and MLE, the correlations between the parameter estimates from PLE and MLE were 0.999 or 1.000 across all simulation conditions with 4 and 6 items, indicating the parameter estimates between PLE and MLE are highly linearly related. There were also found very small mean RMSDiff of the parameter estimates between MLE and PLE, suggesting that that the parameter estimates obtained from MLE and PLE are not only linearly related but also nearly equivalent (i.e., equal to 2 or 3 decimal points).

The results of the empirical studies with small numbers of items also support that PLE behaves similarly to MLE. As illustrated in Figure 30 for a unidimensional model with 4 bullying items and in Figure 34 for a two-dimensional model with 3 bullying and 3 victimization items, all of the estimates from PLE plotted against those from MLE fell on 45 degree line on the figures with $r = .99$, indicating the parameter estimates by PLE are very close to the corresponding estimates by MLE for all parameters.

In summary, it has been shown that the parameter estimates from PLE are almost identical to those from MLE with the same amount of estimation accuracy in both simulation and empirical studies for small numbers of items.

Performance and Estimation Time of PLE for Large Numbers of Items

The third goal of the study was to examine how well PLE overcomes the limitation of MLE when fitting LMA models for large numbers of items and also how fast the parameter estimates of PLE converge for the models. In this section, the findings of the simulation studies for uni- and multidimensional models with large numbers of items are described, followed by the computational time of PLE.

Unidimensional Models for Large Numbers of Items

Simulation studies for unidimensional models with large numbers of items showed that PLE yielded nearly unbiased item parameter estimates and very high correlations between estimates and true parameters in most of the conditions. Especially, PLE showed more excellent performance in the conditions for polytomous items when it recovered the parameters used to simulate the data.

The results of the simulation studies also support that PLE performed as well as MMLE for unidimensional models with large numbers of items. Overall, MMLE consistently yielded smaller mean bias values than PLE. For all conditions with 5 category items, however, most of the mean bias values of PLE were as small as those of MMLE, where the absolute values of the mean bias of both estimation methods are less than 0.02 for location parameters and 0.01 for slope parameters.

When comparing the mean RMSE of PLE with those of MMLE, the mean RMSE values of PLE for polytomous items were slightly larger than those of MMLE. However, for binary items with sample sizes of 500 and 1000, PLE showed smaller mean RMSE values than MMLE, indicating that PLE recovered location and slope parameters used to simulate the data with less amount of estimation errors than MMLE. Even though these patterns were found in only several conditions of the unidimensional models for large numbers of items, the results imply that PLE not only performs well but also it may behave as well as MMLE for large numbers of items.

Multidimensional Models for Large Numbers of Items

PLE resulted in unbiased parameter estimates for location and slope parameters across most of the multidimensional models with large numbers of items. When comparing the mean bias values of PLE by number of categories, the models involving 5 response category items resulted in the most unbiased estimates for location and slope parameters. It appeares that sample size affects estimation accuracy but the number of dimensions doesn’t. Given the test length and the number of categories, smaller mean RMSE values for the estimates of PLE were generally associated with larger sample sizes. When the number of dimensions was varied, however, any consistent pattern was not found in terms of estimation accuracy. For example, there were irregular increase or decrease in mean RMSE values for location parameters but similar mean RMSE values for slope parameters across all dimensions.

In terms of correlation between estimates of PLE and true parameters, PLE consistently showed as high correlations as 0.957 to 0.998 for location parameters and 0.973 to 0.997 for slope parameters. Overall, more numbers of higher correlations were found in slope parameters than in location parameters. As the number of dimensions increased, there were some changes in the correlations for location parameters. The correlations for slope parameters, however, stayed similar across all the dimensions.

In the previous studies (Anderson, Li, & Vermunt, 2007; Li, 2010), it was shown that PLE was feasible for large number of items, but their application of PLE was limited to the models in Rasch family. Anderson (2013) fit more general LMA models where slope parameters were estimated by MLE using nonlinear programming (e.g., Proc NLP in SAS) and reported that the estimation of LMA models for eight 5-category items was successful, but it failed for nine 5-category items. In this study, the extended PLE for more general LMA models successfully estimated both location and slope parameters of LMA models for large numbers of items, overcoming the limitation of MLE. It yielded nearly unbiased estimates and high correlations between the estimates and true parameters and performed better for polytomous items than for binary items. For unidimensional models with large numbers of items, it worked as well as MMLE.

Computational Time of PLE for Large Numbers of Items

The PLE algorithm was implemented in SAS using a series of SAS macros. From the simulation studies for large numbers of items, the computational time of PLE appeared to be reasonably good. For all of the uni- and multidimensional models with 20 items, the estimation time of PLE ranged from 23 seconds to 12 minutes using a desktop computer with a 3.5 GHz CPU and 8 GB of RAM. For 50 items, PLE was performed on a better desktop computer with a 3.6 GHz CPU and 16 GB of RAM and the computational time ranged from 1 minute 45 seconds to 2 hours 45 minutes.

PLE also performed fast with real datasets for large numbers of items. The computer used for empirical studies was equipped with a 3.5 GHz CPU and 8 GB memory. For a unidimensional model with 9 bullying items, 5 response categories, and the sample size of 315, most of the location and slope parameter estimates converged in 15 iterations and it took about 20 seconds for convergence. For 2-dimensional models with 13 items (i.e., 9 bullying and 4 victimization items), 5 response categories, and the sample size of 322, the parameter estimates converged in 1 minute 10 seconds for convergence with 25 iterations.

One of the issues that have been reported often regarding the estimation of MIRT models is that computational time increases exponentially with the number of latent variables (Wang, Chen, & Cheng 2004; Glass 2005). In this study, PLE provides a promising approach for this issue. Of 72 simulation conditions for large numbers of items, the models that would be expected to need much more time for convergence were the ones involving 50 items, 5 categories, and 1000 people. On average, the parameter estimates of PLE for these models converged in 13 iterations and it took less than 50 minutes for a unidimensional model and 3 hours for multidimensional models until convergence. One noticeable thing is that the computational time of PLE doesn’t increase sharply, but moderately, or even stays similar among the models as the number of dimensions increases. The moderate amount of increase in estimation time is assumed to be due to the advantage of LMA models that they don’t require multiple numerical integrations in parameter estimation for MIRT models. When considering the heavy computational work in MMLE/EM algorithm and Bayesian estimation procedure with Markov chain Monte Carlo (MCMC) methods for higher dimensionality, the findings suggest that PLE can be an alternative way to estimate parameters in MIRT models with less heavy computational work and time.

