A Speededness Item Response Model for Associating Ability and Speededness Parameters

By

Immanuel Williams

A dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Graduate School of Education

Written under the direction of

Youngsuk Suh

And approved by

New Brunswick, New Jersey

May 2017

ABSTRACT OF THE DISSERTATION

A Speededness Item Response Model for Associating Ability and Speededness Parameters

by Immanuel James Williams

Dissertation Director:

Youngsuk Suh

Test speededness is defined as the failure to attempt all items on an assessment within a specified time frame. The presence of speededness is an issue known to undermine assessments (Bejar, 1985). Therefore, researchers have developed several approaches to reduce test speededness, including non-statistical methods (e.g., Evans & Reilly, 1972), as well as probabilistic models, such as augmented item response theory (IRT) models (e.g., Cao & Stokes, 2008). However, an assumption about speededness that is often not discussed in the literature is the relationship between speededness and ability of the examinee in the context of IRT modeling. Previous studies have used modified IRT models to reduce test speededness, but none have evaluated the effect of neglecting the association between speededness and ability. In the same regard, only one model has addressed this association.

This study will address four different purposes regarding the relationship between speededness and ability. The first purpose is to propose a new IRT model that associates ability with speededness and to develop an estimation algorithm for the proposed model, which is evaluated in terms of the recovery of model parameters by manipulating certain hyperparameters in the model. The second purpose of this study is to determine the robustness of the proposed IRT model when speededness is not present, and to show the inefficiency of the traditional IRT model when speededness is associated with ability. The third purpose is to examine the impact of ignoring the association between ability and speededness on parameter estimation and to investigate the robustness of the proposed model under conditions when speededness and ability are independent. Lastly, data are generated from an existing model that associated ability and speededness in a different manner to determine how robust the proposed model is under a different speededness schema. These four purposes are used to thoroughly understand the proposed model and its contribution to the research of test speededness.

The Markov Chain Monte Carlo (MCMC) Metropolis Hastings algorithm was implemented to estimate model parameters using C++ and R, an object-oriented language and a statistical software, respectively. The results showed that the proposed model was able to recover the model parameters accurately under various conditions of known hyperparameters. The proposed model was also able to estimate model parameters well when ability was not associated with speededness when there were a large amount of respondents and items. In addition, the proposed model was also able to estimate model parameters well when speededness was not present when the sample size and the number of items were large. Lastly, when speededness and ability were generated under a different method, the proposed model was unable to estimate the model parameters well. In summary, this work allows researchers to further understand the impact of speededness and its association with ability in a variety of conditions.

Acknowledgments

This dissertation would not have been possible without the support of many people. I would like to express sincere gratitude to my adviser, Dr. Youngsuk Suh, for her continued support and guidance in my research. Thank you for being an advocate on my behalf and the support I needed during this journey. My sincere gratitude also goes to my committee members; Dr. Gregory Camilli, Dr. Chia-Yi Chiu, and Dr. William Strawderman. Thank you for your insightful feedback and guidance throughout the development of this dissertation.

I am extremely fortunate for the Graduate Assistance in Areas of National Need (GAANN) fellowship, I would not have been able to complete this dissertation without the financial support. My deepest appreciation to Dr. Jimmy de la Torre, thank you for pushing me to be a better researcher and always keeping me on my toes.

Words are not enough to express my thanks to Dr. Evelyn Erenrich and the Rutgers Graduate School – New Brunswick for their financial support and allowing me to mentor incoming students. I extend my gratitude to Tafia Simpson, your guidance solidified my graduate career. Thank you for continuing to connect me with colleagues and always being there to help. Since 2007 you have been an inspiring mentor providing academic, personal, and professional advice. In the same respect, the students that I have taught during my tenure at CSUMB have inspired to not only complete this but also move past strongholds to become a stronger individual.

Finally, this dissertation would not have been possible without God. All the people who I have crossed paths with throughout my academic career has led to this moment. This dissertation is dedicated to my wife and family for their unconditional love, support, and patience throughout this journey. They believed in me when I had my doubts and always offered words of encouragement, keeping me balanced with work and fun. Special thanks to my friends who often let me

vent especially during times where my model was not giving me expected results.

List of Tables

List of Figures

Contents

Chapter 1

INTRODUCTION

1.1 Definition of Test Speededness

Educational achievement metrics are vital to today's society in many ways. Assessments within educational programs are used to provide instructors with information about students' learning progression or lack thereof. Assessment results are often a core factor used to determine which students gain access to higher learning. These results also provide educators with information that can be used to design curricula, to determine efficacy of classroom methodology, and to evaluate the effects of policy on student learning. Given these applications, the accuracy of assessments are of the utmost importance. However, there are several confounding factors that may potentially undermine or bias the effectiveness of an assessment. A particular factor that commonly limits the accuracy of an assessment is the presence of test speededness. Test speededness is defined as the presence of time constraints on an assessment that prevents a group of examinees from completing a significant portion of the assessment (Bejar, 1985).

Time constraints are essential to the effective administration of an assessment. However, when these time constraints significantly affect test completion, they can undermine the purpose of measuring student ability and thus may negatively affect how test-takers are evaluated. It has been shown that speededness biases model parameter estimates in the context of IRT literature (e.g., Oshima, 1994; Chang, Tsai, & Hsu, 2014; Brown, Li, & Yang, 2013) and also affects the validity of a test (e.g., Lu & Sireci, 2007; Jin & Wang, 2014; Wise & Kingsbury, 2015). Though speededness is most commonly attributed to insufficient time allotment, there are other confounding effects that may prevent an examinee from finishing an assessment within specified time limits.

1.2 The Motivation of the Study

There are many theories as to why an examinee might become speeded during a test, depending on the specific context of an assessment. Time constraints are the most common factor that researchers use to explain test speededness, especially in the context of high-stakes assessments (Bolt, Cohen, & Wollack 2002; Yamamoto & Everson, 1997; Cao & Stokes, 2008; De Boeck, Cho, & Wilson, 2011; Wollack, Suh, & Bolt, 2007). The most common explanation is that examinees may spend too much time on items at the beginning of an assessment, leaving insufficient time to attempt the remaining items on the assessment. This behavior causes students to feel rushed towards the end of an assessment and prevents them from using optimal test-taking strategies, resulting in guessing and/or omitting items (Suh, Cho, & Wollack, 2012). To determine reasonable time constraints for an assessment, test designers take into account various factors, including the number of items on a test, the overall difficulty level of the test, and the test-taking population's level of ability. If these factors are not considered during the process of setting time constraints, the validity of an assessment may be questioned.

Additionally, recent theories suggest that a lack of motivation can inhibit student performance in the context of low-stakes assessments (Cao & Stokes, 2008; Jin & Wang, 2014; Wise & Kingsbury, 2015; Wang & Xu, 2015). Though lack of motivation and time constraints occur during different contexts of an assessment, they both elicit the same behavior of guessing and/or omitting responses. During a low-stakes assessment, this behavior often occurs due to test-takers' awareness of the purpose of the assessment (Cao & Stokes, 2008). Determining a confounding factor that may be related to lack of motivation and speededness due to time constraints may improve the understanding of and response to speededness.

Another perspective on these factors is that speededness can confound the effect of ability on the overall academic performance on an assessment. Researchers often overlook the possibility that the completion of an assessment within specified time constraints may depend on the examinees' ability. That is, examinees with high ability have a greater probability of finishing an exam in time compared to low ability examinees. Whereas, examinees with low ability have a greater probability of spending too much time on items at the beginning of an assessment, thus leaving insufficient time to consider all items compared to high ability examinees. This distinction presents the notion that not all individuals have the same probability of becoming speeded. This particular association is further deepened if the items are not ordered in increasing difficulty; examinees with low ability may spend excessive time on difficult questions received at the beginning of an assessment (Oshima, 1994). The relationship between ability and speededness was shown to have an impact on academic performance during an exam (Musch & Broder, 1999). An examinee with low ability may also experience poor motivation to complete an assessment, especially when the stakes themselves are low.

Given these three plausible and correlated factors driving test speededness (i.e., time, motivation, and ability level), it would be reasonable to state that someone with lower ability has a higher probability of becoming speeded compared to someone with high ability, especially when effected by time constraints and/or poor motivation. It is also important to note that the location on a test at which an examinee becomes speeded may also depend on ability. It is a reasonable assumption that a low-ability examinee will have a higher probability of becoming speeded earlier in the assessment compared to someone with greater ability. Currently, all but one model (Goegebeur, De Boeck, Wollack, & Cohen, 2008) assume an independent relationship between ability and speededness and there is no model that assumes that ability impacts the point at which an examinee will become speeded. Additionally, no studies have been conducted to examine the effect of associating speededness and ability on parameter estimation. A new speededness IRT model is therefore proposed and investigated in this dissertation.

1.3 The Objectives of the Study

There are many assumptions that are made in regards to how speededness affects the behavior of examinees. These assumptions are modeled when researchers attempt to evaluate the effect of speededness on an assessment. Some models assume that speededness occurs only on difficult items (Chang, Tsai, & Hsu, 2014; Cao & Stokes, 2008), whereas other models argue that speededness occurs with respect to latent groups (De Boeck, Cho, & Wilson 2011; Evans & Riley, 1971). However, existing models have not been able to explicitly explain speededness in relation to an examinees' ability levels, and thus have not examined the effect of modeling this association on parameter estimation. If this association is ignored, the estimation of model parameters may be negatively impacted. Therefore, it is critical to investigate these assumptions regarding speededness to improve upon these models.

This dissertation is designed to serve four purposes regarding the assumption of speededness and ability, and thus consists of four related studies. The purpose of the first study is to propose a model that associates ability and speededness, in which ability influences the location of speededness, and to implement an estimation procedure for this new model. This new model will be examined by manipulating certain hyperparameters to determine their impact on the estimation of model parameters. The purpose of the second study is to determine the robustness of the proposed model even when speededness is not present. This study is also used to show how traditional IRT models fall short when data are generated under the proposed model. The purpose of the third study is to discover if ignoring the association between ability and speededness within an existing speededness model impacts the estimation of model parameters, and to verify the robustness of the proposed model when speededness and ability are not associated. Lastly, the purpose of the fourth study is to determine the impacts of ignoring this association between speededness and ability and the robustness of the proposed model when a different speededness model is used to generate data.

The first study is necessary because current literature only describes one proposed model associating ability with speededness (Goegebeur et al., 2008). The relationship between ability and speededness in Goegebeur et al.'s model is not easily interpretable and they employ an estimation technique that is not typically used to estimate model parameters (Bolt, Cohen, & Wollack, 2001; Cao & Stokes, 2008). Therefore, it is necessary to derive a new model that associates ability and speededness in a manner that test developers and practitioners can understand as well as implement a tractable estimation method. In order to evaluate the feasibility of such a model, this study aims to answer the following question:

- How does the parameter recovery of the proposed model perform under various simulation conditions, including several known values of hyperparameters?

The second study is undertaken because, in order for a speededness model to be admissible, the model must be applicable even when speededness is not apparent within data. This is necessary in cases where one suspects the presence of speededness but, in fact, speededness might not exist in reality. In such an exploratory stage, the proposed model should estimate model parameters as accurately as traditional IRT models and find any speededness-related parameters to be minimal. Most speededness modeling studies consider this condition within their study (Bolt, Cohen, & Wollack, 2001; Cao & Stokes, 2008); however, they also incorporate the condition when speededness is present and illustrate how traditional IRT models fall short in the context of the estimation of model parameters. This second study is also important to demonstrate how ignoring the presence of speededness negatively impacts the estimation of model parameters. Therefore, the second study is structured to answer the following questions:

1. How well are the parameters recovered under traditional IRT model compared to the proposed model with respect to speeded conditions, in which ability and speededness are associated?

2. How well are the parameters recovered under the proposed model compared to traditional IRT model with respect to non-speeded conditions, (i.e. speededness is not present)?

The third study is critical to the research of speededness because if ability is associated with speededness within the data, it is unknown whether ignoring this information can lead to biased results or not. In other words, does taking into account this association improve the estimation of model parameters or have a non-significant effect? Although a speededness model that associates ability and speededness has been proposed in the literature (Goegebeur et al., 2008), the research question of that study did not address models that did not associate ability with speededness. Using the same rationale from the second study regarding the proposed model's admissibility, it is important to determine the robustness of the proposed model when speededness is present but not associated with ability. In sum, the third study is designed to answer the following questions:

1. What are the impacts of ignoring the association between test speededness and ability on the estimation of model parameters of a preexisting speededness model when data are generated from the proposed model?

2. How robust is the proposed model when speededness is present but not associated with ability?

The last study provides another way of determining the robustness of the proposed model and a preexisting speededness model in a different context of assuming a relationship between ability and speededness. This work adds to the body of research regarding speededness in that it investigates how the proposed model and a preexisting speededness model recovery model parameters when the association about ability and speededness is present but generated in a different manner.

1. How robust is the proposed model when speededness is associated with ability but generated under a different model?

2. What are the impacts of ignoring the association between ability and speededness when data are generated under a different model?

This dissertation will contribute to the literature by investigating whether ignoring the association between speededness and ability is negligible during the estimation of model parameters, or whether such an assumption needs to be included within the discussion of speededness. Although no model can perfectly describe speededness behavior, it is still important to clarify and validate reasonable assumptions that may arise in the presence of test speededness.

Chapter 2

LITERATURE REVIEW

2.1 Concept of Speededness

Gulliksen (1950) noted that there were two types of assessments: those that tested for power and others that measured speed. Power tests are designed to measure pure ability, whereas speed tests are constructed to measure ability with respect to speed (Lu & Sireci, 2007). Additionally, power tests are created with the intention that test-takers will fully attempt each item. Alternatively, speed tests are designed with the expectation that examinees will not be able to attempt each item within the specified time constraints. This dichotomy plays a vital role not only in the construction of an assessment, but also in its utilization. When designing a test, test developers have to make important decisions regarding this dichotomy. These decisions include but are not limited to the following aspects: time constraints, difficulty of items, number of items, and ordering of items. If these aspects are not given appropriate consideration, the line that distinguishes power and speed tests will become distorted.

When a power test has the characteristics of a speed test (i.e., a significant proportion of examinees do not finish the assessment), it is conceivable that the power test is confounded by speededness. Speededness is defined as time constraints preventing a group of examinees from completing a significant portion of an assessment (Bejar, 1985; Oshima, 1994; van der Linden, 2011). Speededness effects prevents examinees from fully attempting every item, which may induce random guessing or failure to make responses (Brown, Li, & Yang, 2013; Wang & Xu, 2015). The primary focus of this study is to evaluate the effects of test speededness on power tests, not speed tests. In this study, test speededness is also termed speededness effect or speededness.

It is important to note that time constraints are needed for any assessment, as tests cannot be administered with unlimited time. However, if the time constraints of an assessment are not properly defined (i.e., too little time to complete the assessment), a large number of examinees may not attempt some portions of the assessment, which can be strong evidence of the presence of speededness. In addition, if the number of items and /or the order of the difficulty of items are not taken into account, the speededness effect may worsen. For example, increasing the number of items while keeping the time constraints constant may cause severe test speededness. Speededness effects would also worsen if the difficulty of items were randomly ordered (Lawrence, 1993; Oshima, 1994). Additionally, some studies have shown that speededness can result from lack of motivation of examinees (Jin & Wang, 2014; Cao & Stokes, 2008).

2.2 Detecting Speededness

As a metric, speededness can be reported at different levels of severity. Understanding the precise level of test speededness provides vital insights into test evaluation, as certain levels can be negligible while others have a dramatic effect on the quality of an assessment. As a solution to this need, Gulliksen (1950) proposed a method that compared the variance of incorrect answers from two sources: the number of items given incorrect responses and the number of items given no response at all. The former source of incorrect answers is caused by lack of ability, whereas the latter source is a result of lack of time (Lu & Sireci, 2007). If a large amount of variance of incorrect answers is due to incorrect solutions, the test would be deemed unspeeded. On the other hand, if a significant amount of variance of incorrect answers is due to items not reached, the test would be considered speeded. Stafford (1971) introduced a similar technique to detect speededness. This strategy differed from Gulliksen's (1950) approach by directly comparing the number of incorrect items and the number of items not reached.

Swineford (1974) approached this problem by introducing a general heuristic to determine if the presence of speededness is significant. According to Swineford's rule, if more than 20% of the test-taking population does not complete an assessment (items left blank), or if each item is completed by less than 75% of the population, then the test displays a significant amount of speededness. This heuristic provides practitioners with a quick way to verify if speededness is problematic. The Swineford, Gulliksen, and Stafford methods were similar in that their techniques assume speededness to be represented solely by unanswered items, as opposed to alternative models that assume speededness to be represented by both unanswered items and random guessing. Additional probabilistic models, discussed in section 2.4, have also been developed in order to determine signficant speededness (Wise & DeMars, 2006; De Boeck, Cho, & Wilson, 2012; Bolt, Cohen, & Wollack, 2002).

2.3 Consequences of Speededness

Speededness negatively impacts many aspects of an exam, which creates cascading effects on the implications of an assessment. These ramifications are a key concern among researchers and practitioners, who strive to make meaningful conclusions regarding the assessment studied. The elements of an assessment that are affected by significant speededness are discussed below. Most of the elements discussed in this proposal are aligned in the context of IRT.

The use of item parameter estimates in measurement applications (e.g., adaptive testing and equating) is paramount to the administration, and to many implications, of an assessment. However, one of the critical effects of speededness on an assessment is how it alters the calibration of these item parameters (Brown, Li, & Yang, 2013; Oshima, 1994; Suh, Cho, & Wollack, 2012). Consequently, obscuring the ability to properly calibrate item parameters cascades into other applications such as adaptive testing and equating (Bridgeman & Cline, 2004; Kingston & Dorans, 1984; van der Linden, Breithaupt, Chuah, & Zhang, 2007; Wollack, Cohen, & Wells, 2003). If traditional IRT models are utilized for test evaluation and a large amount of speededness occurs, the calibration of items towards the end of a test tends to be biased (Bolt, Cohen, & Wollack, 2002; Yamamoto, 1995). Namely, items that appear towards the end of an assessment would end up more difficult than in non -speeded conditions. Therefore, the estimates of the difficulty parameters tend to be greater (i.e., more difficult) than the true parameters (Bolt, Cohen, & Wollack, 2002). Concurrently, the discrimination parameter estimates are biased to a certain degree (Brown, Li, & Yang, 2013). Once an item is affected by speededness, it loses its power to discriminate between low- and high-ability examinees. These confounding effects mean that speededness can have critical impacts on the proper development and analysis of assessments.

Speededness also has a significant effect on computerized adaptive testing (CAT) and multistage testing (MST) schemes, as estimated item parameters are used within the context of test banks to determine which item or sets of items will be administered during these types of assessments. (Bridgeman & Cline, 2004; Kingston & Dorans, 1984; van der Linden, Breithaupt, Chuah, & Zhang, 2007). If an examinee is performing well during a CAT or a MST, more difficult items should be correspondingly administered to the examinee. However, if the items within the test bank were calibrated under speeded conditions, the examinee may receive easier items, and thus his or her ability level will be inflated. Poor calibration of item parameters due to speededness also affects linking, equating, and scaling. These three processes use item parameter estimates to ensure the consistency of an assessment through multiple administrations. Research has shown that poorly calibrated item parameters due to speededness affects linking coefficients. Biased linking parameters cause the equating function to be skewed, which then causes the scaling between assessments to be distorted (Brown, Li, & Yang, 2013; Wollack, Cohen, & Wells, 2003).

Speededness also introduces biases in the estimation of ability levels in the context of IRT (Wollack, Suh, & Bolt 2007). When ability estimates are not accurate, other factors are affected, such as the local item independence assumption and the validity of the assessment (Yen, 1993). This bias occurs when speededness influences test-takers to change their test-taking strategy (Cao & Stokes, 2008; Jin & Wang, 2014). If examinees use an ability-based strategy during the first portion of an assessment and subsequently implement a guessing strategy when speededness occurs, the estimation of ability will not be inaccurate, because the response pattern during the beginning of an assessment changes undesirably towards the end of the assessment. This directly affects the validity of an assessment because the construct of interest is distorted by construct-irrelevant variance (i.e., speededness).

This construct-irrelevant variance creates a local item dependence structure, thus causing items to have a dependent association. Instead of items being locally independent with respect to ability, items are locally dependent due to a secondary stimulus, speededness. Local item dependence may lead to unreliable estimates of test reliability, item and test information, and standard error estimates (Lee, Kolen, Frisbie, & Ankenmann, 2001; Sireci, Thissen, & Wainer, 1991). In particular, the presence of speededness also causes researchers to question the validity and the reliability of an assessment, as the accuracy of measurement is disrupted by the presence of speededness (Lu & Sireci, 2007). Speededness specifically affects a test's precision (Gullkisen 1950), since reliability tends to be magnified when speededness is present. When an assessment's validity and reliability are called into question, test scores and conclusions drawn from the test scores may be rendered invalid (Lu & Sireci, 2007).

2.4 Remediating Speededness Effects

Due to the negative impacts of test speededness, researchers have derived methods for removing these effects from assessments. Some researchers have developed non-statistical approaches to accomplish this task, while others have proposed probabilistic models to capture the presence of speededness. One of the primary goals of this dissertation is to evaluate the effects of ignoring an assumption about the association between speededness and ability. Therefore, a few non-statistical techniques and probabilistic models that are not directly related to this dissertation are briefly discussed. Subsequently, a more thorough discussion regarding essential probabilistic models is offered, which is foundational to the model proposed in this dissertation. These models are inherently designed to reduce speededness impact on an assessment based on the researchers' understanding of speededness.

Many non-statistical techniques have been presented to counter speededness effects, but each approach has its limitations. One method is to increase test administration time (Lawrence, 1993). Although this is a legitimate approach, additional time may not always lead to improved performance (Evans & Reilly, 1972). Another technique is to remove items that are located towards the end of an assessment, assuming that they are the source of speededness. (Oshima, 1994; Lord, 1980). The main drawback to this approach is the fact that it is nearly impossible to determine the inception of speededness behavior that is applicable to all examinees. Another issue that arises with removing items is that both the content and construct validity of an assessment decrease when items are removed. Lastly, Bejar (1985) attempted to determine when a person becomes speeded, based on the difficulty of items and the performance of the examinee on the assessment, in order to remove all items after this point. This approach is difficult to implement because the performance of the student may not be indicative of his or her true ability if he or she becomes speeded earlier on an assessment. That being said, it is critical to consider other techniques to reduce the presence of speededness in assessments.

As an alternative to the aforementioned non-statistical methods, researchers have developed several probabilistic models to reduce the effects of speededness on item calibration. Unlike non-statistical methods, probabilistic models are not contrived based on convenience but rather on certain assumptions regarding speededness. These models generally assume that speededness causes examinees to guess on items and fail to answer items. (Bolt, Cohen, & Wollack 2002; Yamamoto & Everson, 1997; Cao & Stokes, 2008; De Boeck, Cho, & Wilson, 2011; Wollack, Suh, & Bolt, 2007).

One such model seeks to reduce test speededness on computer-based assessments with respect to response time on items (van der Linden, 2007; van der Linden, 2011). Research by van der Linden (2006) proposed the lognormal model, which can be used to measure the response time it takes a test-taking population to respond to each item on an assessment. This metric is then used to model the total amount of time that a test-taking population requires to complete an entire assessment. The lognormal model is extremely convenient for adaptive tests because the algorithm can select items not only based on the difficulty and content but also on the time expected to complete the item, which may reduce the probability of speededness.

Another model created for computer-based assessments is the effort-moderated model proposed by Wise and DeMars (2006). In this model, if an examinee answers an item within a time frame below a certain threshold (i.e., an unusually short amount of time), it is assumed that the test-taker is guessing rather than responding based on his/her ability. This specified time is noted as an item threshold. However, the effort-moderated model can itself be biased if a test-taking population is composed of multiple subgroups that differ in average ability levels and test-taking strategies. In this case, the threshold parameter will not be invariant, meaning that the expected time it takes one subgroup to answer a question may differ drastically from another. As a result, the threshold used to determine if an examinee is speeded will be inaccurate.

Although assessments are increasingly being administered in a computer-based format, a significant number still use a paper-and-pencil format (Patelis, 2000). This removes the ability to measure time spent on each item, meaning the response time and lognormal models will be inadequate in these contexts. The current models for paper-based tests assume that examinees respond to speededness via various types of guessing mechanisms. Cao and Stokes (2008) proposed models that represent these different guessing approaches. One of their models is the continuous guessing model used for both minimally and highly motivated students. Under the continuous guessing model, guessing behavior is modeled so that the probability of correctness is fixed to a constant and the potential to become speeded occurs randomly throughout an assessment. The second guessing model is based on an assumption that some examinees guess on hard questions but attempt easy questions based on their ability levels. This model was extended by Chang, Tsai, and Hsu (2014), who assumed that students leave the more difficult items until the end of the test while attempting easier items based on their ability. However, these guessing models are limited in that they assume examinees do not omit responses as a result of speededness.

These various non-statistical and probabilistic models are essential to the study of speededness and its reduction. There are some additional models that this paper will examine more in depth, as they are vital to understanding the proposed model and are more robust to handle paper-and-pencil tests and missing responses. In order to model the following assumptions – 1) speededness and ability are associated, and 2) the inception of speededness depends on ability – one must understand the basis of Jin and Wang's (2014) two-parameter logistic mixture (2PLMix) model and the assumptions made under the gradual process change (GPC) model introduced by Goegebeur et al. (2008). The 2PLMix model is an extension of the mixture Rasch model (MRM) (Bolt, Cohen, & Wollack, 2002) and HYBRID model (Yamamoto & Everson, 1997). The 2PLMix model also uses some of the key assumptions found in the IRT-threshold guessing model (Cao and Stokes, 2008). Therefore, an understanding of the models used to derive the 2PLMix is crucial in understanding the mechanism used to model the two new assumptions.

These four models (MRM, HYBRID, IRT-threshold guessing, and 2PLMix) are considered to be mixture models, which means two aspects (test taking strategy and guessing/non-response) are used to determine the probability of correctness based on a latent group. It is also critical to understand the GPC model because this model formulates an association between speededness and ability level (although this association is derived in an esoteric manner), which suggests there are ramifications for ignoring this association. The following will include a description of each model and a summary of results found in each paper, followed by a comparison of the models relating to the 2PLMix.

2.4.1 Gradual Process Change (GPC) Model

The general structure of the GPC model assumes speededness occurs after an examinee's threshold has been surpassed, η_i (see the description below) and gradually impacts the probability of correctness for the remaining items. The GPC model can be seen as:

$$P_{ij} = P(X_{ij} = 1|\alpha_j, \beta_j, \gamma_j, \theta_i)\min\left\{1, \left[1 - \left(\frac{j}{J} - \eta_i\right)\right]^{\lambda_i}\right\} \tag{2.1}$$

where j represents j^{th} item, J is the total number of items on the assessment, η_i denotes the proportion of nonspeeded items answered by examinee i, λ_i is the intensity factor, and X_{ij} denotes the correctness of the response (1 for correct and 0 for incorrect). The probability of a correct response can be seen as the 3-parameter logistic (3PL) expressed as:

$$P(X_{ij} = 1|\alpha_j, \beta_j, \gamma_j, \theta_i) = \gamma_j + (1 - \gamma_j)\frac{\exp(\alpha_j(\theta_i - \beta_j))}{1 + \exp(\alpha_j(\theta_i - \beta_j))} \tag{2.2}$$

where α_j denotes the discrimination parameter for item j, β_j denotes the difficulty parameter for the j^{th} item, γ_j is the guessing parameter, and θ_i is the ability parameter for the i^{th} examinee. This model shows that normal test-taking behavior occurs on and before the η_i proportion of the test for examinee i; however, after the η_i proportion of the assessment has been completed, the probability of obtaining the correct solution begins to diminish. Figure 2.1 illustrates the functioning of Equation 2.1 for a 40 item test, in which an examinee has a threshold of $\eta_i = 0.5$ and intensity factor of $\lambda_i = 4$.

This figure shows the speededness effect that is multiplied by the probability of a correct answer. If the examinee is speeded, then the probability of a correct solution is multiplied by a proportion that is between 0 and 1. If the examinee is not speeded, then the probability of a correct solution is multiplied by 1. Thus, the test becomes progressively more difficult for an examinee to get a correct solution once he or she becomes speeded. The association between ability, θ_i, and speededness variables, η_i and λ_i, is established by using a copula function. The implementation of a copula function is a nonlinear technique used to create an association among independent random variables.

The simulation study in Goegebeur et al. (2008) compared the performance of the 3PL model vs the GPC using four different levels of speededness. The results showed the model parameters were properly estimated under the GPC and outperformed the 3PL in the context of parameter recovery. However, the estimation technique implemented was a version of the SAS NLMIXED procedure, which is not widely used in this line of research, and the method used to associate ability and speededness is not intuitive to practitioners. Therefore, it is ideal to propose a model that addresses these issues within the GPC paradigm. Although the following described models do not address the assumption regarding the relationship between ability and speededness, they are foundational in describing the model

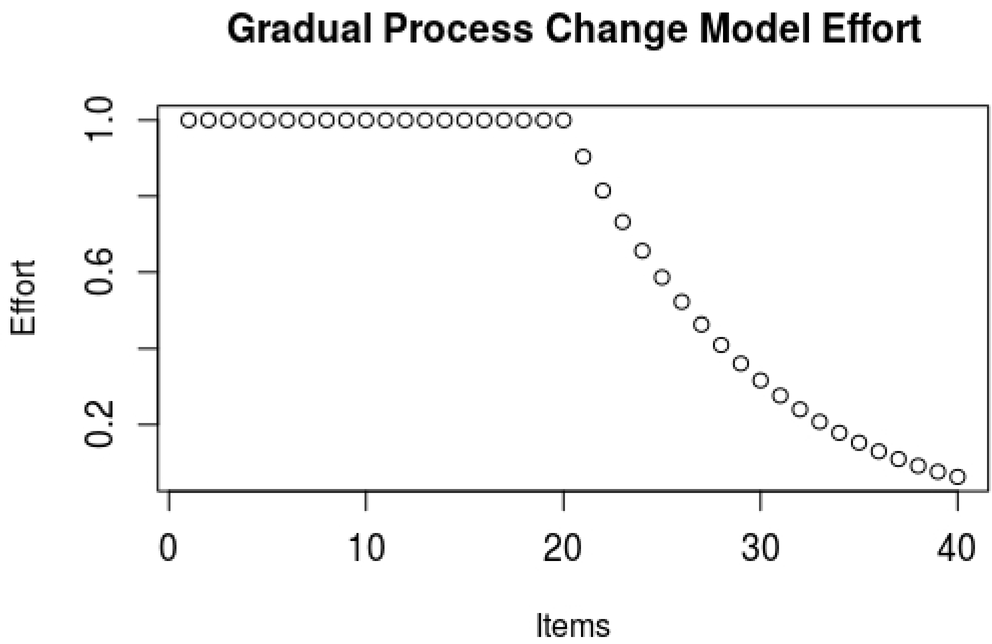

Figure 2.1: Illustration of the Gradual Process Change model where $J = 40$, $\lambda_i = 4$, and $\eta_i = 0.5$. which depicts effort only.

proposed in this dissertation.

2.4.2 Mixture Rasch Model (MRM)

The MRM was first introduced by Rost (1990) to model latent groups within the context of IRT. Bolt et al.'s (2002) study of the MRM was implemented in the context of speededness. The latent groups were specified by one group being nonspeeded and the remaining groups all being speeded. Each latent group was modeled under a Rasch model, with the difficulty levels of speeded items being different among the groups. Finally, the ability distributions between latent groups were set to be non-overlapping. These limitations ensure that a person's response is based on the group's specific difficulty level and ability distribution. The probability of a correct response based on these assumptions can be expressed as:

$$P(X_{ij} = 1|\beta_{jg}, \theta_{ig}) = \frac{\exp(\theta_{ig} - \beta_{jg})}{1 + \exp(\theta_{ig} - \beta_{jg})} \tag{2.3}$$

where g is the number of groups ranging from $1, ..., G$, β_{jg} is the difficulty parameter for item j in group g, and θ_{ig} is i^{th} examinee's ability in group g. Markov Chain Monte Carlo was used to estimate the MRM parameters. The probability of being in group g is denoted as π_g where $\sum_g \pi_g = 1$. Although multiple latent groups are possible in the MRM, Bolt et al. (2002) investigated only two latent groups, a speeded class and a nonspeeded class. The researchers assumed that the nonspeeded group's difficulty parameter was always smaller than the speeded group's difficulty parameter for the speeded items. This constraint ensured that speededness always resulted in an inflated difficulty parameter. The results showed that difficulty parameters were estimated well in the nonspeeded group.

However, by imposing only two latent classes, the MRM misses the effect of various levels of speededness occurring at different points of an assessment. Using data solely from nonspeeded items (i.e., ignoring the speeded items) can be seen as controversial in the context of ensuring a valid assessment that uses all the information in parameter estimation. Wollack et al.'s (2003) study went further and used only the difficulty parameter estimates from the nonspeeded group within the context of equating, which is even more controversial. Due to these limitations of the MRM, there is need for a model that includes more than two latent groups.

2.4.3 HYBRID Model

An alternative model used to measure the effect of various levels of speededness is the HYBRID model by Yamamoto and Everson (1997). The HYBRID model assumes a switching behavior in the test-takers. Initially, examinees are assumed to employ a normal test-taking strategy. However, according to the model, some examinees will switch the normal strategy to a guessing strategy partway through the test and continue this strategy until completion. A group of examinees that makes this switch on the same item forms one latent class. Thus, different latent classes will portray a different number of consecutively guessed items at the end of the test. This model assumes that normal test-taking strategy occurs during nonspeeded conditions of the assessment, while a switch to a guessing strategy indicates that the examinee has become speeded. The HYBRID model can be expressed as:

$$P(X_{ij} = 1|\alpha_j, \beta_j, c_j, \theta_i, g_m) = \big(1 + \exp(-\alpha_j(\theta_i - \beta_j))\big)^{g_m} c_j^{g_m+1} \tag{2.4}$$

where α_j, β_j, and θ_i are the discrimination, difficulty, and ability, respectively and c_j is the probability of examinees in latent class g randomly guessing the correct answer to item j. Additionally, g_m is a binary variable that indicates when a test is taken under nonspeeded ($g_m = -1$) or speeded ($g_m = 0$) conditions. A marginalized maximum likelihood estimation approach was used to by the researchers to estimate model parameters. The simulation design of Yamamoto and Everson (1997) investigated a 70-item assessment in which there were 20 different levels of speededness effects. These levels were arbitrarily defined based on what the researchers deemed to be the inception of speededness. The HYBRID model was able to recover its model parameters well, especially the ability and difficulty parameters, during the speeded sections of the test. The model was also able to accurately determine the location at which speededness started.