Future Direction

One extension of this study is to compare the performance of PLE with the current standard IRT estimation method (i.e., MMLE) which was limited to unidimensional models. To get the estimates of unidimensional models by MMLE, nonlinear mixed procedure in SAS (i.e., Proc NLMIXED) was performed for small numbers of items and flexMIRT and MULTILOG for large numbers of items. Several software packages were developed for MIRT model estimation (e.g., Mplus, IRTPRO, flexMIRT, EQSIRT, etc.) and they are commercially available. They offer different estimation algorithms for MIRT models, which are based on marginal maximum likelihood or Bayesian approaches. It would be interesting to examine and compare the performance of PLE with other standard MIRT estimation procedures with respect to parameter recovery and run time. Even though great progress in development of software packages and computer hardware has enabled MIRT model estimation, the heavy computational work still remains due to numerical integrations, leading to lack of flexibility in estimation such as limitation of dimensionality, categories, sample size, items, and so on.

Another area for future work is the standard error estimation of PLE. The standard errors of the estimates from PLE are underestimated because the dependency among responses from a single person that exists in a stacked dataset is ignored when the MLE of the stacked regression procedure is performed. In this study, jackknife procedure was conducted to obtain correct standard errors of PLE, but only unidimensional models with small numbers of items were considered for the procedure. As known, the jackknife procedure was also extremely time consuming for PLE in this study. PLE algorithm implemented in this study allows us to control the number of iterations for convergence. In this study, the number of iterations was set to 30 to ensure full convergence and three desktop computers were used. With 5 categories and the sample size of 1000, it required about 8 hours for 4 items and 13 hours for 6 items per replication. Since each simulation condition was replicated 30 times, almost seven month was required to complete jackknife procedures for 18 unidimensional models with small numbers of items. Since jackknife procedure continues by removing one observation from the sample at a time and re-estimating parameters until the very last observation of the sample is left out, the computational time increases exponentially as sample size gets larger. Improvement in efficiency of the current SAS macros that were written to implement the PLE algorithm would partially relieve the computational time to run jackknife procedure. In addition to this, more theoretical and technical explorations are needed to find a reasonable way to correct standard error of PLE.

Another area for future work is in more efficient programming of the algorithm in SAS or possibly R. There also exist a few of technical limitations in computer hardware and SAS macros. As mentioned earlier, when fitting models with 50 items by PLE, a better computer was required than the one used for 20 items. In this study, a desktop computer with a 3.6 GHz CPU and 16 GB of RAM was used for 50 items with any number of categories and sample sizes. It implies that much better computer would be needed when the number of items gets larger than 50 items. However, this is only temporary limitation because computers keep getting faster. Regarding the limitation of the SAS macros, no termination rules are implemented in the current PLE algorithm. The current SAS macros allow users to set the number of iterations by defining the value as one of SAS global macro variables for PLE execution. The parameter estimation is terminated when the algorithm has been iterated as many times as the number of iterations defined by users. During execution, SAS macros generate a series of datasets necessary for re-iterative process of the algorithm and history datasets for each item that contain up-dated estimates and log-likelihood ratio from each iteration cycle. By examining the history datasets for each item, users can decide whether or not item parameter estimates converge based on the convergence criteria described in Chapter 4, and if they converged, in what iteration number the convergence occurs. Users can reset the number of iterations with any new value in case the convergence doesn't occur. When considering that most of software packages employ termination rules for estimation and the estimation process ends automatically when the convergence of estimates is fully achieved, the current way of terminating PLE algorithm needs an improvement. Based on the termination criteria in Chapter 4, termination rules should be implemented in the algorithm so that some hassles caused by the current way can be removed.

Finally, throughout the study, it has been shown that PLE is easy to implement by using conditional logistic regression and can handle models with high dimensionality for polytomous items. For future studies, the algorithm could add options for LMA models with covariates, ordinal constraints on response categories (i.e., ν_{ij}'s), linearly transformed ν_{ij}'s , and so on. These inclusions in LMA models were examined by Anderson (2013) to propose LMA models as a MIRT model. The models with covariates, ordinal constraints, and linear transformation, however, were fit by MLE in the study. The future studies that include PLE will not only add the justification for the advantages of LMA models as latent variable models relative to traditional factor analytic and item response theory methods but also provide a solution that leads to better measurement.

References

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

Alasuutari, P., Bickman, L. B., & Brannen, J. (2008). *The Sage handbook of social research methods*. London, England: Sage Publications.

Andersen, E. B. (1995). Polytomous Rasch models and their estimation. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 271–291). New York, NY: Springer.

Anderson, C. J. (2002). Analysis of multivariate frequency data using graphical models and generalizations of the *RC*(*M*) association model. *Psychological Methods*, *7*, 447–467.

Anderson, C. J. (2013). Multidimensional item response theory models with collateral information as Possion regression models. *Journal of Classification*, *30*, 276-303.

Anderson, C. J., & Böckenholt, U. (2000). Graphical regression models for polytomous variables. *Psychometrika*, *65*, 497-509.

Anderson, C. J., & Vermunt, J. K. (2000). Log-multiplicative association models as latent variable models for nominal and/or ordinal data. *Sociological Methodology*, *30*, 81-122.

Anderson, C. J., & Yu, H. T. (2007). Log-multiplicative association models as item response models. *Psychometrika*, *72*, 5-23.

Anderson, C. J., Li, Z., & Vermunt, J. K. (2007). Estimation of models in a Rasch family for polytomous items and multiple latent variables. *Journal of Statistical Software*, *20(6)*. Retrieved from http://www.jstatsoft.org/v20/i06/paper.

Anderson, C. J., Verkuilen, J., & Peyton, B. L. (2010). Modeling polytomous item responses using simultaneously estimated multinomial logistic regression models. *Journal of Educational and Behavioral Statistics*, *35*, 422-452.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*, 561-573.

Arnold, B.C., & Strauss, D. (1991). Pseudolikelihood estimation: Some examples. *Sankhyā: The Indian Journal of Statistics*, *Series B*, *53*, 233-243.

Baker, F. B., & Kim, S.-H. (2004). *Item response theory: Parameter estimation techniques*. Monticello, NY: Marcel Dekker, Inc.

Bartholomew, D. J., & Knott, M. (1999). *Latent variable models and factor analysis* (2nd ed.). New York, NY: Oxford University Press.

Becker, M. P. (1989). On the bivariate normal distribution and association models for ordinal categorical data. *Statistics and Probability Letters, 8*, 435–440.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B*, *36*, 192-236.

Besag, J. (1975). Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society, Series D*, *24*, 179-195.