Like the MRM approach, the HYBRID model is based on a mixture model, but accounts for multiple levels of speededness via various speeded latent classes. The HYBRID model also differs from the MRM in that each latent class does not obtain a specific difficulty parameter but rather a random response pattern. Although the HYBRID model includes more than two latent classes and has the ability to estimate more latent classes, the simulation study only investigated speededness occurring at 20 different locations. The number of classes can be easily increased, but the estimation time is heavily affected by the number of classes. Also, the model assumes each latent group has the same response pattern, which is not realistic. Therefore, a model that can overcome these limitations needs to be introduced.

2.4.4 Item Response Theory-Threshold Guessing (IRT-TG)

The IRT threshold guessing (IRT-TG) model presented by Cao and Stokes (2008) differs from the previously discussed models by 1) allowing all items as locations for the inception of speededness within the simulation study and 2) implementing the possibility of random guessing for each item. The location of speededness within a test is not based on belonging to a latent class but rather is specific to each individual test-taker. The IRT-TG assumes an examinee from a subpopulation of test-takers uses his or her knowledge to answer questions up to a given threshold, and then guesses on the remaining items. The remaining population responds normally until the end of the assessment. The IRT-TG is expressed in the following equation:

$$P(X_{ij} = 1|\alpha_j, \beta_j, c_j, \theta_i, \delta_i) = \begin{cases} \frac{\exp(\alpha_j(\theta_i - \beta_j))}{1+\exp(\alpha_j(\theta_i-\beta_j))} & j \leq \delta_i \\ \frac{\exp(c_j)}{1+\exp(c_j)} & j > \delta_i \end{cases} \tag{2.5}$$

where α_j, β_j, and θ_i are the discrimination, difficulty, and ability, respectively, δ_i is the last item at which examinee i uses normal ability, and c_j is a guessing parameter. Markov Chain Monte Carlo was used to estimate IRT-TG model parameters. It is important to note that the guessing parameter corresponds to the item and not the person. Similarly to the HYBRID and MRM approaches, the IRT-TG assumes J (the total number of items on an assessment) locations at which speededness begins. Moreover, the IRT-TG uses a stochastic method to determine the location, δ_i, at which speededness behavior begins. The distribution used to determine the probability of being speeded on the j^{th} item can be expressed as:

$$\pi_j = \frac{j^\omega - (j-1)^\omega}{(J-1)^\omega}(1 - \pi_J) \tag{2.6}$$

where ω is a hyperparameter and π_J is the probability of not being speeded. The probability distribution for an examinee becoming speeded on a test can be seen in vector form as $[\pi_1, \pi_2, \ldots, \pi_J]$. Another way of understanding this distribution is to state that $\delta_i \sim [\pi_1, \pi_2, \ldots, \pi_J]$ It is important to note that the probability of an examinee being speeded on a given item is determined by $(1-\pi_J)$, and thus the probability of being speeded on an assessment is distributed among the $(1, \ldots, J-1)$ items. Consequently, an examinee not being speeded on an assessment is represented by becoming speeded on the J^{th} item (i.e., the last item). The results of Cao and Stokes (2008) showed that the IRT-TG was able to recover the model parameters well in both conditions of the presence and absence of speededness.

Although the HYBRID and IRT-TG models are similar in that they both implement a switching strategy approach, they differ with respect to how the random responses are generated. The IRT-TG incorporates a guessing parameter, c_j, to generate a random response, whereas the HYBRID model's random response is based on a latent class membership. Finally, the IRT-TG model does a better job representing real-world situations by modelling guessing behavior on an individual item basis rather than by latent groups. However, this guessing behavior is only impacted by the item itself, which may not always be the case. Guessing behavior may depend on the location at which speededness behavior begins, which is affected by both the item and the specific test-taker. This shortfall is addressed by the next model.

2.4.5 Two Parameter Logistic Mixture Model (2PLMix)

Assumptions found in the MRM, HYBRID, and IRT-TG models were used by Jin & Wang (2014) to develop the 2PLMix model. The 2PLMix implements each item as a possible location for the inception of speededness similarly to the previous two models. The method used to determine the probability of becoming speeded is similar to the one discussed in the IRT-TG model. However, the 2PLMix model additionally proposes that the logit of the IRT model is impacted by a decrement factor once speededness occurs, meaning that the probability of correct solution decreases by a calculated factor rather than being determined by a general guessing factor. The 2PLMix can be expressed as:

$$P(X_{ij} = 1|\alpha_j, \beta_j, \theta_i, \eta_{ij}) = \frac{\exp(\alpha_j(\theta_i - \beta_j - \eta_{ij}))}{1 + \exp(\alpha_j(\theta_i - \beta_j - \eta_{ij}))} \tag{2.7}$$

where α_j, β_j, and θ_i, are the discrimination, difficulty, and ability, respectively, and η_{ij} is the decrement parameter. The decrement parameter is determined by the following function:

$$\eta_{ij} = \begin{cases} 0 & j \leq \delta_i \\ \gamma_{\delta_i} & j > \delta_i \end{cases} \tag{2.8}$$

where δ_i is the item at which examinee i becomes speeded and γ_{δ_i} is the speededness effect. The value of $n_{ij} = 0$ when the examinee is not under speeded conditions (that is, when $j \leq \delta_i$). However, n_{ij} resolves to γ_{δ_i} when $j > \delta_i$. The equation used to determine the speededness effect can be expressed as:

$$\gamma_{\delta_i} = \kappa(J - \delta_i) \tag{2.9}$$

where κ is a parameter used to moderate the speededness impact. This speededness effect (γ_{δ_i}) is item and person specific, meaning that examinee i becomes speeded at item j. Once an examinee becomes speeded, the decrement parameter (η_{ij}) has a constant effect on the probability of obtaining the correct solution on the remaining items. Equation 2.9 also shows that the earlier an examinee becomes speeded on a test, the larger the speededness effect will be and vice versa. An illustration of this concept will be demonstrated later in section 3. Besides the 2PL version of the model, Jin and Wang (2014) also proposed the 1PL, 3PL, and graded-response mixture models in their study.

The 2PLMix assumes latent classes of varying levels of speededness, similarly to the HYBRID and IRT-TG models previously discussed. However, the 2PLMix adopts and implements multiple levels of speededness based on the number of items, much like the IRT-TG model within the simulation study. In other words, each item on the assessment represents a location at which an examinee may become speeded. The 2PLMix also implements Equation 2.6 to determine the location, δ_i, at which an examinee will become speeded, but with a slight modification. It is important to note that the 2PLMix model parameters were estimated with Markov Chain Monte Carlo simulations. The 2PLMix assumes that the exponents must always be greater than 1 to ensure the concavity of the probability mass function. Instead of assuming ω to be always greater than 1, the 2PLMix model increases the exponent by 1. The modified equation is denoted as:

$$\pi_j = \frac{j^{\omega+1} - (j-1)^{\omega+1}}{(J-1)^{\omega+1}}(1 - \pi_J) \tag{2.10}$$

which represents the probability of being speeded on the j^{th} item. Again, the probability of being speeded is $(1-\pi_J)$, which is distributed amongst the $1, \ldots J-1$ items. Equation 2.10 is a non-decreasing probability function which implies the probability of becoming speeded increases as an examinee completes the test. Equation 2.10 also implies that the location at which examinee i becomes speeded, δ_i, follows a distribution denoted in vector form as $[\pi_1, \pi_2, \ldots, \pi_J]$. If $\delta_i = J$, examinee i did not become speeded during the assessment. The mechanism of this probability mass function is further discussed in an example in section 3.

The model used within Jin and Wang's (2014) simulation study was the 3PL-GPCMix model, which included the 3PL and the general partial credit models. The item parameters were properly estimated with respect to the parameter recovery. However, ability, θ_i, was briefly discussed, and δ_i, the parameter of the location at which speededness occurs, was not mentioned in terms of parameter recovery. Though the models have progressed in terms of accounting for more realistic assumptions, some general notions can be found in all four models.

Jin and Wang (2014) highlighted four assumptions about the four mixture models (MRM, HYBRID, IRT-TG, and 2PLMix) that are important to understand the general nature of these models and the model presented in this dissertation. The first assumption is that examinees answer items based on the original order of the assessment. This means that items cannot be skipped and returned to. Second, a proportion of examinees use their ability up to a certain point and then began to implement a guessing strategy, while the remaining examinees fully attempt the entire test. The third assumption is that different examinees become speeded at different locations. This assumption may not seem to hold for the MRM and the HYBRID, but these models assume that real world applications will see a greater number of latent groups than those presented within their respective simulation studies. Lastly, these models assume that once an examinee becomes speeded, the examinee will remain speeded until the end of the assessment.

Though these models share some common characteristics, there are a few assumptions that inherently make these models different. All the mixture models discussed so far (except the HYBRID) account for the impact of speededness within the logit; however, each model implements this slightly differently. The MRM assumes that the difficulty parameter is different between the two groups. This is not a favorable assumption because it reflects that speededness only impacts item difficulty and not ability level. The IRT-TG model assumes only the guessing parameter within the logit when speededness is present. Since this model removes the item and ability parameters from the logit, its mechanism may not allow the guessing parameter to accurately account for speededness and its effect on the other model parameters. The implementation of latent classes within the HYBRID model implies that examinees within each latent class has the same response pattern during the speeded portion of a test, which is not realistic. The 2PLMix model assumes that a parameter representing both the item and the person reflects the impact of speededness. This is more realistic because if speededness occurs, then this parameter will account for the bias that occurs according to the item and the examinee. The proposed model takes several features from the discussed models and attempts to reflect the effect of speededness while resolving the various limitations of these approaches.

Chapter 3

Methodology

3.1 Overview

The models highlighted in the literature review lay the foundation for the development of the model introduced in this dissertation, which correlates ability and speededness. Based on previous research, there is evidence of a potential relationship between ability and speededness. Ignoring such a relationship in the data may have many negative ramifications and has not been thoroughly evaluated in the literature. Therefore, it is important to determine what happens to quality of the estimation process when the association between ability and speededness is ignored.

In section 3.2, the description of a new speededness model and relevant parameters will be discussed, followed by a comparison between the new model and the 2PLMix in section 3.3. Next, a comparison between the GPC and the new model will be provided in section 3.4. An overview will then be given regarding Markov Chain Monte Carlo (MCMC) in section 3.5. Lastly, section 3.6 will offer a discussion regarding the implementation of MCMC to the estimation of model parameters of the models presented within the simulation study.

3.2 A Modified 2PLMix

The model introduced in this section, denoted as the M2PLMix model, is a modification of the 2PLMix model. This modification assumes that the probability of being speeded depends on ability. The M2PLMix model's probability function of getting an item correct is identical to the 2PLMix model probability function denoted as follows:

$$P(X_{ij} = 1|\alpha_j, \beta_j, \theta_i, \eta_{ij}) = \frac{\exp(\alpha_j(\theta_i - \beta_j - \eta_{ij}))}{1 + \exp(\alpha_j(\theta_i - \beta_j - \eta_{ij}))} \tag{3.1}$$

where α_j is the discrimination parameter, β_j is the difficulty parameter, θ_i is the ability parameter, and η_{ij} is the decrement parameter. Similarly to the 2PLMix model, the decrement parameter can be represented as:

$$\eta_{ij} = \begin{cases} 0 & j \leq \delta_i \\ \gamma_{\delta_i} & j > \delta_i \end{cases} \tag{3.2}$$

and the γ_{δ_i}, the speededness effect parameter, is expressed by the following equation:

$$\gamma_{\delta_i} = \kappa(J - \delta_i) \tag{3.3}$$

It is important to reiterate that Equations 3.1-3.3 are identical to Equations 2.7-2.9, which are used to determine the impact of speededness on the probability of obtaining a correct answer and thus have the same meaning. The only difference between these two models is that in the M2PLMix, the probability of not being speeded depends on the ability level, whereas in the 2PLMix model the probability of not being speeded is fixed across all examinees regardless of ability. Explicitly, the location of speededness, δ_i, directly depends on the ability level, θ_i, for the M2PLMix. It should be also noted that regardless of the association between speededness and ability, if the probability of not being speeded is known, the probability of being speeded is also known, (i.e., $P(S) = 1 - P(S)$, where P(S) represents the portability of being speeded).

The probability of not being speeded within the 2PLMix is denoted by the parameter, π_J, which is the same value for the entire test-taking population. This assumption not only makes the parameter not applicable under all testing contexts (i.e., low- and high-stakes assessments), but also makes the use of this parameter unrealistic. The probability of not being speeded on a test in the 2PLMix is a property of the test and not a characteristic of the test-taking population. However, within the M2PLMix model, a probability function is used to describe the probability of not being speeded based on ability which is expressed as:

$$\pi(\theta_i) = \lambda + (1 - \lambda)\frac{1}{1 + \exp(-\theta_i)}, \tag{3.4}$$

where λ is the threshold probability of not being speeded. The threshold probability is the baseline probability of not being speeded for the test-taking population. In other words, every examinee in the test-taking population has the same baseline probability of not being speeded. If it is known that the stakes of an assessment are low and students' motivation levels are low, it is safe to assume λ is small. The baseline probability may also be small if the time constraints on a high-stakes assessment are improperly defined. The baseline probability can be adjusted based on the researcher's or practitioner's beliefs about the assessment. Though λ is typically estimated, under different contexts of an assessment, it is expected that this value is small. Equation 3.4 allows the probability of not being speeded to not only be a characteristic of the test but also of the population because of the inclusion of ability.

Equation 3.4 accomplishes this by involving the logit of the ability level multiplied by $(1 - \lambda)$. This product is added to λ, which signifies an increase in the probability of not being speeded. The higher the ability level is, the larger the logit will be, and thus the higher the probability of not being speeded. However, the probability of not being speeded remains close to the baseline if the ability level is low. In direct terms, the higher the ability level, the lower the probability of being speeded. On the other hand, the lower the ability level, the higher the probability of being speeded. Furthermore, λ is not equivalent to the 2PLMix model parameter, π_J. Although they both denote the probability not being speeded, λ is an intermediate parameter, whereas π_J describes the probability of not being speeded for an entire test-taking population.

Figure 3.1 shows the implementation of Equation 3.4, in which the number of quadrature nodes (between -3 and 3) for the ability parameter is 25 and the baseline probability of not being speeded is $\lambda = 0.4$. This baseline is based on the research conducted by Jin and Wang (2014). This figure shows that low-ability examinees have a low probability of not being speeded (i.e., a high probability of being speeded). Conversely, high-ability examinees have a high probability of not being speeded and thus a low probability of being speeded. It is very important to emphasize that each examinee has a distinct probability of not being speeded. Equation 3.4 does not add any new parameters to the model compared to the 2PLMix, but rather replaces the probability of not being speeded with a more realistic parameter using information already found in the 2PLMix model (the ability level of each examinee).

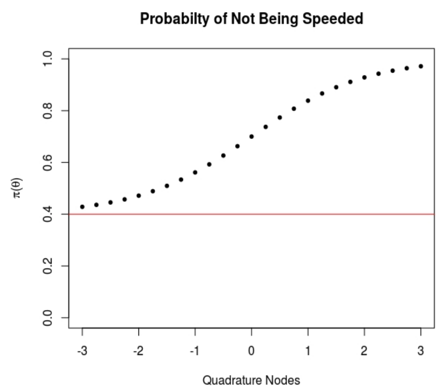

Figure 3.1: The probability of not being speeded based on 25 quadrature nodes between -3 and 3 and the baseline probability (red line) of not being speeded being $\lambda = 0.4$

Within the 2PLMix framework, the location at which an examinee becomes speeded is random. However, the M2PLMix model uses Equation 3.4, which influences the location where an examinee may become speeded. In other words, the ability not only influences whether an examinee becomes speeded but also the location at which an examinee becomes speeded. If an examinee has low ability, he or she has a higher probability of being speeded earlier on in the assessment compared to someone who has higher ability. This has a dramatic effect on the number of examinees who are speeded within a population, which will be further discussed within the following section. Equation 3.4 impacts the probability of being speeded on an assessment with respect to examinee i being speeded on item j, which can be expressed as:

$$\pi_j(\theta_i) = \frac{j^{\omega+1} - (j-1)^{\omega+1}}{(J-1)^{\omega+1}}(1 - \pi(\theta_i)) \tag{3.5}$$

where ω and J have the same definition as noted in the 2PLMix model and $\pi(\theta_i)$ denotes the probability of not being speeded for examinee i. Equation 3.5 is very similar to Equation 2.10, but includes the ability parameter, with Equation 3.4 embedded within Equation 3.5. Equations 3.4 and 3.5 imply that the location at which an examinee i becomes speeded, δ_i, follows a distribution denoted in a vector form as $[\pi_1(\theta_i), \pi_2(\theta_i), \ldots, \pi_J(\theta_i)]$. As stated with the 2PLMix model, if $\delta_i = J$, the i^{th} examinee is not speeded on any items. The higher the examinee's ability, the later the location of speededness (if it occurs at all).

Figures 3.2 and shows the relationships between parameters for the M2PLMix, 2PLMix, and 2PL models. Then 2PLMix displays how the decrement parameter, η_{ij}, depends on the the location of speededness, δ_i, and the speededness impact parameter, γ_{δ_i}, under the 2PLMix model. The location of speededness depends on a hyperparameter, ω, and the probability of being speeded, π_j. Under the M2PLMix model the decrement parameter is based on the location of speededness, δ_i, and the speededness impact parameter, γ_{δ_i}, under the M2PLMix model. However, δ_i depends on the probability of being speeded, $\pi_J(\theta_i)$, based on examinee i ability level, θ_i, ω, and the baseline probability of not being speeded, λ. The modified model is more realistic in that examinees with lower ability levels will become speeded more often and earlier on an assessment compared higher-ability examinees. The 2PL model shows that ability and item parameters effect X_{ij} which is common in the other two models.

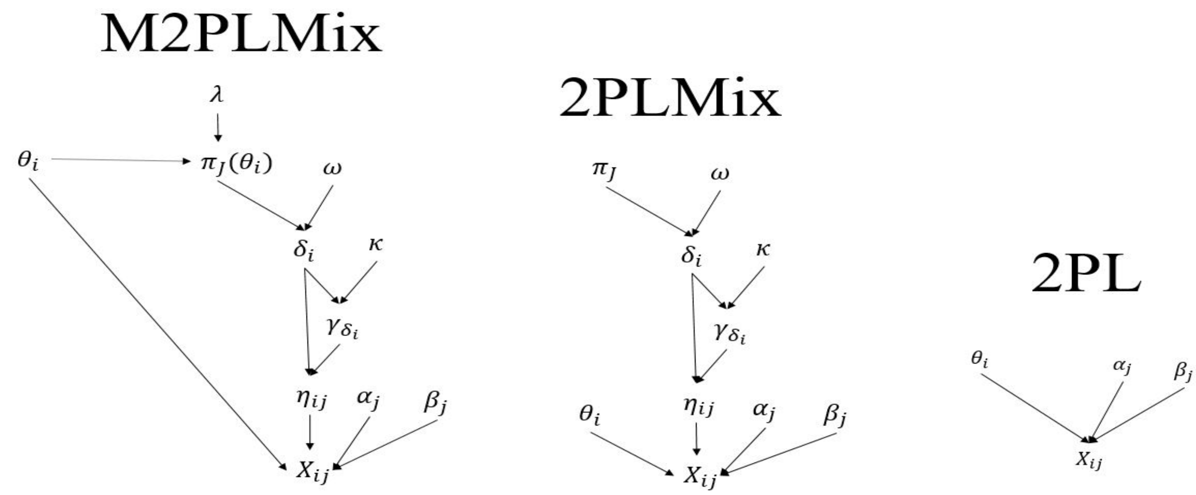

Figure 3.2: Displays the relationship between the parameters and how they influence the probability of X_{ij} within the M2PLMix, 2PLMix, and 2PL models.

3.3 Comparing the 2PLMix and M2PLMix

A clear way to illustrate the difference between the 2PLMix and M2PLMix models is to work through an example and applying each model. Assume there are $J = 20$ items, the probability of not being speeded is $\pi_J = 0.4$ for the 2PLMix, and the baseline probability of not being speeded is $\lambda = 0.4$ for the M2PLMix. The speededness impact factor is $\kappa = 0.2$ and the hyperparameter within the probability function is set to $\omega = 1$. These values are selected from Jing and Wang (2014). As noted in the previous section within the 2PLMix model, δ_i follows the discrete distribution of a Multinomial$[\pi_1, \pi_2, \ldots, \pi_{20}]$; however within the M2PLMix model, δ_i follows the discrete distribution of a Multinomial$[\pi_1(\theta_i), \pi_2(\theta_i), \ldots, \pi_{20}(\theta_i)]$. Table 3.1 represents the probability of being speeded on items $16 - 19$ for a person with high ability ($\theta = 3$) under M2PLMix, a person with low ability ($\theta = -3$) under M2PLMix, and a person generated under 2PLMix. This table also shows the speededness effect on the logit for both models once speeded.

Table 3.1: 2PLMix and M2PLMix probability distribution of speededness for $J = 20$ items and the speededness effect.

Probability of Being Speeded for Items 16-19				
	π_{16}	π_{17}	π_{18}	π_{19}
M2PLMix ($\theta = 3$)	$\pi_{16} = 0.002$	$\pi_{17} = 0.002$	$\pi_{18} = 0.002$	$\pi_{19} = 0.003$
M2PLMix ($\theta = -3$)	$\pi_{16} = 0.049$	$\pi_{17} = 0.052$	$\pi_{18} = 0.055$	$\pi_{19} = 0.058$
2PLMix	$\pi_{16} = 0.051$	$\pi_{17} = 0.054$	$\pi_{18} = 0.058$	$\pi_{19} = 0.061$
Speededness Effect	$\eta_{ij} = 0.8$	$\eta_{ij} = 0.6$	$\eta_{ij} = 0.4$	$\eta_{ij} = 0.2$

As noted in the previous section, the 2PLMix model assumes every examinee has the same probabilities of being speeded on the items of an assessment, whereas the M2PLMix demonstrates how ability modifies the probabilities of being speeded on the items of an assessment. The probability of becoming speeded on the 16^{th} item within the 2PLMix is 0.051 for all test-takers. In contrast, under the M2PLMix, an examinee with an ability level of $\theta = -3$ has a probability of 0.049 of becoming speeded on the 16^{th} item and an examinee with an ability of $\theta = 3$ has a probability of 0.002 for becoming speeded on that same item. This exemplifies how ability plays a role not only in an examinee becoming speeded but also where they become speeded. The probability of becoming speeded for each item is greater for an examinee with lower ability than an examinee with higher ability, which shows that someone with higher ability is less likely to become speeded compared to a low-ability examinee. The M2PLMix probability distribution for an examinee with low ability is comparable to the 2PLMix probability distribution. This occurs due to the logit of a low ability examinee, which translates to a small increase in the baseline probability of not being speeded. Table 3.1 draws attention to the fact that an examinee with high ability (e.g., a meticulous student) may still become speeded, and a low-ability examinee may be able to complete a test without being speeded.

A common factor between both models is the speededness effect, which shows that if an examinee became speeded on the 16^{th} item, the speededness effect would be 0.8 (i.e., $\eta = 0.8$) for items 17 though 20. Since both models assume the same speededness effect with respect to speededness location, the same values are used regardless of the models. Though table 3.1 shows how the ability level influences the location of the inception of speededness, another illustration will help clarify this notion.

A preliminary simulation analysis was implemented to illustrate the mechanism of the 2PLMix and M2PLMix models with respect to the location at which speededness began based on the models' assumptions, respectively. This simulation considered three different ability levels, $\theta = 3$ (high ability), $\theta = 0$ (medium ability), and $\theta = -3$ (low ability), so that generalizations about low-, medium-, and high-ability examinees can be made. The number of examinees within each ability level was 300. The baseline probability of not being speeded was $\lambda = 0.4$ within the M2PLMix, and the probability of not being speeded was $\pi_J = 0.4$ for the 2PLMix. Though these values have different meaning with respect to their corresponding models, they still contribute to the probability of not being speeded in a similar manner. This simulation assumed $J = 40$ items; however, simulating location of speededness for all 40 items is cumbersome. Therefore, for the summary of the results, the test was broken into five sections: the first section denotes becoming speeded on items 1-10, the second section denotes becoming speeded on items 11-20, the third section denotes becoming speeded on items 21-30, the fourth section denotes becoming speeded on items 31-40, and the fifth section is $\delta = 40$ which implies not becoming speeded.

Within the simulation, the location at which an examinee became speeded was generated 100 times, and the average was evaluated for each section of the test corresponding to the ability level for both models. Figure 3.3 displays the results from this simulation. The x-axis denotes the location at which an examinee became speeded, the y-axis represents the number of speeded examinees, and the three ability levels, high, medium, and low, are colored by black, red, and blue, respectively. The bar plot on the left panel is the M2PLMix, and the bar plot on

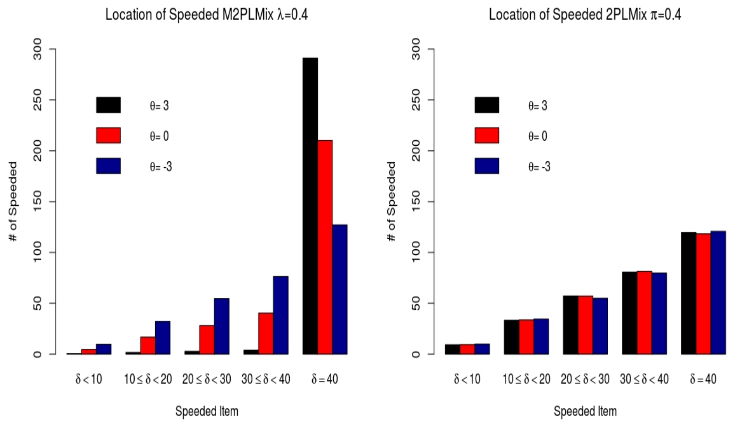

Figure 3.3: Locations at which examinees will become speeded according the M2PLMix (left panel) and the 2PLMix (right panel).

the right panel is the 2PLMix. The number of examinees that become speeded increases from one section to the next regardless of the models. However, there is a general trend for the M2PLMix within each section regarding the first four sections, in which the number of speeded examinees is the largest for low ability patrons, followed by medium ability, and then finally high ability levels has the lowest number of people affected by speededness. However, this does not occur for the fifth section of the test, in which the ranking order in magnitude is high-, medium-, and low-ability levels. Therefore, ability not only has an impact on whether an examine is speeded but also the location at which one becomes speeded, and those who with higher ability will be more likely not to be speeded.

This is not the case for the 2PLMix model, in which, within all five sections, almost the same number of examinees from each ability level become speeded within each section. This implies that no matter the ability level, the location at which an examinee becomes speeded is constant. Using the M2PLMix model dramatically changes the number of examinees affected by speededness compared to the 2PLMix model. This occurs because within the M2PLMix model, the high- and medium-ability examinees have a higher probability of not becoming speeded, and thus the total number of speeded examinees decreases. However, the location at which speededness occurs within the 2PLMix is not influenced by ability levels.Further comparisons of the proposed model to another existing model is provided in the next section to have a deeper understanding of the M2PLMix.

3.4 Comparing the M2PLMix and GPC Models

Although the M2PLMix and GPC models both account for the association between ability and speededness, this association is carried out in two very distinct manners. The mechanism of the M2PLMix and GPC models differ on three primary factors: 1) the level probabilistic theory required, 2) the practicality of the number of people affected by speededness, and 3) the impact of speededness on probability.

Equation 3.4 demonstrates the probabilistic theory required to understand how ability is related to speededness within the M2PLMix model. More specifically, ability and speededness are simply inversely proportional to one another, meaning that as ability increases, the probability of being speeded decreases, and as the ability decreases, the probability of being speeded increases. As mentioned previously this also has an impact on which item is likely to become speeded on. This is in contrast to the GPC model, where understanding the relationship between ability and speededness requires advance knowledge of a copula function. The copula function forces marginal distributions of random variables to have a dependence structure based on some known correlation matrix. The copula function can be represented as

$$G(\theta, \eta, \lambda) = C(G_1(\theta), G_2(\eta), G_3(\lambda)) \tag{3.6}$$

where C is the copula function, G_1 is the marginal distribution for θ (ability), G_2 is the marginal distribution for η (inception of speededness), and G_3 is the marginal distribution for λ (speededness intensity). A correlation structure must be provided in order to force an association between the marginal distributions. In Goegebeur et al. (2008), the marginal distributions for the θ, η, and λ were set as $N(0, 1)$, $\beta(\alpha, \beta)$, and $\log N(\mu_\lambda, \sigma_\lambda)$, respectively. These distributions and hyperparameters are used to illustrate the model which may need to change to reflect real life scenarios. These hyperparameters make the GPC model more convoluted than the M2PLMix model.

The second difference stems from the number of people shown to be affected by speededness under each model. Under comparable conditions found in the M2PLMix model, the GPC model's correlation structure can be seen as $\rho(\theta, \eta) = 0.5$ (correlation between ability and the location at which speededness occurs) and $\rho(\eta, \lambda) = \rho(\theta, \lambda) = 0.2$, in which 99% of the sample is calculated to be speeded. In contrast, creating an association between ability and speededness such that the M2PLMix model comparable to the GPC model ($\lambda = 0.4, \omega = 2$), 40% of the sample is calculated as speeded. This contrast can be seen in Figure 3.4. This figure shows the proportion of the test completed with respect to ability level. This figure demonstrates that nearly the entire test population would be experience test speededness under the GPC model, while the M2PLMix has a majority of the sample completing the test.

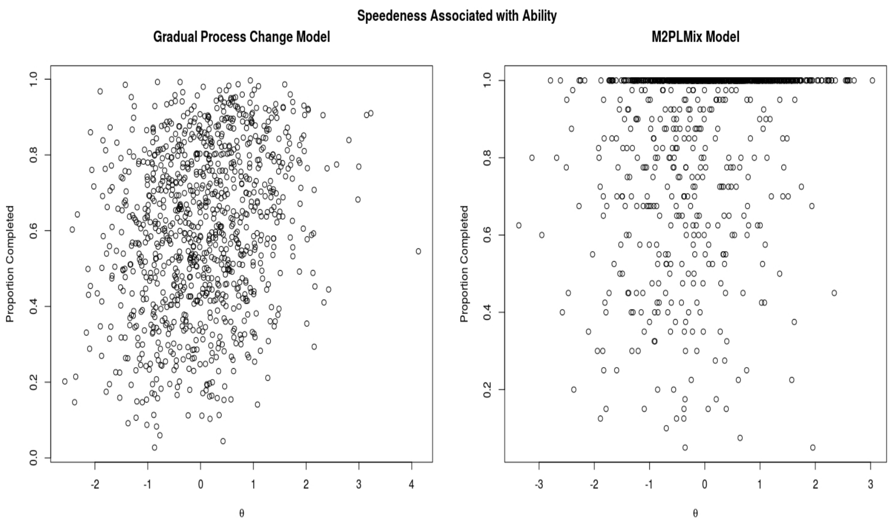

Figure 3.4: The proportion of test completed with respect to ability under the GPC (left panel) and M2PLMix models (right panel).

The third difference between the two models is the manner by which speededness affects the probability of a correct response. The GPC directly impacts the probability function of obtaining the correct solution, seen in equation 2.1. In contrast, the M2PLMix speededness impacts the logit function, indirectly affecting the probability of a correct answer. Figure 3.5 shows the probability of a correct answer for a 40 item test. Both models assumed that for an ability level of 0, speededness occurs halfway through the assessment (the vertical line shows speededness location). This figure shows that the impact of speededness on correctness can be equivalent under both models. However, the rate of the individuals affected by speededness within the GPC model is much greater than the M2PLMix model, so much that the former becomes unrealistic. Lastly, the complexity behind understanding copula functions could prevent test developers from correctly implementing the GPC model. These differences are the key to understanding how they are used to model reality.

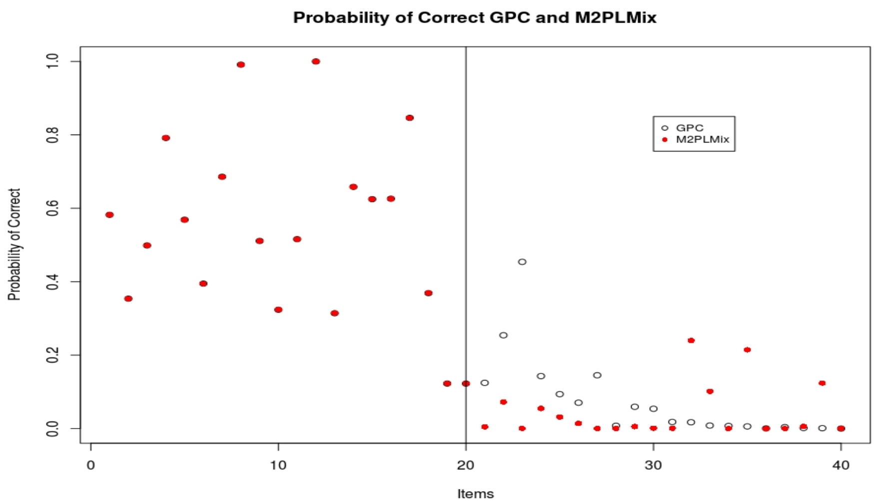

Figure 3.5: Probability of Correct under the GPC (black circle, $\lambda = 4$) and M2PLMix models (red circle).