Bock, D.R., Gibbons, Schilling, S.G., Muraki, E., Wilson, D.T., and Wood, R. (2003). *TESTFACT 4.0 Computer Software and Manual*, Lincolnwood, IL: Scientific Software International.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37*, 29-51.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, *46*, 443-459.

Bock, R. D., & Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika*, *35*, 179-197.

Cai, L. (2013). flexMIRT$_R$ version 2: Flexible multilevel multidimensional item analysis and test scoring [Computer software]. Chapel Hill, NC: Vector Psychometric Group.

Chang, H. H. (1996). The asymptotic posterior normality of the latent trait for polytomous IRT models. *Psychometrika*, *61*, 445-463.

Chang, H. H., & Stout, W. (1993). The asymptotic posterior normality of the latent trait in an IRT model. *Psychometrika*, *58,* 37-52.

Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika*, *40*, 5–32.

Clogg, C. C., & Eliason, S. R. (1987). Some common problems in log-linear analysis. *Sociological Methods and Research, 16*, 8-44.

Coleman, J.S. (1964). *Introduction to mathematical sociology*. Glencoe, IL: Free Press.

De Boeck, P., & Wilson, M. (2004), *Explanatory item response models: A generalized linear and nonlinear approach.* New York, NY: Springer.

De la Torre, J., & Patz, R. J. (2005). Making the most of what we have: A practical application of multidimensional IRT in test scoring. *Journal of Educational and Behavioral Statistics*, *30*, 295-311.

Embretson, S. E. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika*, *56*, 495–515.

Espelage, D. L. and Holt, M. K. (2001), Bullying and victimization during early adolescence: Peer influences and psychosocial correlates. *Journal of Emotional Abuse*, *2*, 123–142.

Espelage, D. L., Holt, M. K., & Henkel, R. R. (2003). Examination of peer-group contextual effects on aggression during early adolescence. *Child Development*, *74*, 205–220.

Fischer, G. H. (1995). The derivation of polytomous Rasch models. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 293-305). New York, NY: Springer.

Glass, C.A.E. (2005). Review of de Bock, P. and Wilson, M. *Explanatory Item Response Theory Models: A Generalized Linear and Nonlinear Approach*", *Journal of Educational Measurement*, *42*, 303–307.

Gelman, A., & Speed, T. P. (1993). Characterizing a joint probability distribution by conditionals. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *55*, 185–188.

Geys, H., Molenberghs, G., & Ryan, L. (1999). Pseudolikelihood modeling of multivariate outcomes in developmental toxicology. *Journal of the American Statistical Association*, *94*, 734–745.

Goodman, L. A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association*, *74*, 537–552.

Goodman, L. A. (1985). The analysis of cross-classified data having ordered and/or unordered categories: Association models, correlation models, and asymmetry models for contingency tables without missing entries. *The Annals of Statistics*, *13*, 10–69.

Heinen, T. (1993). *Discrete latent variable models*. Tilburg, The Netherlands: Tilburg University Press.

Heinen, T. (1996). *Latent class and discrete latent trait models: Similarities and differences*. Thousand Oaks, CA: Sage Publications.

Hessen, D. J. (2012). Fitting and testing conditional multinormal partial credit models. *Psychometrika*, *77*, 693-709.

Holland, P. W. (1990). The Dutch Identity: A new tool for the study of item response models. *Psychometrika*, *55*, 5-18.

Joe, H., & Liu, Y. (1996). A model for multivariate binary response with covariates based on conditionally specified logistic regressions. *Statistics & Probability Letters*, *31*, 113–120.

Jöreskog, K.G., & Sörbom, D. (2004). LISREL 8.8 [computer software]. Chicago, IL: Scientific Software international, Inc.

Junker, B. W., & Sijtsma, K. (2000). Latent and manifest monotonicity in item response models. *Applied Psychological Measurement*, *24*, 65–81.

Kelderman, H., & Rijkes, C. P. M. (1994). Loglinear multidimensional IRT models for polytomously scored items. *Psychometrika, 59*, 149–176.

Lauritzen, S. L., & Wermuth, N. (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative. *Annals of Statistics, 17*, 31-57.

Li, Z. (2010). *Loglinear models as item response models*. Unpublished doctoral dissertation. University of Illinois, Urbana-Champaign.

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

McKinley, R. L., & Reckase, M. D. (1983). *An extension of the two-parameter logistic model to the multidimensional latent space* (Research Report ONR 83-2). Iowa City, IA: The American College Testing Program.

Molenberghs, G., & Verbeke, G. (2005). *Models for discrete longitudinal data*. New York, NY: Springer.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied psychological measurement*, *16*, 159-176.

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

Muthén, B.O., & Muthén, L.K. (2012). Mplus Version 7.0 [computer software]. Los Angeles, CA: Muthén and Muthén.

Ostini, R., & Nering, M. L. (2006). Polytomous item response theory models. Thousand Oaks, CA: Sage Publications.

Parzen, M., Lipsitz, S.R., Fitzmaurice, G.M., Ibrahim, J.G., Troxel, A., & Molenberghs, G. (2007). Pseudo-likelihood methods for the analysis of longitudinal binary data subject to nonignorable non-monotone missingness. *Journal of Data Science, 5*, 1-21.

Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement, 9*, 401–412.

Reckase, M. D. (2009). *Multidimensional item response theory.* New York, NY: Springer.

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

Rijmen, F., Tuerlinckx, F., De Boeck, F., & Kuppens, P. (2003). A nonlinear mixed model framework for item response theory. *Psychological Methods, 8*, 185–205.

Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores*. Iowa City, IA: Psychometric Society.

Sapiro, V., Rosenstone, S. J., & the National Election Studies. (1999). National Election Studies, 1998 Pilot Study [dataset]. Ann Arbor: University of Michigan, Center for Political Studies [producer and distributor].

Sheu, C.F., Chen, C.T., Su, Y.H., & Wang, W.C. (2005). Using SAS PROC NLMIXED to fit item response theory models. *Behavior Research Methods*, *37*, 202–218.

Smit, A., & Kelderman, H. (2000). Pseudolikelihood estimation of the Rasch model. *Journal of Outcome Measurement*, *4*, 513-523.

Strauss, D. (1992). The many faces of logistic regression. *The American Statistician*, *46*, 321-327.

Strauss, D.,& Ikeda, M. (1990). Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, *85*, 204-212.

Thissen, D., Chen, W.H., & Bock, R.D. (2003). Multilog (version 7) [Computer software]. Lincolnwood, IL: Scientific Software International.