3.5 Estimation of Model Parameters: Markov Chain Monte Carlo Algorithm

The technique used to estimate the parameters in the 2PL, 2PLMix, and M2PLMix models is discussed in this section. When estimating speeded model parameters, researchers have implemented either the Marginal Maximum Likelihood Estimation (MMLE) algorithm or the Markov Chain Monte Carlo (MCMC) algorithm, a Bayesian technique. The parameters estimated within the HYBRID model proposed by Yamamoto and Everson (1997) used the computer program HYBIL (Yamamoto, 1989), which implemented MMLE. In contrast, the other speeded models discussed in this dissertation, including the proposed M2PLMix, used a variation of the MCMC algorithm to estimate parameters. MCMC is a technique used to generate samples from a posterior distribution, which is then used to find the estimates of the parameters within a model. In Bayesian statistics, the posterior distribution is the probability of the parameters within a model given data. The posterior distribution can be expressed as the joint probability of the parameters within the model and the data multiplied by the probability of the data (Junker, Patz, & VanHoudnos, 2012; Gelman, Carlin, Stern, & Rubin, 1995). This can be expressed as:

$$f(\tau|X) = f(\tau, X)f(X) \tag{3.7}$$

where X is the data and τ represents the model parameters. Using Bayes theorem, the posterior distribution can also be shown as:

$$f(\tau|X) = \frac{f(X|\tau)f(\tau)}{\int f(X|\tau)f(\tau)d\tau} \propto f(X|\tau)f(\tau). \tag{3.8}$$

In modern statistics, it is quite challenging to compute the integral of the denominator; therefore, the numerator is typically used as a proxy for the posterior distribution. Once the posterior distribution is approximated, there are many different methods for finding the estimates of the parameters. One method is to determine the posterior mean of the estimate of τ, formally known as the expected a posteriori (EAP) of τ, which can be expressed as:

$$E(\tau|X) = \int \tau f(\tau|X)d\tau. \tag{3.9}$$

Another method is to find the maximum a posteriori (MAP) of τ, which is the posterior mode, and can be expressed as:

$$\underset{\tau}{\text{argmax}} = \frac{f(\tau|X)f(X)}{\int f(\tau|X)f(X)dx}. \tag{3.10}$$

For this study, the EAP and MAP are used to summarize the posterior distribution, depending on the nature of the parameter of interest (continuous vs. discrete). In order to find the posterior distribution, a sampling technique or Monte Carlo integration must be implemented. However, this dissertation only discusses sampling techniques. In this case, the process of finding the posterior distribution involves drawing a series of samples generated from a sampling distribution denoted as $\tau^1, \ldots, \tau^k$, where τ^k is parameter estimate at the k^{th} iteration. Within MCMC, there are two types of sampling procedures: Gibbs Sampling and Metropolis Hastings.

3.5.1 Gibbs Sampling

Gibbs Sampling (Gelfand & Smith, 1990) uses complete conditional distributions to directly sample and find the posterior distribution. There are many methods of directly sampling from the posterior distribution; this dissertation discusses two methods: inversion sampling and rejection sampling (Junker, Patz, & VanHoudnos, 2012). Inversion sampling involves sampling a random number, u, from a uniform distribution $U(0, 1)$. The inverse of the cumulative distribution function (CDF) of the random number, $F^{-1}(u)$, would be drawn from the desired distribution. Inversion sampling can be used only when both the CDF and the inverse of CDF can be expressed in a closed form.

When the inverse of the CDF cannot be expressed in a closed form, rejection sampling can be used. Rejection sampling draws a random variable, z, from the distribution $g(z)$, assuming that $cg(z) > f(z)$ for the support of z, where c is some constant. From here, a random number, u, is sampled from a uniform distribution $U(0, 1)$ and is compared to $R = f(z)/cg(z)$. If $u < R$, then z is accepted from

$f(z)$; otherwise, it is rejected. These Gibbs Sampling methods are quite robust, but are not useful when the complete conditional distribution is unknown (Junker, Patz, & VanHoudnos, 2012).

3.5.2 Metropolis Hastings

The Metropolis Hastings algorithm (Chib & Greenberg, 1995; Hasting, 1970; Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953) presents a viable alternative if one cannot sample from the complete conditional distribution. The first step is to generate τ from a proposal distribution $g(\tau|\tau^{k-1})$. This proposal distribution in this estimation algorithm is the prior distribution of τ^{k-1}. It is typical to allow the proposal distribution to be chosen such that it is independent of the previous iteration of τ^{k-1} and follows a normal distribution $N(\tau^{k-1}, \sigma)$. Once a random draw is computed, the following computation can be expressed:

$$\alpha(\tau, \tau^{k-1}) = min\left\{\frac{f(\tau|X)g(\tau^{k-1}|\tau)}{f(\tau^{k-1}|X)g(\tau|\tau^{k-1})}, 1\right\} \tag{3.11}$$

The minimum of equation 3.11 is then compared to a random number, u, which is sampled from a uniform distribution $U(0, 1)$. It can be concluded that if $u \leq \alpha(\tau, \tau^{k-1})$, then $\tau^k = \tau$, otherwise $\tau^k = \tau^{k-1}$. Once the sample of draws is large enough (burn-in is established), the techniques used to summarize the posterior distribution sample (e.g., EAP or MAP) may be used.

3.5.3 Facets of MCMC

There are several concepts within the MCMC method that are pertinent to the study at hand. Within the process of estimating parameters, starting near true value, using the right amount of burn-in period, and measuring the convergence of estimated parameters are critical in determining if estimates are near the true parameter value. Another important concept is blocking, which is used to efficiently estimate parameters in groups rather than a single parameter at a single time.

Initial Parameters In order to avoid using too many iterations or risk extreme tails of the posterior distribution (which may cause a chain not to converge) in estimating model parameters, it is ideal to choose initial values that are close to true parameters based on the data (Spiegelhalter et al, 1996; Thomas & Gauderman, 1996) or preconceptions about the model parameters. For example, in Patz and Junker's (1999) MCMC algorithm used to estimate IRT model parameters, all the ability and difficulty parameters were initialized to be 0 and discrimination parameters were initialized at 1. This was implemented based on the generation of the model parameters.

Burn-In When implementing MCMC it is imperative that a burn-in period is allowed so that the true value is approached. In MCMC estimation procedure it is expected that estimation chains eventually converge to the ideal stationary distribution, which is the target distribution. This can be done after a number of iterations is thrown away and the remaining are used to determine the estimated parameters. A typical length of burn-in period is 5000 iterations (e.g., de la Torre, 2009). The following two paragraphs discuss ways of determining if the parameter estimate obtained after a burn-in period has stabilized to a specific value.

Convergence There are many ways to visually determine if an estimated value has stabilized or converged to a specific value. Traceplots are used to display iteration number versus the estimated value of the draw of the parameter at each iteration. If the draw remains consistent around one particular value, then convergence is found. An autocorrelation plot is also used to measure the lag effect of the estimated parameter for each iteration. It is ideal for the plot to have a high correlation during its inception and show a rapidly decreasing pattern. Lastly, a density plot can be used to detect if a parameter has a unimodal distribution, which is ideal for estimation.

To measure convergence analytically, the Geweke diagnostic (Geweke, 1992) and Gelman and Rubin statistic (Gelman & Rubin, 1992) can be used. The

Geweke diagnostic is used to verify convergence for one chain with respect to the estimated parameters. This approach produces a z-score for every parameter estimated, and if the z-score is within ± 2, the parameter is assumed to have converged. Alternatively, the Gelman and Rubin statistic, $\sqrt{R}$, uses multiple chains to determine whether the proportion error found in the estimated parameter is attributed to Monte Carlo error or lack of convergence. It is ideal that the $\sqrt{R} < 1.2$ for each estimated parameter (de la Torre & Douglass, 2004).

Blocking Blocking is a technique that handles the issue of complete conditional densities which use univariate densities and require each parameter in the model to be estimated one at a time. That being said, it allows for estimation to occur for a group of parameters simultaneously. Blocking also avoids the problem of low rates of acceptance. It is ideal to block parameters based on elements that are related to each other (e.g., item parameters).

3.6 Estimation of the 2PL, 2PLMix, and M2PLMix Models

The majority of speededness models that implement MCMC in the literature have used WinBUGS (Spiegelhalter, Best, Carlin, & van der Linde, 2002), a computer software that uses Gibbs sampling, as the methodology of calibrating item parameters, ability levels and speededness parameters (e.g., Bolt, Cohen, & Wollack, 2002; Jin & Wang, 2014; Suh, Cho, & Wollack, 2012). Currently, no speededness model has been implemented using MCMC Metropolis Hastings to estimate parameters within the speeded model. In this study, the MCMC Metropolis Hastings algorithm is applied to estimate the parameters of the 2PL, 2PLMix, and M2PLMix models. The software used to generate data and analyze the estimated parameters was R (R Development Core Team, 2012), and the estimation of model parameters was implemented in C++ using various libraries, such as Armadillo and Boost (Sanderson & Curtain, 2016; Schaeling 2008). The item parameters, ability level, and the location at which speededness begins, δ_i, were estimated within the speeded models, whereas the baseline probability of not being speeded, λ, and probability of not being speeded, π_J, κ and ω are assumed to be known.

The parameters were initialized by setting all the discrimination parameters to 1 and setting all δ_is to the maximum number of items (not speeded). The former was based on the fact that the discrimination parameter has to be greater than 0. The latter, however, was set so that the estimation models can gather enough evidence from the data to signify that speededness has occurred. In contrast, an ad-hoc method was implemented to initialize the ability and difficulty parameters. This process involved finding the proportion correct for each item (difficulty) and the proportion correct for each examinee (ability) from the data data. These proportions were used to find z-scores, these z-scores were then multiplied by 0.95 and increased by 0.05 to ensure no values were close to 0. The initial difficulty estimates were multiplied by -1 to ensure that these values start near the true parameter. This method is better than setting all the initial values for β and θ to 0 or random numbers because if the item is difficult or the ability level high, the initial parameter will be positive, otherwise it will be negative.

In order to implement the MCMC Metropolis Hastings algorithm, the posterior distribution (likelihood) must be stated. The following explanation of the likelihood is shown only for the M2PLMix model, but assuming δ_i and θ_i are independent would suffice for the 2PLMix model and ignoring δ_i would suffice for the 2PL model. The posterior distribution is proportional to the conditional distribution of data given the unknown parameters multiplied by the prior distributions of the unknown parameters which can be expressed as follows:

$$f(X|\boldsymbol{\tau})f(\boldsymbol{\tau}) \tag{3.12}$$

where $\boldsymbol{\tau} = (\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\delta})$. The ability parameters were $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_N)$, the discrimination parameters were $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_J)$, the difficulty parameters were $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_J)$ and the location at which speededness begins were $\boldsymbol{\delta} = (\delta_1, \delta_2, ..., \delta_N)$. The conditional distribution, $f(X|\boldsymbol{\tau})$, can be expressed as

$$f(X|\boldsymbol{\tau}) = \prod_{i=1}^{N}\prod_{j=1}^{J} f(X_{ij}|\alpha_j, \beta_j, \theta_i, \delta_i), \tag{3.13}$$

in which the probability distribution for X_{ij} is the following

$$f(X_{ij}|\alpha_j, \beta_j, \theta_i, \delta_i) = P_{ij}^{X_{ij}}(1 - P_{ij})^{(1-X_{ij})}, \tag{3.14}$$

where the probability of the correct response (X_{ij}) can be seen as:

$$P_{ij} = P(X_{ij} = 1|\alpha_j, \beta_j, \theta_i, \delta_i) = \frac{\exp(\alpha_j(\theta_i - \beta_j - \eta_{ij}))}{1 + \exp(\alpha_j(\theta_i - \beta_j - \eta_{ij}))}. \tag{3.15}$$

The prior distribution for the M2PLMix model parameters, $f(\boldsymbol{\tau})$, can be expressed as:

$$\begin{aligned} f(\boldsymbol{\tau}) &= f(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\delta}) \\ &= f(\boldsymbol{\alpha}) f(\boldsymbol{\beta}) f(\boldsymbol{\delta}, \boldsymbol{\theta}) \\ &= f(\boldsymbol{\alpha}) f(\boldsymbol{\beta}) f(\boldsymbol{\delta}|\boldsymbol{\theta}) f(\boldsymbol{\theta}) \\ &= \prod_{j=1}^{J} f(\alpha_j) \prod_{j=1}^{J} f(\beta_j) \prod_{i=1}^{N} f(\delta_i|\theta_i) \prod_{i=1}^{N} f(\theta_i) \end{aligned} \tag{3.16}$$

where the hyperparameters within each prior distribution are based on previous studies (Jin & Wang, 2014; Junker, Patz, & VanHoudnos, 2012). The prior distributions for each parameter can be seen as:

$$\begin{aligned} f(\theta_i|0, 1) &= N(\theta_i|0, 1) \\ f(\alpha_j|0.3, 1) &= \text{log-}N(\alpha_j|0.3, 1) \\ f(\beta_j|0, 1) &= N(\beta_j|0, 1) \\ f(\delta_i|\theta_i, \lambda) &= \text{Multinomial}[(\pi_1(\theta_i), ..., \pi_J(\theta_i)] \end{aligned} \tag{3.17}$$

The complete likelihood function can be expressed as:

$$f(X|\boldsymbol{\tau})f(\boldsymbol{\tau}) = \prod_{i=1}^{N}\prod_{j=1}^{J} f(X_{ij}|\alpha_j, \beta_j, \theta_i, \eta_i(\delta_i)) \prod_{j=1}^{J} f(\alpha_j) \prod_{j=1}^{J} f(\beta_j) \prod_{i=1}^{N} f(\delta_i|\theta_i) \prod_{i=1}^{N} f(\theta_i). \tag{3.18}$$

The complete conditional densities for each individual parameter can be seen as:

$$f(\theta_i|rest) \propto \prod_{i=1}^{N} P_{ij}^{X_{ij}}(1-P_{ij})^{(1-X_{ij})} N(\theta_i|0,1)\ \forall i = 1, ..., N, \tag{3.19}$$

$$f(\alpha_j|rest) \propto \prod_{j=1}^{J} P_{ij}^{X_{ij}}(1-P_{ij})^{(1-X_{ij})} \text{log-}N(\alpha_j|0.3,1)\ \forall i = 1, ..., J, \tag{3.20}$$

$$f(\beta_j|rest) \propto \prod_{j=1}^{J} P_{ij}^{X_{ij}}(1-P_{ij})^{(1-X_{ij})} N(\beta_j|0,1)\ \forall i = 1, ..., J, \tag{3.21}$$

$$f(\delta_i|rest) \propto \prod_{i=1}^{N} P_{ij}^{X_{ij}}(1-P_{ij})^{(1-X_{ij})} \text{Multinomial}[\pi_1(\theta_i), ..., \pi_J(\theta_i)]\ \forall i = 1, ..., N. \tag{3.22}$$

where the prior probability of each parameter is multiplied by the distribution of X_{ij} for each item or examinee, respectively. The term, $rest$, denotes the remaining parameters that are being estimated in the model. The algorithm is seen appendix G.

Chapter 4

Simulation Study

4.1 Description of Four Studies

As mentioned previously, to evaluate the viability of the M2PLMix model, four studies were investigated via a simulation study. The first study was used to determine the implications of assuming the hyperparameters (λ, ω, and κ) were known during the estimation of M2PLMix model parameters. This was done by generating data in which specific values for the hyperparameters were chosen. This state was referred to as the normal condition because these values were used in studies 2-4 for the M2PLMix and 2PLMix models. Next, data were generated in which each hyperparameter was modified individually (further discussed in Data Generation). Lastly, the model parameters (α, β, θ, & δ) were then estimated under each condition. The model parameter estimates with the modified hyperparameters were then compared to the model parameters under the normal condition. This was done to examine the recovery of model parameters when speededness hyperparameters differ within the generation of data and during the estimation of model parameters.

The second study was designed to determine if the M2PLMix model can recover model parameters when speededness was not present. The design of this study was also intended to show that the 2PL (a traditional IRT model) does not estimate model parameters well when the data were generated under the M2PLMix model. The third study was used to determine if the 2PLMix model (ability and speededness were not associated) can estimate model parameters in a case where ability and speededness were associated. This study was also used to verify if the M2PLMix can estimate model parameters well when ability and speededness were not associated. Lastly, the fourth study was designed to determine how the M2PLMix and 2PLMix perform when speededness and ability were associated, but the data were generated under the GPC model. Studies 2 and 3 considered the same simulation factors to answer their respective questions. The factors studied include sample size, test length, probability of not being speeded (baseline) and item difficulty ordering.

4.1.1 Generation of Data

The discrimination parameters for the items were generated from a lognormal distribution with mean and standard deviation of 0.3 and 1, respectively. Furthermore, the item difficulty parameters and the examinees' ability levels were both generated from a normal distribution with mean and variance of 0 and 1, respectively. The item parameters can be found in Appendix E. The location of the inception of speededness was generated using both the M2PLMix and 2PLMix models discussed in the previous chapters in which the hyperparameters were $\omega = 2$, $\kappa = 0.2$, and the probability of not being speeded (baseline) was a factor. The distributions for these parameters used to generate the data within this dissertation were based on the generation of data found in Jin and Wang's (2014) study.

The item and ability level parameters were generated to be consistent under each condition and replication for the first 3 simulation studies. For example, the examinees' abilities were the same across all conditions and replications when for a particular sample size (i.e., $N = 500$) for all models. This was not the case for the ability levels within the fourth study. The generation of data under the

GPC model, used in the fourth study, required the ability level and speededness parameters to be generated through a known correlation structure. The data generation algorithm for each model used to generate data is found in appendix F.

The first study generated data in which the hyperparameters were $\lambda = 0.4$, $\omega = 2$, and $\kappa = 0.2$ based on Jin and Wang's (2014) study. These hyperparameters were then modified by the following design: λ and κ by ± 0.05 and ω by ± 0.5 to generate data. These deviations were based on retaining the functionality of the M2PLMix model and each deviation occurred in isolated events. Meaning that if κ became 0.25 then the following parameters remained $\lambda = 0.4$ and $\omega = 2$. Therefore, the normal condition composed of all unmodified parameters and the two modifications of each hyperparameter implied seven different hyperparameter conditions. The sample size was $N = 1000$ and number of items was $J = 40$ for all seven conditions. The number of replications for each condition was set to $R = 15$. The M2PLMix model parameters were then estimated assuming $\lambda = 0.4$, $\omega = 2$, and $\kappa = 0.2$. The potential values of the hyperparameters were based on the conditions found in Cao and Stokes (2008) and Jin and Wang (2014).

The factors investigated under studies 2 and 3 were sample size, test length, probability of not being speeded (baseline) and item difficulty ordering. There were three levels of sample size: $N = 500$, 1000, and 2000; three levels of test length: $J = 20$, $J = 40$, and 80; three levels of probability (baseline) of not being speeded: $\pi = 0.2, 0.4$, and 0.6 (2PLMix model) or $\lambda = 0.2, 0.4$, and 0.6 (M2PLMix model); and the item difficulty was either randomly ordered or ordered from easiest to most difficult. It should be noted that when the data were generated under the 2PL model, the speededness factors were not included. It was imperative to study various sample sizes and test lengths because no research has evaluated how the number of examinees or the lengths of a test impact the calibration of item parameters and the estimation of ability parameters under the speededness conditions. Though multiple studies have evaluated different types of speededness, no work has cross evaluated speededness and item difficulty ordering in an comprehensive manner. Thus studying these factors adds to the field of education measurement research in that this work poses questions that have yet to be answered.

Data generated under the speeded models within studies 2 and 3 implied 54 unique conditions for each speeded model. Additionally, data generated under the 2PL model resulted in 18 distinct conditions. The number of replications for each condition was set to $R = 40$. The simulation factors and the corresponding levels were based on previous studies that were used to research the effects of speededness (Cao & Stokes, 2008; Jin & Wang, 2014; Oshima, 1994; Goegebeur et al., 2008).

The second study included the generation of data under the M2PLMix and 2PL models in which the M2PLMix and 2PL models parameters were estimated for both types of generation of data. This design was also implemented for the third study in which data were generated under the M2PLMix and 2PLMix models in which the M2PLMix and 2PLMix models parameters were estimated for both types of generation of data. This design was imperative to answer the research questions proposed.

The generation of the data under the fourth study was created by correlating the speededness parameters within the GPC model using the following three correlation structures:

- $\rho(\theta, \eta) = 0.5, \quad \rho(\eta, \lambda) = 0.2, \quad \rho(\theta, \lambda) = 0.2$
- $\rho(\theta, \eta) = 0.6, \quad \rho(\eta, \lambda) = 0.2, \quad \rho(\theta, \lambda) = 0.2$
- $\rho(\theta, \eta) = 0.7, \quad \rho(\eta, \lambda) = 0.2, \quad \rho(\theta, \lambda) = 0.2$

where ρ denotes the correlation between two parameters. These values were used to mimic the capability of the M2PLMix model in that there was an association between ability and speededness via $\rho(\theta, \eta)$. Since neither ability and speededness intensity $(\rho(\theta, \lambda))$ nor speededness location speededness intensity $(\rho(\eta, \lambda))$ were associated within the M2PLMix model, the correlation between these values were set to 0.2. Furthermore, three different levels of association were used to depict the relationship between ability and the speededness location to imitate the varying levels of the baseline probability of not being speeded under the M2PLMix model.

The marginal distributions for θ, η, and λ were $N(0,1)$, $\beta(2,2)$, and $\log N(0, 0.01)$, respectively. The parameters used within these distributions were based on Goegebeur's et. al (2008) study. The three structures were evaluated in order to determine how the M2PLMix and 2PLMix models performed when the association between speededness and ability was generated from a different model.

The fourth study primarily focused on how ability and speededness were associated differently; however, the intensity of speededness was assumed to have the same relationship between ability and location of speededness. The sample size was $N = 1000$ and number of items was $J = 40$. $R = 15$ replications were generated for each condition. Once the data were generated under the GPC model, both the M2PLMix and 2PLMix models were used to estimate their model parameters, respectively.

4.1.2 Estimation of Model Parameters

All four studies exclusively used the proposed algorithm written in C++ and various libraries (i.e., Armadillo, Boost) to estimate the 2PL, 2PLMix, and M2PLMix model parameters (2PL code found in Appendix D). This algorithm was implemented using parallel programming to increase the number of data sets the CPU could estimate at a time. Amazon Web Services Elastic Cloud Compute was also used to enhance the number of programs that could run in parallel. Code for the parallel programming is found in Appendix D.

The parameters of interest within each estimation algorithm, regardless of the model, were the discrimination (α), difficulty (β), and ability (θ) parameters. However, within the speeded models, an additional parameter of interest was the location at which speededness begins (δ). There were two chains created for each parameter. For each replication of a particular condition, the number of iterations was $10,000$, in which the burn-in was $5,000$ iterations. The EAP method was used to estimate all model parameters except for δ. This parameter is discrete; therefore, the MAP method was used.

4.2 Evaluating Simulation Studies

Since MCMC was used to estimate the model parameters, multiple aspects of the estimation process were investigated to ensure the model parameter estimates were near expected (true) parameters. Three questions were posed to evaluate the estimated parameters from each model:

1. Does the Markov chain converge to a value for each parameter?
2. Were the estimates of the model parameters close to the original (true) values?
3. Does the model (estimated parameters) fit the data well?

The first question was used to examine whether the model parameters converged, whereas the second and third questions were used to compare the models themselves. These evaluative aspects are typical concepts discussed in recent research of speededness (Jin & Wang, 2014; Cao & Stokes,2008; Chang, Tsai, & Hsu, 2014), and thus are discussed in detail in the following subsections.

4.2.1 Model Convergence

When estimating parameters within the implementation of MCMC, chain convergence is key to the evaluation and stability of the estimated parameters. There are many techniques that exist to determine if a chain converged, which was discussed in section 3. The Gelman and Rubin $\sqrt{R}$ statistic was implemented to analytically show convergence in this dissertation. The average percentage of non-converged parameters for each condition was calculated for each IRT model parameter estimation (i.e., α, β, θ). Autocorrelation and trace plots of each parameter are traditionally used to ensure proper convergence graphically. However, due to the number of plots required, none will be provided.

4.2.2 Evaluation Criterion

Once it is known that the model parameters have properly converged, it is imperative to determine how well each parameter was recovered. The measures used to evaluate the recovery of the traditional IRT model parameters were the bias and root mean squared error (RMSE). The bias and RMSE of a parameter, ξ, where ξ can represent α, β, or θ parameters, are denoted as the following:

$$bias(\xi) = \sum_{r=1}^{R} \frac{(\xi - \hat{\xi})}{R} \tag{4.1}$$

$$RMSE(\xi) = \sum_{r=1}^{R} \sqrt{\frac{(\xi - \hat{\xi})^2}{R}} \tag{4.2}$$

where $\hat{\xi}$ is the estimated parameter, ξ is the true parameter, and R is the number of replications. These measures are common tools to evaluate the accuracy of the calibration of item parameters and estimation of the ability parameters within speeded models in the literature (e.g., Lee & Ying, 2015; Chang, Tsai, & Hsu, 2014; Brown, Li, & Yang, 2013). The standard deviation of each criterion was determined as well to show the variation in estimating a parameter over multiple replications.

Several different methods were proposed to investigate the recovery of the speededness location parameter. Five distinct techniques were implemented to measure the accuracy of the inception of speededness. These measures are as follows:

1. M_1 denotes the average proportion of correct specification

$$M_1 = \sum_{i=1}^{N} \frac{\mathbb{1}(\hat{\delta}_i = \delta_i)}{N}, \tag{4.3}$$

2. M_2 denotes the average proportion of correct specification of the inception for speeded examinees

$$M_2 = \sum_{i=1}^{N} \frac{\mathbb{1}(\delta_i \neq J)\mathbb{1}(\hat{\delta}_i \neq J)}{\sum_{i=1}^{N} \mathbb{1}(\delta_i \neq J)}, \tag{4.4}$$

3. M_3 denotes the average proportion of correct specification of the inception for non-speeded examinees

$$M_3 = \sum_{i=1}^{N} \frac{\mathbb{1}(\delta_i = J)\mathbb{1}(\hat{\delta}_i = J)}{\sum_{i=1}^{N} \mathbb{1}(\delta_i = J)}, \tag{4.5}$$

4. M_4 represents the average of relative bias (Forero & Maydeu-Olivares, 2009)

$$M_4 = \frac{\sum_{i=1}^{N} \frac{\delta_i - \hat{\delta}_i}{\delta_i}}{N}, \tag{4.6}$$

5. M_5 denotes the average proportion difference between the true proportion of speeded examinees and the estimated proportion of speeded examinees

$$M_5 = \sum_{i=1}^{N} \frac{\mathbb{1}(\delta_i \neq J)}{N} - \sum_{i=1}^{N} \frac{\mathbb{1}(\hat{\delta}_i \neq J)}{N}, \tag{4.7}$$

where $\mathbb{1}$ represents an indicator function (i.e. equal to 1 when the statement within the parenthesis is true and 0 when the statement is false), N is the sample size and J is the number of items. These values were averaged with respect to all replications for each condition.

M_1, M_2, and M_3 were very similar in that they were used to measure correct specification of being speeded or not being speeded, using the location parameter, δ . M_1 was able to determine if the specification of the inception of speededness or being not being speeded was correct. Further, M_2 verifies those that were speeded to be estimated as speeded. M_3 verifies those that were truly not speeded to be estimated as not speeded. Since these measures were proportions, the range of possible values were between 0 and 1. The closer these statistics were to 1, the better the estimation of the inception of speededness. M_1 & M_3 were conservative statistics, in that if $\hat{\delta}_i$ was not exactly δ_i, a misspecification will have occurred. In contrast, $\hat{\delta}_i$ does not have to be exactly δ_i to be a correct specification for M_2.

M_4 represents the difference between δ_i and $\hat{\delta}_i$ in relation to δ_i, meaning that the closer the estimated value is to the true value, the smaller the relative bias becomes. M_4 ranges from $-\frac{1-J}{J}$ to $(J-1)$. However, this statistic has an intrinsic penalty for being speeded earlier on an assessment and obtaining an incorrect $\hat{\delta}_i$ in terms of estimation. For example, if two examinees were speeded at two different locations, say item 4 and item 35, and the model estimated the speeded locations to be 6 and 37 respectively, the relative bias for the first examinee is 0.5 and the second is 0.05. This is important to note when evaluating the models in the results section. It is also important to note that the implementation of M_4 was proposed by Forero and Maydeu-Olivares (2009) in which they defined any value greater than 0.2 was not ideal. The last measure, M_5, measures the proportion difference between the true and estimated δ_i parameters with respect to the proportion of specifying a person as speeded. This measure's possible values were between -1 to 1 and the closer M_5 is to 0 the better.

Currently, no research has created a robust index to measure the location of where speededness occurs for each individual. This is probably attributed to the fact that measuring the exact location in which speededness begins is very difficult and imprecise. However, these measurements were used to provide an initial attempt of evaluating the accuracy of measuring speededness location.

4.2.3 Model Comparison

Model comparison analyses were conducted to determine which model fitted the data best among the models being compared. Note that the model comparison was only conducted in studies 2 and 3. Multiplying -2 by the log-likelihood function ($-2 \log L$) is a typical technique used to compare two models. The likelihood uses the estimated parameters within equation **??** can be seen as:

$$-2 \log L = -2 \log L(X|\hat{\xi}) \tag{4.8}$$

where $\hat{\xi}$ denotes the estimated model parameters and X is the data. Other measures can be used to determine the fit of parameters to the data such as the Akaike Information Criterion (AIC) which is used when investigating the complexity of a model versus how well the model fits the data. This measure can be seen as:

$$AIC = 2k - 2 \log L(X|\hat{\xi}) \tag{4.9}$$

where k denotes the number of estimated parameters. A correction is often needed when estimating many parameters, therefore the Akaike Information Criterion Correction (AIC_c) is provided as:

$$AIC_c = AIC + \frac{(2k(k+1))}{(n-k-1)}, \tag{4.10}$$

where n denotes the sample size. Another measure named the Bayesian Information Criterion (BIC) is used to reduce over fitting which can be represented as:

$$BIC = -2 \log L(X|\hat{\xi}) + k \log(n). \tag{4.11}$$

Lastly, the adjusted Bayesian Information Criterion is used as another method of comparing the models:

$$AdjBIC = -2 \log L(X|\hat{\xi}) + k \log(\frac{n+2}{24}). \tag{4.12}$$

4.3 Results

Each study was presented in a different format to ensure that results patterns were recognized and the differences between the models were observed easily. For the first and fourth studies, the majority of the results are presented within the main body of the dissertation. However, the second and third studies conditions that are checked in Table 4.1 are shown in the following sections and the remaining are provided in the Appendix in terms of the recovery of parameters and model fit. Therefore, there are only 18 conditions when the data were generated under a speededness model (i.e. M2PLMix & 2PLMix) and there are 6 conditions when the data were generated under the 2PL model presented in the results section. This was carried out due to space, but generalizations are extended to the results shown in the appendices. The plots of the bias and RMSE of the estimated model parameters do not include the ability parameter and are shown for $N = 1000$ and $J = 40$ within studies 2 and 3, due to the number of plots. The convergence of the α, β and θ parameters are only shown within studies 2 and 3, for the same condition due to space limitations.

Table 4.1: Results presented within Dissertation for Study 2 and 3

		Test Length		
		$J = 20$	$J = 40$	$J = 80$
	$N = 500$	✓	-	-
Sample Size	$N = 1000$	-	✓	-
	$N = 2000$	-	-	✓

4.3.1 Study 1 Outcomes

This study was used to determine the ramifications of generating the data under a set of hyperparameters and to determine if the recovery of the model parameters would be affected by using different sets of hyperparameters during the estimation process. Table 4.2 shows the bias, standard deviation of the bias, RMSE, and standard deviation of the RMSE for the traditional 2PL IRT parameters (α, β, θ).

Table 4.2: IRT model parameters: Generating data through changing hyperparameters and estimating M2PLMix model parameters with hyperparameters set to $\lambda = 0.4$, $\omega = 2$, and $\kappa = 0.2$

Estimating Under $\lambda = 0.4$, $\omega = 2$, and $\kappa = 0.2$ Model N = 1000, J = 40									
		Normal				-			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
	α	0.04	0.05	0.1	0.16	-	-	-	-
Normal	β	-0.09	0.07	0.11	0.18	-	-	-	-
	θ	-0.04	0.19	0.28	0.13	-	-	-	-
		$\omega = 1.5$				$\omega = 2.5$			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
	α	0.02	0.05	0.1	0.15	0.08	0.04	0.12	0.16
ω	β	-0.1	0.07	0.12	0.18	-0.08	0.1	0.12	0.19
	θ	-0.04	0.21	0.29	0.15	-0.04	0.15	0.26	0.1
		$\kappa = 0.15$				$\kappa = 0.25$			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
	α	0.14	0.06	0.16	0.18	-0.02	0.08	0.12	0.19
κ	β	-0.08	0.13	0.14	0.21	-0.09	0.08	0.12	0.18
	θ	-0.04	0.16	0.28	0.12	-0.04	0.19	0.27	0.13
		$\lambda = 0.35$				$\lambda = 0.45$			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
	α	0.05	0.04	0.11	0.17	0.06	0.04	0.1	0.16
λ	β	-0.09	0.09	0.12	0.19	-0.08	0.07	0.11	0.18
	θ	-0.04	0.15	0.27	0.1	-0.04	0.19	0.28	0.13

Each variation of a hyperparameter shows its recovery of α, β, and θ in Table 4.2. Again, normal denotes when the hyperparameters were all unmodified (normal condition). When ω was modified (with respect to ± 0.5), the recovery of the model parameters were the same as if the data were generated with no modifications made to the ω (i.e., the normal case) except for one parameter. The exception was when $\omega = 2.5$, where the discrimination parameters were slightly underestimated compared to the normal case.

The λ hyperparameter also behaved similarly; parameter recovery was consistent, regardless of the variation in λ. This was not the case for κ, as the discrimination parameters were underestimated when $\kappa = 0.15$. The rationale for this result stems from κ's direct impact on the amount speededness that affects the probability of correctness in which more items will be answered correctly when $\kappa = 0.15$. Since the discrimination parameter was estimated with $\kappa = 0.2$, the discrimination parameter loses its power to discriminate between high and low ability examinees, because $\kappa = 0.2$ assumes more incorrect solutions. The variations of the κ parameter also had an impact on the difficulty parameter, increasing the standard deviation of the bias for all items. However, κ was not problematic in the recovery of the ability parameter.

Figures 4.1 - 4.6 display the RMSE and bias for each item's α and β parameter when the data were generated with modified hyperparameters $\omega = 2.5$, $\kappa = 0.25$, $\lambda = 0.45$ with the normal condition as the control. The bottom two panels display when the data were generated and estimated under the normal condition and the top panels shows when the data were generated with a modification and estimated under the normal case. The outcomes for when $\omega = 1.5$, $\kappa = 0.15$, $\lambda = 0.35$ are found in Appendix A.