Troxel, A. B., Lipsitz, S. R., and Harrington, D. P. (1998). Marginal models for the analysis of longitudinal measurements subject to nonignorable non-monotone missing data. *Biometrika, 85, 661-672*.

Vermunt, J. K. (1997). LEM 1.0: A General Program for the Analysis of Categorical Data [Computer Software], Tilburg, The Netherlands. Published online at http://members.home.nl/jeroenvermunt/.

Wang, W. C., Chen, P. H., & Cheng, Y. Y. (2004), Improving measurement precision of test batteries using multidimensional item response models. *Psychological Methods*, *9*, 116–136.

Wasserman, S., & Pattison, P (1996). Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p^*. *Psychometrika*, *61*, 401–425.

Wong, R. S. (2011). *Asymptotic, Jackknife, and Bootstrap Standard Errors for RC(M) Multidimensional Association Models*. Paper presented at the annual meeting of the American Sociological Association Annual Meeting, Las Vegas, NV.

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

Zhang, J., & Stout, W. (1997). On Holland's Dutch Identity conjecture. *Psychometrika, 62*, 375–392.

Zwinderman, A. H. (1995). Pairwise parameter estimation in Rasch models. *Applied Psychological Measurement*, *19,* 369-375.

Appendix. SAS Macros for PLE

```
/*******************************************************
  Basic Macro 1 : Data Conversion for PLE - Polytomous items
*******************************************************/

%macro Data_Poly_MDC(responses,  ItemTraitAdj, traitAdj, items, ncat = );

proc iml;

use &items;
read all var _char_ into varnames;
close &items;

use &responses;
read all var _all_ into personByitem;
close &responses;

use &traitAdj;
  read all into traitA;
close &traitAdj;

use &ItemTraitAdj;
  read all into ItemTraitA;
close &ItemTraitAdj;


/******************* Set-up  *****************************/

/* Basic information necessary for data conversion */

nitems = nrow(ItemTraitA);
npersons = nrow(personBYitem);
ncat = &ncat;
ntraits = nrow(traitA);
MaxNphi= (ntraits##2-ntraits)/2 + ntraits;


/* Index for person */

nrepeats = ncat*nitems;
personi = (1:npersons)`;
personhold=J(1,1,0);
do person = 1 to npersons;
  persontmp = REPEAT(personi[person],nrepeats);
  personhold = personhold // persontmp;
end;
ni = nrow(personhold);
personid = personhold[2:ni];

/* Index for case*/

nrepeats = ncat;
casei = (1:npersons*nitems)`;
casehold=J(1,1,0);
```

```
do case = 1 to npersons*nitems;
  casetmp = REPEAT(casei[case],nrepeats);
  casehold = casehold // casetmp;
end;
ni = nrow(casehold);
caseindex = casehold[2:ni];

/* Index for items */

itemi = 1:nitems;
itemhold = J(1,1,0);
do item = 1 to nitems;
  itemtmp= REPEAT(itemi[item],ncat);
  itemhold = itemhold // itemtmp;
end;
ni = nrow(itemhold);
itemtmp = itemhold[2:ni];
itemindex = REPEAT(itemtmp,npersons);

/* Index for categories */

cati = (1:ncat)`;
catindex = REPEAT(cati,nitems*npersons);
Nstack = nrow(catindex);


/* Creating effect codes for responses on each item */

do person =1 to Npersons;

ECodeR_1 = J(nitems*ncat, nitems*ncat, -1);
ij=1;

do item = 1 to nitems;
   do cat = 1 to (ncat);
     if cat = personBYitem[person, item] then ECodeR_1[, ij]=1;
       ij=ij+1;
   end;
   end;
   ECodeR=ECodeR//ECodeR_1;
   end;

/* Need some names for effect codes for responses in the data set */

Letter_E={"E"};
Letter_cat={"cat"};
ECRList = char(J(1,nitems*ncat,0));
clevels = char(1:ncat);

ij = 1;
do i=1 to nitems;
 do j=1 to (ncat);
  tmp = concat(Letter_E, varnames[1,i],Letter_cat,clevels[1,j]);
  ECRList[1,ij]=rowcatc(tmp,1,1);
  ij = ij+1;
 end;
```

```
end;

/* Creating effect codes for location indicators*/

blocki = I(ncat-1) ;
minusonerow = J(1,ncat-1,-1) ;
miblock=minusonerow // blocki ;

block1 = J(1, 1, 1) ;
block2 = J(ncat-1, 1, 0) ;
onecol=block1//block2 ;

unit = onecol ||miblock ;

iblocks =block(unit, unit);

do i=3 to nitems;
  iblocks = block(iblocks,unit);
end;

ELamda= REPEAT(iblocks,npersons,1);

/* Need some names for location parameters in the data set */

Letter_A={"Lamda"};
ItemNo = char(1:nitems);
CatLevels = Char(1:ncat);
Lamda_Names = char(J(1, nitems*ncat, 0));

p=1;

do i = 1 to nitems;
  do j = 1 to ncat;
    tmp = concat(Letter_A, ItemNo[1,i], CatLevels[1,j]);
    Lamda_Names[1,p] = rowcatc(tmp, 1, 1);
    p=p+1;
   end;
  end;

/* Creating variables for Lamda estimates on each item */

LamdaEst = J(NStack, nitems*ncat, 0);

/* Need some names for Lamda estimates in the data set */

Letter_Lamda={"NewLam"};
ItemNo = char(1:nitems);
CatLevels = Char(1:ncat);
Lamda_Estimates = char(J(1, nitems*ncat, 0));

p=1;

do i = 1 to nitems;
  do j = 1 to ncat;
    tmp = concat(Letter_Lamda, ItemNo[1,i], CatLevels[1,j]);
    Lamda_Estimates [1,p] = rowcatc(tmp, 1, 1);
```

```
    p=p+1;
   end;
  end;


  /* Response vector Y */

Y = J(NStack,1,0);
istack=1;

do person =1 to Npersons;
 do item = 1 to nitems;
   do cat = 1 to (ncat);
     if cat = personBYitem[person,item] then Y[istack,1]=1;
       istack=istack+1;
   end;
  end;
end;

/**** Setting starting value of phis ****/

/* Create item by correlated trait matrix and variable names */

itemCorTraitA = itemTraitA * TraitA - itemTraitA;

Letter_CorTrait={"CorTrait"};
CorTList = char(J(1,ntraits,0));
theta = char(1:ntraits);

do i=1 to ntraits;
   tmp = concat(Letter_CorTrait, theta[1, i]);
  CorTList[1,i]=rowcatc(tmp,1,1);
 end;

itemByphi=J(nitems, MaxNphi, 0);

do item = 1 to nitems;
  do j=1 to MaxNphi;
   iphi=1;
      do p=1 to ntraits;
        do q=p to ntraits;
          if q=p then do; itemBYphi[item,iphi]=itemTraitA[item,q]; end;
          else; do;
             if (itemTraitA[item,p] & itemCorTraitA[item,q]=1) then
itemBYphi[item,iphi]=traita[p,q]*0.5;
             else if  (itemtraita[item,p]=0 & itemCorTraitA[item,p]=1) then
itemBYphi[item,iphi]=traita[p,q]*0.5*itemCorTraitA[item,p]; end;
             iphi=iphi+1;
             end;
           end;
         end;
   end;


do i=1 to npersons;
do item=1 to nitems;
```

```
Phis_1=repeat(ItemBYphi[item ,], ncat, 1);
Phis = Phis//Phis_1;
end;
end;

/* Need some names for the phi parameters in the data set */

Letter_Phi={"Phi"};
TraitNo1 = char(1:ntraits);
TraitNo2 = char(1:ntraits);
Phis_Names = char(J(1, MaxNphi, 0));
i=1;

do p = 1 to ntraits;
   do q = p to ntraits;
    tmp = concat(Letter_Phi, TraitNo1[1,p], TraitNo2[1,q]);
    Phis_Names[1,i] = rowcatc(tmp, 1, 1 );
      i=i+1;
   end;
end;


/* Setting starting value of categories on each item */

/* ncat = 3 */

%if &ncat=3 %then %do;

do person =1 to Npersons;

Nu_1 = J(nitems*ncat, nitems, -1);
i=1;

do item = 1 to nitems;
  do cat = 2 to (ncat);
     if cat = personBYitem[person, item] then Nu_1[, i]= cat*0.2;
      end;
       i=i+1;
 end;
Nu = Nu//Nu_1;
end;
%end;

/* ncat = 4 */

%if &ncat=4 %then %do;

do person =1 to Npersons;

Nu_1 = J(nitems*ncat, nitems, -1);
i=1;

do item = 1 to nitems;
  do cat = 2 to (ncat);
     if cat = personBYitem[person, item] then Nu_1[, i]= cat*0.111;
      end;
```

```
      i=i+1;
 end;
Nu = Nu//Nu_1;
end;
%end;


/* ncat = 5 */

%if &ncat=5 %then %do;

do person =1 to Npersons;

Nu_1 = J(nitems*ncat, nitems, -1);
i=1;

do item = 1 to nitems;
  do cat = 2 to (ncat);
     if cat = personBYitem[person, item] then Nu_1[, i]= cat*0.1-0.1;
      end;
       i=i+1;
 end;
Nu = Nu//Nu_1;
end;

%end;


/* Need some names for Nu in the data set */

Letter_Nu = {"Nu"};
ItemNo = char(1:nitems);
Score_Nu = char(J(1, nitems, 0));

do i = 1 to nitems;
    tmp = concat(Letter_Nu, ItemNo[1,i]);
    Score_Nu[1,i] = rowcatc(tmp, 1, 1);
end;


/********* Concatenate All *******************/


/* Create data file for ItemCorTraitA */

create itemCorTraitA from itemCorTraitA [colname=CorTList];
append from itemCorTraitA;
close itemCorTraitA;

/*Create datafiles for master data file*/

create personid from personid [colname="personID"];
append from personid;
close personid ;

create caseindex from caseindex [colname="caseID"];
```

```
append from caseindex ;
close caseindex ;

create itemindex from itemindex [colname="itemID"];
append from itemindex;
close itemindex ;

create catindex from catindex [colname="catID"];
append from catindex;
close catindex ;

create Y from Y [colname="Y"];
append from Y ;
close Y ;

create ECodeR from ECodeR [colname=ECRList];
append from ECodeR;
close ECodeR ;

create ELamda from ELamda [colname=Lamda_Names];
append from ELamda;
close ELamda ;

create LamdaEst from LamdaEst [colname=Lamda_Estimates];
append from LamdaEst;
close LamdaEst ;

create Phis from Phis [colname=Phis_Names];
append from Phis;
close Phis ;

create Nu from Nu [colname=Score_Nu];
append from Nu;
close Nu ;

/*Create master datafile*/

data PolyMaster ;
merge personid caseindex itemindex catindex Y ECodeR ELamda LamdaEst Phis Nu
;
run ;


%mend Data_Poly_MDC;

/*******************************************************
  Basic Macro 2 : Computing Weighted Rest Scores
*****************************************************/


%macro ComputeWrest (responses, TraitAdj, ItemTraitAdj, ItemCorTraitA,
itemNum= );

proc sql;
create table Nu (drop = itemID)  as
select *
```

```
from &masterdata (keep= itemID Nu1-Nu&nitems)
where itemID=&itemNum;
quit;

proc sql;
create table Phis (drop = itemID)  as
select *
from &masterdata (keep= itemID phi: )
where itemID=&itemNum;
quit;

proc iml;

use Nu;
read all var _all_ into Nu ;
close Nu;

use Phis;
read all var _all_ into Phis ;
close Phis;

use responses;
read all var _all_ into personByitem;
close responses;

use TraitAdj;
  read all into TraitA;
close TraitAdj;

use ItemTraitAdj;
  read all into ItemTraitA;
close ItemTraitAdj;

use ItemCorTraitA;
  read all into ItemCorTraitA;
close ItemCorTraitA;

/* Basic information necessary for data conversion */

nitems = nrow(ItemTraitA);
npersons = nrow(personBYitem);
ncat = &ncat;
ntraits = nrow(traitA);
MaxNphi= (ntraits##2-ntraits)/2 + ntraits;
nstack=npersons*nitems*ncat;

/** Compute individual's weighted rest score **/

personBYnu=J(npersons, nitems, 0);

do p = 1 to npersons;
personBYnu[p ,]=nu[p*ncat ,];
end;

/* Rest score */
```

```
Nu_total = personBYnu*itemTraitA;

Rest_1 = J(1, MaxNphi,0);
item=&itemNum;

do i= 1 to npersons;
   iphi=1;
    do p=1 to ntraits;
        do q=p to ntraits;
          if p=q then do;
          Rest_1[1,iphi] =  (Nu_total[i,p]-personBYnu[i,
item])*itemTraitA[item,p]; end;
            else if q=p+1 then do;
            if (itemTraitA[item,p] & ItemCorTraitA[item,q]=1) then Rest_1[1,
iphi] = Nu_total[i,q]*traitA[p,q];
            else if (itemTraitA[item,p]=0 & ItemCorTraitA[item,p]=1) then
Rest_1[1, iphi] = (Nu_total[i,q-1]*traitA[p,q])*itemTraitA[item,q]; end;
          else if q=p+2 then do;
            if (itemTraitA[item,p] & ItemCorTraitA[item,q]=1) then Rest_1[1,
iphi] = Nu_total[i,q]*traitA[p,q];
            else if (itemTraitA[item,p]=0 & ItemCorTraitA[item,p]=1) then
Rest_1[1, iphi] = (Nu_total[i,q-2]*traitA[p,q])*itemTraitA[item,q]; end;
            else if q=p+3 then do;
            if (itemTraitA[item,p] & ItemCorTraitA[item,q]=1) then Rest_1[1,
iphi] = Nu_total[i,q]*traitA[p,q];
            else if (itemTraitA[item,p]=0 & ItemCorTraitA[item,p]=1) then
Rest_1[1, iphi] = (Nu_total[i,q-3]*traitA[p,q])*itemTraitA[item,q]; end;
          iphi=iphi+1;
           end;
        end;
         Rest=Rest//Rest_1;
    end;


/* Weighted rest score : Phi times Rest score */

itemBYphi=J(1, maxnphi, 0);
iphi=1;

do j=1 to maxnphi;
itemBYphi[1, iphi]=phis[1, j];
iphi=iphi+1;
end;

Wrest=J(npersons, 1, 0);

do i=1 to npersons;
    Wrest[i, 1]=itemBYphi[1 ,]*Rest[i ,]`;
end;

catBYitemRest=J(npersons*ncat, ncat, 0);
istack=1;

do i=1 to npersons;
   do j=1 to ncat;
     catBYitemRest[istack,j]=diag(wrest[i, 1]);
```

```
     istack=istack+1;
   end;
end;


istack=1;
do i=1 to npersons;
       catBYitemRest[istack, 2:ncat]=-1*catBYitemRest[istack, 1];
       istack=istack+ncat;
end;

Letter_Wrest={"Wrest"};
WrestList = char(J(1,ncat,0));
clevels = char(1:ncat);

do i=1 to ncat;
   tmp = concat(Letter_Wrest, clevels[1, i]);
  WrestList[1,i]=rowcatc(tmp,1,1);
 end;

 /* Create data file for weighted rest score */

create Wrest from catBYitemRest [colname=WrestList];
append from catByitemRest;
close Wrest;


/* Create dataset for Step 1 */

proc sql;
create table Itemdata as
select *
from &masterdata (keep= caseID itemID catID Y Lamda&itemNum: phi: Nu1-
Nu&nitems)