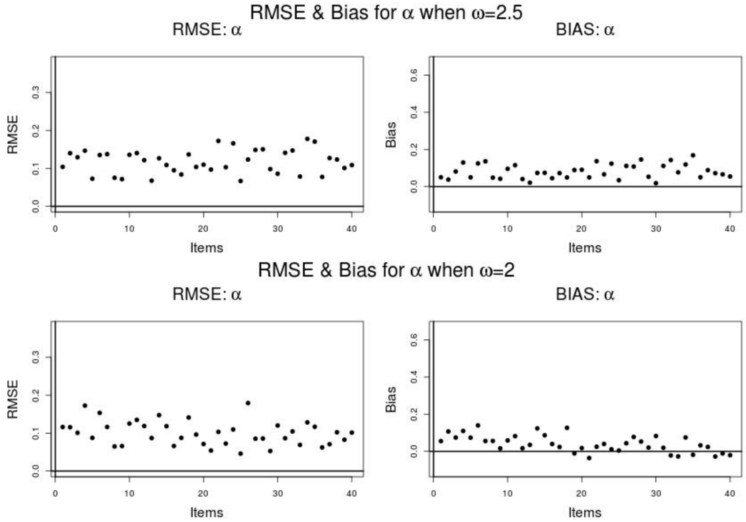

Figure 4.1: Bias and RMSE for α, the estimation model with $\lambda = 0.4$, $\omega = 2$, and $\kappa = 0.2$, data generation with $\omega = 2.5$ (top) and Normal (bottom) models, $N = 1000$, $J = 40$

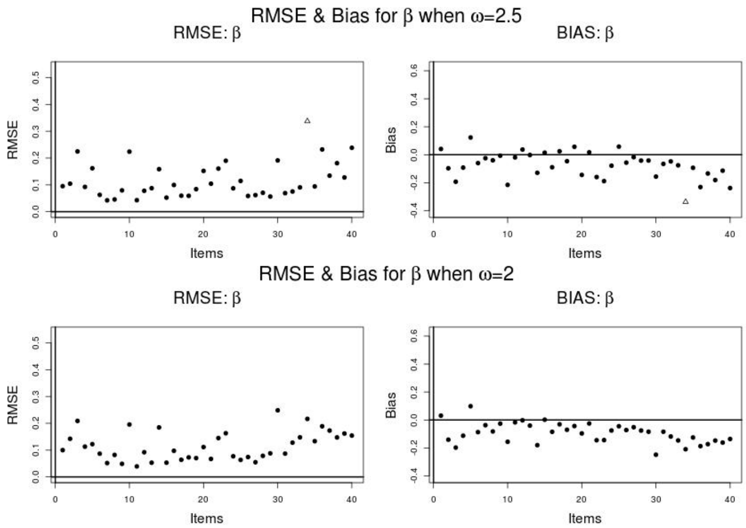

Figure 4.2: Bias and RMSE for β, the estimation model with $\lambda = 0.4$, $\omega = 2$, and $\kappa = 0.2$, data generation with $\omega = 2.5$ (top) and Normal (bottom) models, $N = 1000$, $J = 40$

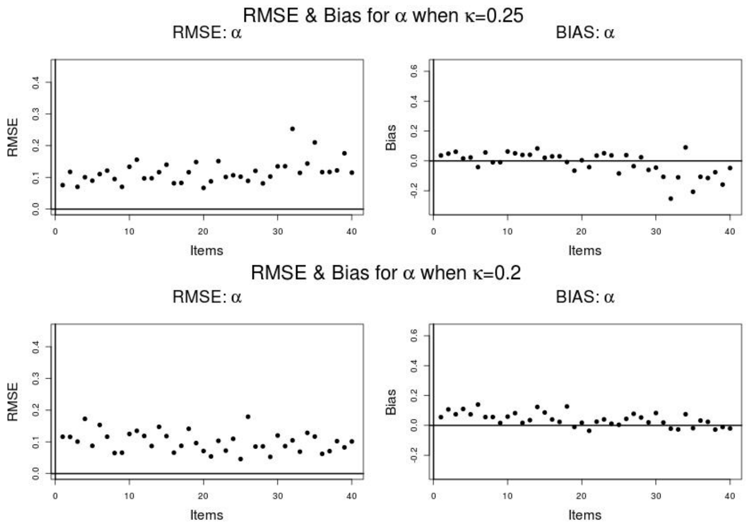

Figure 4.3: Bias and RMSE for α, the estimation model with $\lambda = 0.4$, $\omega = 2$, and $\kappa = 0.2$, data generation with $\kappa = 0.25$ (top) and Normal (bottom) models, $N = 1000$, $J = 40$

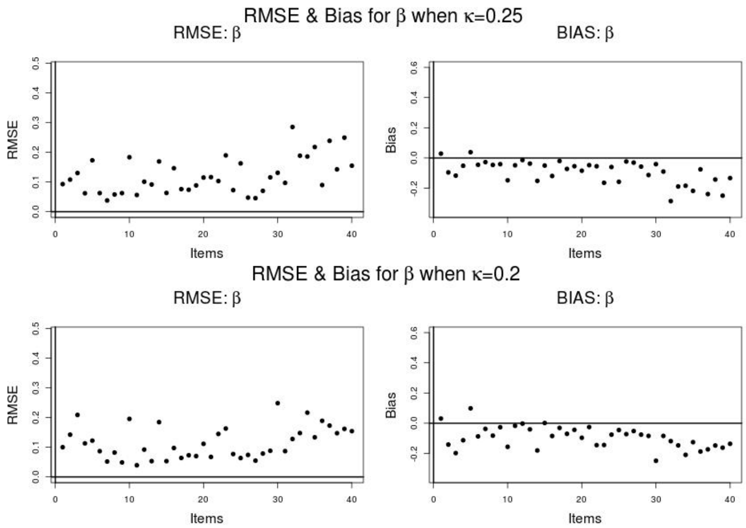

Figure 4.4: Bias and RMSE for β, the estimation model with $\lambda = 0.4$, $\omega = 2$, and $\kappa = 0.2$, data generation with $\kappa = 0.25$ (top) and Normal (bottom) models, $N = 1000$, $J = 40$

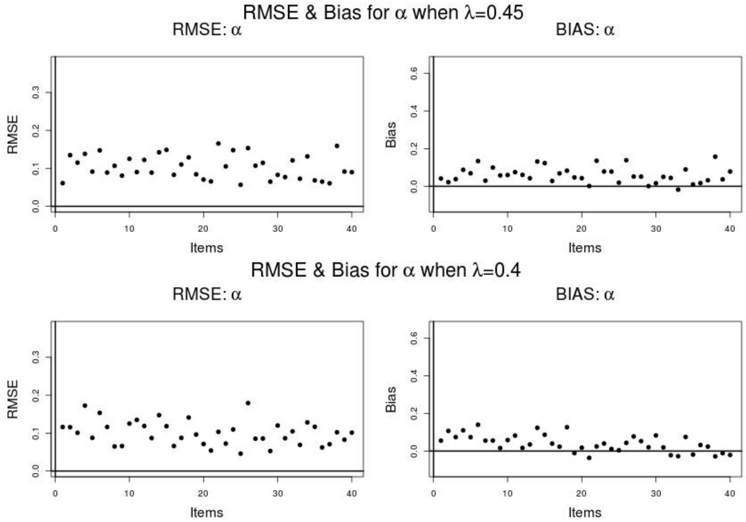

Figure 4.5: Bias and RMSE for α, the estimation model with $\lambda = 0.4$, $\omega = 2$, and $\kappa = 0.2$, data generation with $\lambda = 0.45$ (top) and Normal (bottom) models, $N = 1000$, $J = 40$

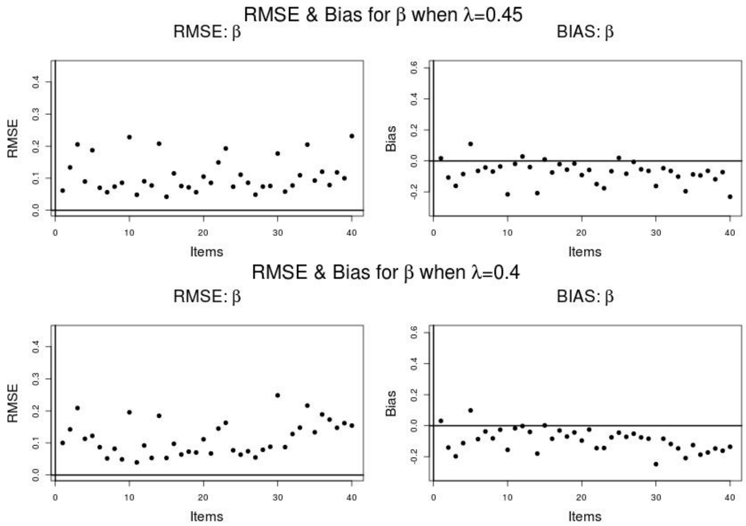

Figure 4.6: Bias and RMSE for β, the estimation model with $\lambda = 0.4$, $\omega = 2$, and $\kappa = 0.2$, data generation with $\lambda = 0.45$ (top) and Normal (bottom) models, $N = 1000$, $J = 40$

A gauge was created to determine if the bias or RMSE per item was in a reasonable range: if the absolute value of an item parameter's bias or RMSE was above an arbitrary number, then it was classified as an open triangle, otherwise it was classified as a closed circle. Throughout all four studies this gauge was arbitrarily set to 0.3 for the RMSE and bias for each estimated item parameter.

Figure 4.1 shows when $\omega = 2.5$, all discrimination parameters were underestimated. However, under the normal condition, all the discrimination parameters were overestimated with some exceptions at the end of the test. Figure 4.2 demonstrates that when $\omega = 2.5$, difficulty parameters tended to be overestimated and were better estimated than the discrimination parameters but the last few items showed large bias and RMSE. The results were very similar to those of the normal condition. Figure 4.3 depicts that when $\kappa = 0.25$, item discrimination parameters were overestimated towards the end of the test, whereas discrimination parameters for the normal condition tended to be underestimated in the beginning of the test. However, RMSE values were similar between the two conditions. Figure 4.4 depicts when $\kappa = 0.25$, difficulty parameters were overestimated towards the end of the test, matching the recovery of the M2PLMix model under normal conditions. Figure 4.5 reveals that when $\lambda = 0.45$, most item discrimination parameters were underestimated which was very similar to the normal condition. Figure 4.6 shows when $\lambda = 0.45$, difficulty parameters were overestimated towards the end of the test, matching the recovery of the M2PLMix model under normal conditions.

In general, the graphs indicate that modifying these hyperparameters causes slight over- or underestimation depending on parameter types. However, the recovery of these parameters only slightly deviated from the normal condition, meaning that hyperparameters modification did not have an crucial impact.

All five measures used to determine the accuracy of the inception of speededness are seen in Table 4.3. M_1 was used to determine the proportion of potential speededness locations that were correctly specified. The classification rates were similar for all 7 conditions, which was around 0.7. M_2 measures the proportion of those that accurately identified as speeded. M_2 ranged from 0.23 to 0.38 on average, which was a relatively poor result. This implies that the model identifies speeded examinees as not being speeded. This result is corroborated in studies 2 and 3. M_2 performs poorly especially when $\kappa = 0.15$ and $\omega = 2.5$. Setting $\kappa = 0.15$ appears to mask the effect of speededness, especially when compared to the normal condition, yielding the failure of M_2 to be able to correctly identify speeded examinees. Setting $\omega = 2.5$ impacts the convexity of the probability of not being speeded in such a way that less examinees were seen as speeded. That being said, if less people are specified as speeded during the estimation of δ then this would lower the classification rate of those that are speeded.

M_3 denotes the proportion that accurately identified those that were not speeded and had a value of 1 under all variations of the hyperparameters. This result implies that the M2PLMix was proficient in determining unspeeded examinees. For M_4 (relative bias) in absolute value, values of relative bias larger than $|0.2|$ were considered unacceptable. When $\omega = 1.5$, the M_4 (relative bias) was relatively large, implying that the less convex the baseline probability distribution of not being speeded, the more inaccurate the location of speededness. However it is important to be aware of the inherent bias found in this statistic, discussed previously. More specifically, when ω was small during generation of data the probability of being speeded occurs earlier. For other conditions, the M_4 statistic was close to or smaller than the normal condition. M_5 was used to gauge the proportion of speeded baseline examinees compared to the true proportion of speeded examinees. M_5 was fairly comparable throughout all variations of the hyperparameters. The difference was high because of the M2PLMix model's failure to correctly identify speeded examinees (which will be discussed further in studies 2 and 3).

It was important to note that these indices have not been vetted and were based on the nature of the inception of speededness. It was also important to note that there was no absolute way to determine when an examinee switches their test taking strategy, therefore it was expected for these rates to be imperfect. This study allows us to understand the effect of assuming known hyperparameters and the ramifications of generating the data with different hyperparameters.

Table 4.3: Location of Speededness: Generating data through changing hyperparameters and estimating M2PLMix model parameters with hyperparameters set to $\lambda = 0.4$, $\omega = 2$, and $\kappa = 0.2$ as the normal condition.

Estimating under $\lambda = 0.4$, $\omega = 2$, and $\kappa = 0.2$ for N = 1000, J = 40											
		Normal					-				
		M_1	M_2	M_3	M_4	M_5					
Normal	δ	0.7	0.3	1	-0.18	-0.21					
		$\omega = 1.5$					$\omega = 2.5$				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
ω	δ	0.69	0.38	1	-0.29	-0.2	0.7	0.25	1	-0.08	-0.23
		$\kappa = 0.15$					$\kappa = 0.25$				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
κ	δ	0.68	0.23	1	-0.12	-0.25	0.7	0.35	1	-0.12	-0.2
		$\lambda = 0.35$					$\lambda = 0.45$				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
λ	δ	0.69	0.37	1	-0.09	-0.2	0.72	0.34	1	-0.14	-0.18

M_1: correct specification of the location, M_2: correct specification for examinees, M_3: correct specification for nonspeeded examinees, M_4: relative bias, M_5: proportion difference of specified speeded between true and estimated.

4.3.2 Study 2 Outcomes

Data generated under M2PLMix Model when $\lambda = 0.2$

This section discusses the results of the 2PL and M2PLMix when data were generated under the M2PLMix model with the baseline probability of not being speeded $\lambda = 0.2$. Table 4.4) shows the parameter recovery results of the two models. When the sample size and the number of items were $N = 500$ and $J = 20$ respectively and the data were generated under the M2PLMix, the M2PLMix model performed better than the 2PL. The M2PLMix model resulted in lower RMSE and bias values than the 2PL in terms of α, β, and θ as expected.The RMSE of β was the smallest, and the RMSE of θ was the largest across all conditions, which is a typical observation in other IRT parameter recovery studies. When the items were ordered by difficulty level, the parameters were better estimated in both models, compared to the random order condition. This was natural because once an examinee became speeded, the items were already difficult. Therefore, it was expected for these students to underperform during the latter part of the assessment.

As the sample size ($N = 1000$, $N = 2000$) and the length of the test ($J = 40$, $J = 80$) increased, the M2PLMix produced better estimates for α, β and θ, while the 2PL did not. Though the number of unspeeded examinees increased (when the number of observations increases), the number of speeded students increased as well. Therefore, ignoring speededness was problematic no matter the sample size or the number of items. As the sample size and the number of items increased, the standard deviations of the bias and RMSE of the α, β, and θ parameters tended to decrease in the M2PLMix model, while the opposite pattern was observed in the 2PL model.

Table 4.4: IRT model parameters: Generated Model M2PLMix; Estimated Models the M2PLMix and 2PL with $\lambda = 0.2$.

Generating Model M2PLMix N = 500, J = 20 $\lambda = 0.2$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.16	0.2	0.24	0.17	-0.27	0.36	0.37	0.27
	β	-0.16	0.12	0.19	0.19	-0.25	0.22	0.28	0.23
	θ	-0.02	0.4	0.4	0.21	-0.02	0.5	0.44	0.25
Ordered	α	-0.14	0.12	0.2	0.16	-0.24	0.19	0.28	0.18
	β	-0.07	0.08	0.13	0.22	-0.1	0.11	0.16	0.19
	θ	-0.03	0.37	0.38	0.19	-0.02	0.41	0.39	0.22
Generating Model M2PLMix N = 1000, J = 40 $\lambda = 0.2$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.03	0.08	0.12	0.19	-0.28	0.39	0.38	0.29
	β	-0.07	0.08	0.11	0.17	-0.24	0.3	0.27	0.3
	θ	-0.03	0.17	0.28	0.12	-0.02	0.49	0.4	0.26
Ordered	α	-0.09	0.09	0.13	0.19	-0.21	0.19	0.25	0.22
	β	-0.05	0.03	0.08	0.15	-0.11	0.08	0.14	0.14
	θ	-0.03	0.19	0.28	0.12	-0.02	0.4	0.34	0.24
Generating Model M2PLMix N = 2000, J = 80 $\lambda = 0.2$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0	0.03	0.07	0.06	-0.29	0.44	0.41	0.35
	β	-0.03	0.04	0.06	0.06	-0.25	0.33	0.28	0.34
	θ	-0.02	0.1	0.2	0.15	-0.02	0.54	0.41	0.33
Ordered	α	-0.03	0.07	0.08	0.11	-0.2	0.21	0.25	0.21
	β	-0.01	0.03	0.05	0.06	-0.08	0.1	0.12	0.08
	θ	-0.02	0.11	0.19	0.15	-0.02	0.38	0.28	0.27

The measures used to determine the location of speededness can be seen in Table 4.5. M_1 across all conditions shows that the proportion of examinees that were classified correctly occurred at very similar rates in both models. However, since the 2PL classifies the entire test-taking population as not speeded, M_2 was 0, as expected. In contrast, M_2 increases as the number of observations (i.e., N & J) increased for the M2PLMix model. The measurement used to determine the proportion of properly classified nonspeeded examinees, M_3, were consistently 1 for the M2PLMix and 2PL models.