where itemID=&itemNum;
quit;

data itemdata ;
set itemdata ;
merge itemdata Wrest;
run;

%mend ComputeWrest ;

/***********************************************************************
  Basic Macro 3 : Step 1 - Conditional Multinomial Logistic Regression for
Each Item
***********************************************************************/

%macro Step1 (itemNum=, print= );


/****** Step 1 : Conditional Multinomial Logistic Regression for Each Item
********/

/* Conduct Proc MDC for each item */
```

```
proc mdc data=itemdata outest=itemest &print ;
model &response = Lamda&itemnum&cat2-Lamda&itemNum&ncat wrest2-wrest&ncat /
type=clogit  nchoice=&ncat covest=hess;
id &idnum;
run;

/* Create tempory data file for estimated Nu for each category */

data tmpNu ;
set itemest ;
keep wrest1- wrest&ncat;
wrest1=-1*sum(of wrest2 - wrest&ncat);
run;

/* Reorder variable names in tmpNU dataset : wrest1 comes first  */

data tmpNu;
retain wrest1 - wrest&ncat;
set tmpnu;
run;

/* Create dataset for estimated Nu of all items : This will be used in Step 2
*/


proc append base=tmpItemNu data=tmpNu;
run;


/* Dataset : Effect code for responses on each item */

proc sql;
create table EResponse as
select *
from &masterdata (keep= Ey&itemNum:) ;
quit;

/* Update Nu on master dataset */

proc iml;

/* Read datafile for basic information */

use responses;
read all var _all_ into personByitem;
close responses;

use TraitAdj;
  read all into TraitA;
close TraitAdj;

use ItemTraitAdj;
  read all into ItemTraitA;
close ItemTraitAdj;
```

```
use ItemCorTraitA;
  read all into ItemCorTraitA;
close ItemCorTraitA;


/* Basic information */

nitems = nrow(ItemTraitA);
npersons = nrow(personBYitem);
ncat = &ncat;
ntraits = nrow(traitA);
MaxNphi= (ntraits##2-ntraits)/2 + ntraits;
nstack=npersons*nitems*ncat;


/* Read information necessary for updating Nu */

use Eresponse;
read all var _all_ into ER ;
close Eresponse;

use tmpNu;
read all var _all_ into tmpNu ;
close tmpNu;

/* Update examinee's Nu score for item based on selected category */

istack=1;

do i = 1 to npersons;

New_NuScore_1 = J(nitems*ncat,1,0);

    do j = 1 to ncat;
    if ER[istack, j]=1 then New_NuScore_1[, 1] = repeat(tmpNu[1,j],
nitems*ncat, 1);
    end;

New_NuScore=New_NuScore//New_NuScore_1;
istack=istack+(nitems*ncat);

end;

/* Update Nu estimates for each category */

New_CatNu = J(nstack, ncat, 0);
istack = 1;

do i=1 to npersons;
do item=1 to nitems;
do j = 1 to ncat;
New_CatNu[istack, j] = diag(tmpNu[1,j]);
istack=istack+1;
end;
end;
end;
```

```
/* Create names for each nu estimate */

Letter_Nu={"NewNu"};
ItemNo = char(&itemNum);
CatLevels = Char(1:ncat);
Nu_Estimates = char(J(1, ncat, 0));

icat=1;
 do j = 1 to ncat;
    tmp = concat(Letter_Nu, ItemNo, CatLevels[1,j]);
    Nu_Estimates[1,icat] = rowcatc(tmp, 1, 1);
    icat=icat+1;
  end;


/* Concatenate new Nu score for each person and new Nu estimates for each
category */

All_New_Nu = New_NuScore || New_CatNu;
Nu_Name={"Nu&itemNum"};
colnames=Nu_Name || Nu_Estimates;

create All_New_Nu from All_New_Nu[colname=colnames];
append from All_New_Nu;
close All_New_Nu;


/* Put new Nu score and estimates in Master dataset */

data &masterdata;
set &masterdata;
merge All_New_Nu;
run;


 /* Up-date item history */

proc append base=history&itemNum data=itemest ;
run;

%mend Step1;


/*************************************************
  Basic Macro 4 : Step 2 - Stacked Logistic Regression
*************************************************/


%Macro Step2 (print= ) ;

/****** Step 2 : Stacked Conditional Multinomial Logistic Regression for All
Items ********/

proc sql;
create table NewNu as
```

```
select *
from &masterdata (keep= Nu1-Nu&nitems )
quit;

/** Compute Rest scores for each item and trait **/

/* Read basic matrices */

Proc iml;

use responses;
read all var _all_ into personBYitem;
close responses;

use TraitAdj;
  read all into TraitA;
close TraitAdj;

use ItemTraitAdj;
  read all into ItemTraitA;
close ItemTraitAdj;

use ItemCorTraitA;
  read all into ItemCorTraitA;
close ItemCorTraitA;

use tmpItemNu;
read all var _all_ into ItemNu ;
close tmpItemNu;

use NewNu;
read all var _all_ into NewNu;
close NewNu;

/* Basic information */

nitems = nrow(ItemTraitA);
npersons = nrow(personBYitem);
ncat = &ncat;
ntraits = nrow(traitA);
MaxNphi= (ntraits##2-ntraits)/2 + ntraits;
nstack=npersons*nitems*ncat;


/* Create new Nu socres by person for Step 2 */

personBYnewnu=J(npersons, nitems, 0);

do p = 1 to npersons;
personBYnewnu[p ,]=NewNu[p*nitems*ncat, 1:nitems];
end;

Nu_total = personBYnewnu*itemTraitA;

nphi = J(Nstack,MaxNphi, 0);
```

```
istack=1;
do i= 1 to npersons;
  do item= 1 to nitems;
     do j = 1 to ncat;
       iphi=1;
         do p=1 to ntraits;
            do q=p to ntraits;
             if p=q then do;
             nphi[istack,iphi] =  ItemNu[item,j]*(Nu_total[i,p]-
personBYnewnu[i,item])*itemTraitA[item,p]; end;
               else if q=p+1 then do;
                if (ItemTraitA[item,p] &  ItemCorTraitA[item,q]=1) then
nphi[istack,iphi] = ItemNu[item,j]*Nu_total[i,q]*traitA[p,q];
                else if (ItemTraitA[item,p]=0 &  ItemCorTraitA[item,p]=1)
then nphi[istack,iphi] = ItemNu[item,j]*(Nu_total[i,q-
1]*traitA[p,q])*ItemTraitA[item,q]; end;
             else if q=p+2 then do;
                if (ItemTraitA[item,p] &  ItemCorTraitA[item,q]=1) then
nphi[istack,iphi] = ItemNu[item,j]*Nu_total[i,q]*traitA[p,q];
                else if (ItemTraitA[item,p]=0 &  ItemCorTraitA[item,p]=1)
then nphi[istack,iphi] = ItemNu[item,j]*(Nu_total[i,q-
2]*traitA[p,q])*ItemTraitA[item,q]; end;
               else if q=p+3 then do;
                if (ItemTraitA[item,p] &  ItemCorTraitA[item,q]=1) then
nphi[istack,iphi] = ItemNu[item,j]*Nu_total[i,q]*traitA[p,q];
                else if (ItemTraitA[item,p]=0 &  ItemCorTraitA[item,p]=1)
then nphi[istack,iphi] = ItemNu[item,j]*(Nu_total[i,q-
3]*traitA[p,q])*ItemTraitA[item,q]; end;
             iphi=iphi+1;
               end;
          end;
          istack=istack+1;
       end;
     end;
  end;

/* Need some names for the nphi values in the data set */

Letter_nphi={"nphi"};
TraitNo1 = char(1:ntraits);
TraitNo2 = char(1:ntraits);
 nphis_Names = char(J(1, MaxNphi, 0));
 i=1;

do p = 1 to ntraits;
   do q = p to ntraits;
    tmp = concat(Letter_nphi, TraitNo1[1,p], TraitNo2[1,q]);
    nphis_Names[1,i] = rowcatc(tmp, 1, 1 );
      i=i+1;
   end;
end;

/* Create dataset for nphis */

create New_nphis from nphi[colname=nphis_Names];
append from nphi;
```

```
close New_nphis;


/* Put new nphi values in Master dataset  and create stacked dataset*/

data stack;
set &masterdata;
merge &masterdata New_nphis;
run;

/* Conduct Stacked Regression */

proc mdc data=stack outest=stackest &print ;
    model &response = &mymodel &allnphis / type=clogit nchoice=&ncat
covest=hess;
      id &idnum;

      restrict &scaling_constraint ;

run;


/*Temporary dataset for estimated phis*/

proc sql;
create table tmpPhi as
select *
from stackest (keep= nphi: );
quit;

/* The number of item by estimated phis */

data MorePhi;
set tmpPhi;
 do i=1 to &nitems;
 output;
 end;
drop i;
run;

/* Update estimated phis */

proc iml;

use responses;
read all var _all_ into personByitem;
close responses;

use TraitAdj;
  read all into TraitA;
close TraitAdj;

use ItemTraitAdj;
  read all into ItemTraitA;
close ItemTraitAdj;
```

```
use ItemCorTraitA;
  read all into ItemCorTraitA;
close ItemCorTraitA;

use tmpItemNu;
read all var _all_ into ItemNu ;
close tmpItemNu;

use MorePhi;
read all var _all_ into MorePhi ;
close MorePhi;


/* Basic information */

nitems = nrow(ItemTraitA);
npersons = nrow(personBYitem);
ncat = &ncat;
ntraits = nrow(traitA);
MaxNphi= (ntraits##2-ntraits)/2 + ntraits;
nstack=npersons*nitems*ncat;


*tmpOne=J(nitems, MaxNphi, 1);

itemBYallPhi=J(nitems, MaxNphi, 0);

do item = 1 to nitems;
  do j=1 to MaxNphi;
   iphi=1;
      do p=1 to ntraits;
        do q=p to ntraits;
          if q=p then do; itemBYallPhi[item,iphi]=ItemTraitA[item,q]; end;
          else; do;
             if (ItemTraitA[item,p] & itemCorTraitA[item,q]=1) then
itemBYallPhi[item,iphi]=TraitA[p,q];
             else if  (ItemTraitA[item,p]=0 & itemCorTraitA[item,p]=1) then
itemBYallPhi[item,iphi]=TraitA[p,q]*itemCorTraitA[item,p]; end;
             iphi=iphi+1;
             end;
           end;
         end;
   end;

/* Need some names for phi parameters */

Letter_Phi={"Phi"};
TraitNo1 = char(1:ntraits);
TraitNo2 = char(1:ntraits);
Phis_Names = char(J(1, MaxNphi, 0));
i=1;

do p = 1 to ntraits;
   do q = p to ntraits;
    tmp = concat(Letter_Phi, TraitNo1[1,p], TraitNo2[1,q]);
    Phis_Names[1,i] = rowcatc(tmp, 1, 1 );
```

```
      i=i+1;
   end;
end;

/*Create basic datasets necessary for updated phis*/

create EstPhi from MorePhi [colname=phis_names];
append from MorePhi;
close EstPhi;

create itemPhi from itemBYallPhi [colname=phis_names];
append from itemBYallPhi;
close itemPhi;

proc iml;

/* Read basic matrices */

use responses;
read all var _all_ into personByitem;
close responses;

use TraitAdj;
  read all into TraitA;
close TraitAdj;

use ItemTraitAdj;
  read all into ItemTraitA;
close ItemTraitAdj;

use ItemCorTraitA;
  read all into ItemCorTraitA;
close ItemCorTraitA;


/* Basic information */

nitems = nrow(ItemTraitA);
npersons = nrow(personBYitem);
ncat = &ncat;
ntraits = nrow(traitA);
MaxNphi= (ntraits##2-ntraits)/2 + ntraits;
nstack=npersons*nitems*ncat;

use EstPhi;
read all var _all_ into EstPhi;
close Estphi;

use itemPhi;
  read all var _all_ into itemPhi;
close itemPhi;

upPhi_1 = EstPhi#ItemPhi;

do i=1 to npersons;
    do item=1 to nitems;
```

```
      upPhi_2=repeat(upPhi_1[item ,], ncat, 1);
      upPhi = upPhi//upPhi_2;
     end;
end;


/* Need some names for phi parameters */

Letter_Phi={"Phi"};
TraitNo1 = char(1:ntraits);
TraitNo2 = char(1:ntraits);
Phis_Names = char(J(1, MaxNphi, 0));
i=1;

do p = 1 to ntraits;
   do q = p to ntraits;
    tmp = concat(Letter_Phi, TraitNo1[1,p], TraitNo2[1,q]);
    Phis_Names[1,i] = rowcatc(tmp, 1, 1 );
      i=i+1;
   end;
end;

create updatedPhi from upPhi [colname=phis_names];
append from upPhi;
close updatedPhi;


/* Update phi values in master dataset */

data &masterdata;
 merge &masterdata updatedPhi;
 run;

/* Up-date item history */

proc append base=stackhistory data=stackest ;
run;

/* Delete tmpItemNu for proper use in the following iterations */

proc datasets lib=work noprint ;
delete tmpItemNu;
quit;

%mend Step2;


/***********************
   Execution Macro
 ***********************/

%macro Execute_PLE  ;

%Data_Poly_MDC(responses,  ItemTraitAdj, traitAdj, items, ncat =&ncat );
```

```
/*******************************************************
  Uni : The main loop to iterate Step 1
 ********************************************************/

%if &step=1 %then %do;

%do i = 1 %to &iterations ;

     %do item=1 %to &nitems;

         %ComputeWrest (responses, TraitAdj, ItemTraitAdj, ItemCorTraitA,
itemNum=&item);

         %Step1 (itemNum=&item, print=noprint );

      %end;

   %end;

/*****************************************************
  Final estimates of Nu and Phi
 *****************************************************/

  %do item=1 %to &nitems;

   title "Item &item : Final Est. of Nu" ;

         %ComputeWrest (responses, TraitAdj, ItemTraitAdj, ItemCorTraitA,
itemNum=&item);

         %Step1 (itemNum=&item, print= );

   %end;

 %end;

/*******************************************************
  Multi : The main loop to iterate between Step 1 and Step 2
 ********************************************************/

%if &step=2 %then %do;

%do i = 1 %to &iterations ;

     %do item=1 %to &nitems;

         %ComputeWrest (responses, TraitAdj, ItemTraitAdj, ItemCorTraitA,
itemNum=&item);

         %Step1 (itemNum=&item, print=noprint );

      %end;

      %Step2 (print=noprint );

%end;
```

```
/*****************************************************
  Final estimates of Nu and Phi
 *****************************************************/

  %do item=1 %to &nitems;

   title "Item &item : Final Est. of Nu" ;

         %ComputeWrest (responses, TraitAdj, ItemTraitAdj, ItemCorTraitA,
itemNum=&item);

         %Step1 (itemNum=&item, print= );

   %end;

   title "Final estimates of Phi" ;

      %Step2 (print= ) ;

%end;

%mend Execute_PLE  ;
```

Table A. PLE Execution Macro: Descriptions and examples for global variables

Variable name	Description	Value (or Characer) & example	Note
masterdata	master dataset to be used during execution	'masterdata = polymaster' ; default	
cat2	Assigns '2' to 'Lamda&itemnum&cat2' in MDC procedure	'cat2 = 2' ; default	
response	response variable for MDC procedure	'response = y' ; default	
idnum	id variable for MDC procedure	'idnum = caseID' ; default	
step	'1' = Perform only Step 1 '2' = Perform full steps	'step = 1' for unidimensional models or 'step = 2' for multidimensional models	
nitems	Specifies the number of item in the model	'nitems = any number' (e.g., 'nitems = 6' for 6 items in the model)	
ncat	Specifies the number of response categories per item	'ncat = any number between 2 and 5' (e.g., 'ncat = 5' for 5 categories per item in the model)	
iterations	Specifies the number of iterations that users want to run	'iterations = any number' (e.g., 'iterations = 30')	
scaling_constraint	Impose scaling constraints by setting 'nphi##' to 1;The number of 'nphi##' depends on dimensionality.	'scaling_constraint = nphi11=1, nphi22=1, … nphi##=1' for #-dimensional models	Only for multidimensional models
allnphis	Specifies all possible association parameters; the number of 'allnphis' depends on dimensionality	For 2-dimensional models, 'allnphis = nphi11, nphi12, nphi22'. For 3-dimensional models, 'allnphis = nphi11, nphi12, nphi13, nphi22, nphi23, nphi33', and so on	Only for multidimensional models
mymodel	Specifies all location parameters to be estimated in the model; the number of location parameters depends on the number of items and categories per item	For six 3-category items, 'mymodel = lamda12 lamda13 lamda22 lamda23 … lamda62 lamda63'	Only for multidimensional models