The relative bias and the difference between the true and estimated proportions of speeded examinees, M_4 and M_5, tended to decrease as the length of the test and sample size increased for the M2PLMix. On the other hand, these factor remain did not affect the results of M_4 and M_5 for the 2PL under all conditions. M_4 and M_5 tended to performed better under the M2PLMix model than the 2PL, regardless of item order. This tendency was more apparent with the random order condition than the ordered condition.

Table 4.5: Location of Speededness: Generated Model M2PLMix; Estimated Models the M2PLMix and 2PL with $\lambda = 0.2$.

Generating Model M2PLMix N = 500, J = 20 $\lambda = 0.2$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.59	0.02	1	-0.41	-0.4	0.59	0	1	-0.43	-0.41
Ordered	δ	0.61	0.01	1	-0.39	-0.39	0.61	0	1	-0.4	-0.39
Generating Model M2PLMix N = 1000, J = 40 $\lambda = 0.2$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.61	0.24	1	-0.22	-0.3	0.6	0	1	-0.37	-0.4
Ordered	δ	0.58	0.1	1	-0.25	-0.38	0.58	0	1	-0.37	-0.42
Generating Model M2PLMix N = 2000, J = 80 $\lambda = 0.2$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.6	0.57	1	-0.05	-0.17	0.59	0	1	-0.41	-0.41
Ordered	δ	0.6	0.24	1	-0.13	-0.3	0.6	0	1	-0.35	-0.4

M_1: correct specification of the location, M_2: correct specification for speeded examinees, M_3: correct specification for nonspeeded examinees, M_4: relative bias, M_5: proportion difference of specified speeded between true and estimated.

Data generated under M2PLMix Model when $\lambda = 0.4$

This section provides the results in which the data were generated under the M2PLMix model with the baseline probability of not being speeded was 0.4. Model parameters were estimated under the 2PL and M2PLMix models. As the baseline probability of not being speeded (λ) increases, the parameters become better estimated in both models, which can be seen in Table 4.6. This was caused by less examinees being affected by speededness. Though both models performed better when the λ increases, the M2PLMix model recovered α, β, and θ better than the 2PL model.

Item ordering caused the difficulty parameter to be better estimated no matter the sample size, number of items, and the baseline probability of not being speeded for both models. The standard deviation of the RMSE and the bias decreased as sample size and number of items increased for the M2PLMix model but not for the 2PL model. In addition, the standard deviation of the bias was consistently larger for θ across all conditions for both models due to the number of parameters. The standard deviation of bias for α, β, and θ regarding the M2PLMix model was always lower than that of the 2PL model.

Table 4.6: IRT model parameters: Generated Model M2PLMix; Estimated Models the M2PLMix and 2PL with $\lambda = 0.4$.

Generating Model M2PLMix N = 500, J = 20 $\lambda = 0.4$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.11	0.11	0.19	0.2	-0.19	0.22	0.27	0.19
	β	-0.13	0.12	0.18	0.2	-0.19	0.2	0.24	0.19
	θ	-0.02	0.34	0.38	0.16	-0.02	0.44	0.41	0.21
Ordered	α	-0.1	0.09	0.16	0.21	-0.18	0.14	0.22	0.19
	β	-0.06	0.04	0.11	0.2	-0.08	0.07	0.12	0.17
	θ	-0.02	0.3	0.35	0.14	-0.02	0.32	0.35	0.16
Generating Model M2PLMix N = 1000, J = 40 $\lambda = 0.4$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.04	0.07	0.11	0.18	-0.22	0.31	0.31	0.24
	β	-0.06	0.07	0.1	0.15	-0.21	0.25	0.24	0.27
	θ	-0.03	0.18	0.27	0.12	-0.02	0.49	0.39	0.29
Ordered	α	-0.06	0.06	0.11	0.17	-0.15	0.13	0.19	0.19
	β	-0.03	0.03	0.08	0.1	-0.06	0.07	0.11	0.1
	θ	-0.03	0.18	0.26	0.14	-0.02	0.32	0.29	0.2
Generating Model M2PLMix N = 2000, J = 80 $\lambda = 0.4$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.01	0.04	0.07	0.06	-0.2	0.31	0.3	0.25
	β	-0.04	0.04	0.06	0.07	-0.19	0.26	0.23	0.27
	θ	-0.03	0.09	0.19	0.14	-0.02	0.47	0.35	0.3
Ordered	α	-0.04	0.05	0.08	0.1	-0.18	0.16	0.22	0.18
	β	-0.02	0.02	0.06	0.04	-0.07	0.08	0.11	0.06
	θ	-0.03	0.12	0.2	0.16	-0.02	0.37	0.27	0.27

Figures 4.7 - 4.10 display the bias and RMSE for α and β where the sample size was $N = 1000$ and the number of items was $J = 40$ for random and ordered item

difficulty with respect to both M2PLMix and 2PL models. The bias and RMSE for each item discrimination parameter with random difficulty ordering tended to be estimated well during the beginning of the assessment and increased towards the end of the test for both models, as seen in Figure 4.7. However, the 2PL model performs worse compared to the M2PLMix towards the end of the test. As mentioned in the first study, if parameters are poorly recovered (bias or RMSE are greater than 0.3), triangles are used to display their values. There were no poorly recovered items under the M2PLMix model, but 43% of the item discrimination parameters were poorly calibrated within the 2PL model. When item difficulty was ordered, modeling speededness was still found to impact the recovery of α, however at a lower rate, and only 13% of item discrimination parameters were poorly calibrated under the 2PL model.

Figure 4.9 displays a similar pattern in the recovery of difficulty parameters, in that poorly calibrated items primarily occurred towards the end of an assessment for both models. But items were only classified as poorly calibrated only under the 2PL model. In fact, no items were considered to be poorly calibrated when the item difficulty was ordered, as found in Figure 4.10. Under his ordered condition, it appeared that item difficulty parameters were estimated well during beginning of the test, then become underestimated, and then become estimated better again. This pattern becomes stronger under the 2PL model. This result was probably due to the fact that medium difficulty items were affected by speededness, whereas truly difficult items were more impacted by the difficulty of the items.

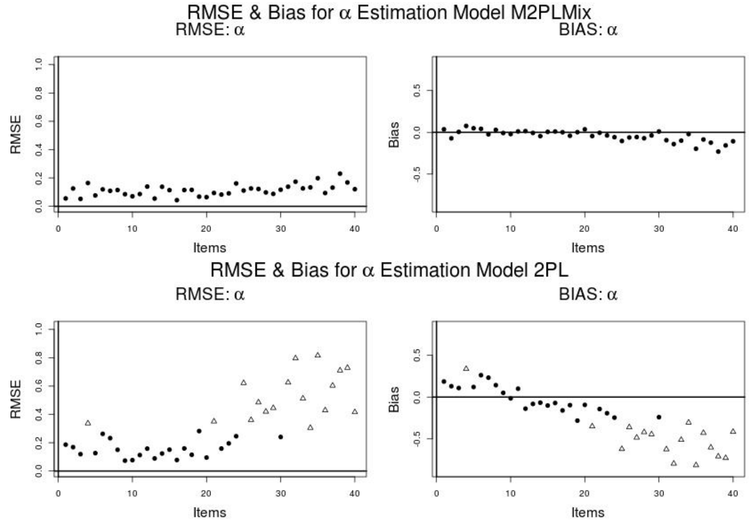

Figure 4.7: Bias and RMSE for α, data generation under M2PLMix model, estimation were the M2PLMix and 2PL models, N =1000, J = 40 $\lambda = 0.4$, Item Ordering was Random

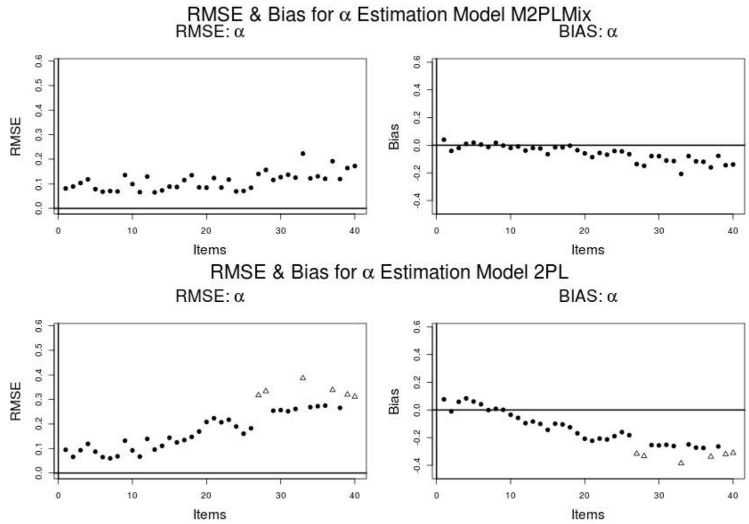

Figure 4.8: Bias and RMSE for α, data generation under M2PLMix model, estimation were the M2PLMix and 2PL models, N =1000, J = 40 $\lambda = 0.4$, Item Ordering was Ordered

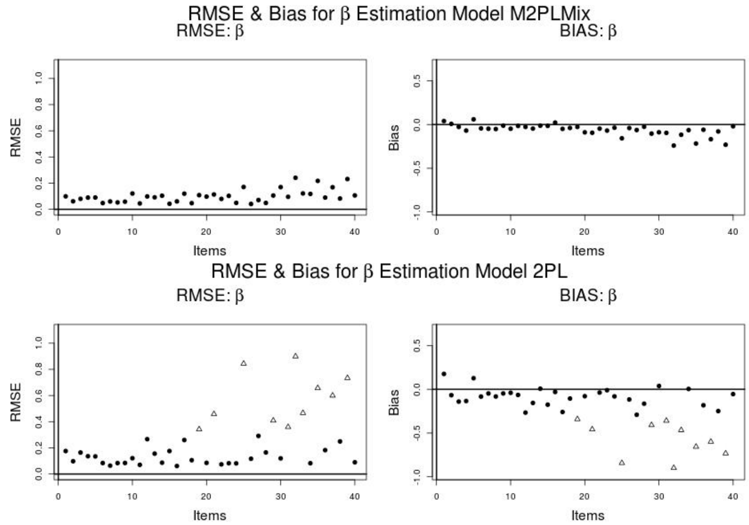

Figure 4.9: Bias and RMSE for β, data generation under M2PLMix model, estimation were the M2PLMix and 2PL models, N =1000, J = 40 $\lambda = 0.4$, Item Ordering was Random

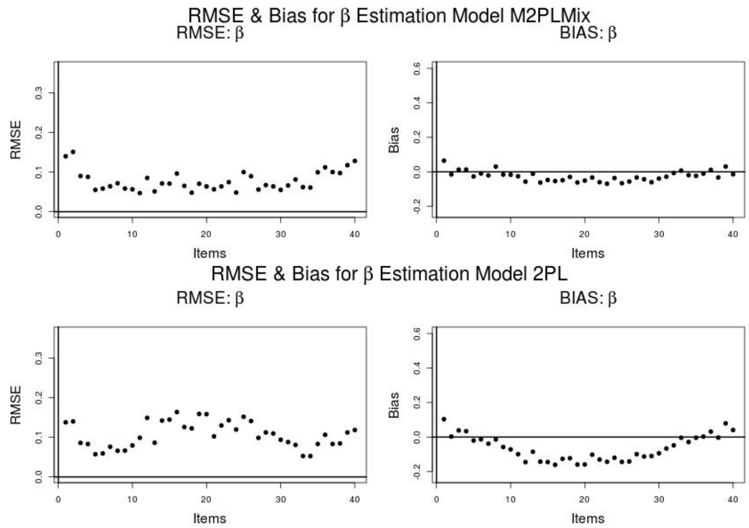

Figure 4.10: Bias and RMSE for β, data generation under M2PLMix model, estimation were the M2PLMix and 2PL models, N =1000, J = 40 $\lambda = 0.4$, Item Ordering was Ordered

Table 4.7 shows the measures of inception of speededness when $\lambda = 0.4$. In the same manner as Table 4.5, the M_1 was similar between both models, and M_3 was always 1 for both models. Compared to Table 4.5, M_1 increased because the number of people affected by speededness decreased. As the sample size and the number of items increase, M_2, M_4, and M_5 are better estimated under the M2PLMix model.

Table 4.8 denotes the convergence rate of α, β, and θ. Each numerical value represents the average of parameters that did not converge. This table shows that when the M2PLMix model was used to estimate the model parameters, almost all parameters converge using the Gelman-Rubin criteria under most conditions. In a similar respect, under the 2PL model used to estimate the model parameters, all the model parameters were able to converge with respect to the Gelman-Rubin criteria. The parameters that did not converged was included within the results (recovery of parameters and model fit) and did not affect the general pattern of the results.

Table 4.7: Location of Speededness: Generated Model M2PLMix; Estimated Models the M2PLMix and 2PL with $\lambda = 0.4$.

Generating Model M2PLMix N = 500, J = 20 $\lambda = 0.4$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.7	0.01	1	-0.27	-0.29	0.7	0	1	-0.27	-0.3
Ordered	δ	0.67	0	1	-0.27	-0.33	0.67	0	1	-0.27	-0.33
Generating Model M2PLMix N = 1000, J = 40 $\lambda = 0.4$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.71	0.25	1	-0.17	-0.22	0.71	0	1	-0.31	-0.29
Ordered	δ	0.7	0.08	1	-0.22	-0.27	0.7	0	1	-0.28	-0.3
Generating Model M2PLMix N = 2000, J = 80 $\lambda = 0.4$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.72	0.55	1	-0.04	-0.13	0.71	0	1	-0.27	-0.29
Ordered	δ	0.69	0.24	1	-0.16	-0.24	0.69	0	1	-0.35	-0.31

M_1: correct specification of the location, M_2: correct specification for speeded examinees, M_3: correct specification for nonspeeded examinees, M_4: relative bias, M_5: proportion difference of specified speeded between true and estimated.

Table 4.8: Percentage of Non-Convergence of IRT Parameters, α, β, and θ for M2PLMix and 2PL models, generated under M2PLMix model, using the Gelman-Rubin Criteria

$N = 500,\ J = 20,\ \lambda = 0.4$			
		M2PLMix R	2PL R
Random	α	0	0
	β	0	0
	θ	0.008	0
Ordered	α	0	0
	β	0.05	0
	θ	0.012	0
$N = 1000,\ J = 40,\ \lambda = 0.4$			
		M2PLMix R	2PL R
Random	α	0	0
	β	0.1	0
	θ	0.01	0
Ordered	α	0	0
	β	0	0
	θ	0.022	0
$N = 2000,\ J = 80,\ \lambda = 0.4$			
		M2PLMix R	2PL R
Random	α	0	0
	β	0	0
	θ	0.005	0
Ordered	α	0	0
	β	0	0
	θ	0.032	0

Data generated under M2PLMix Model when $\lambda = 0.6$

This section provides the results in which the data were generated under the M2PLMix model with the baseline probability of not being speeded $\lambda = 0.6$, and model parameters were estimated under the 2PL and M2PLMix models. Even though the baseline probability of not being speeded was $\lambda = 0.6$, hence less examinees were affected by speededness, the 2PL model was still unable to recover the item parameters, compared to the M2PLMix model, as seen in Table 4.9. Furthermore, both models improved in terms of recovery of item parameters and ability level due to a lower number of speeded examinees. As with the other λ levels, model parameters were better estimated when item difficulty was ordered.

The standard deviation for bias and RMSE decreased as the sample size and the number of items increased. Of all three baseline probabilities of not being speeded, $\lambda = 0.6$ has the lowest RMSE and bias due to the low number of individuals that were impacted by speededness.

Table 4.9: IRT model parameters: Generated Model M2PLMix; Estimated Models the M2PLMix and 2PL with $\lambda = 0.6$.

Generating Model M2PLMix N = 500, J = 20 $\lambda = 0.6$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.11	0.12	0.18	0.17	-0.18	0.18	0.24	0.18
	β	-0.09	0.11	0.15	0.18	-0.13	0.17	0.19	0.2
	θ	-0.03	0.31	0.35	0.15	-0.02	0.36	0.37	0.17
Ordered	α	-0.04	0.06	0.14	0.19	-0.1	0.1	0.17	0.18
	β	-0.04	0.06	0.11	0.26	-0.05	0.06	0.11	0.2
	θ	-0.02	0.29	0.34	0.15	-0.02	0.31	0.35	0.16
Generating Model M2PLMix N = 1000, J = 40 $\lambda = 0.6$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.01	0.05	0.1	0.16	-0.1	0.2	0.2	0.21
	β	-0.05	0.04	0.09	0.13	-0.16	0.17	0.18	0.21
	θ	-0.02	0.16	0.26	0.1	-0.02	0.4	0.33	0.25
Ordered	α	-0.04	0.06	0.1	0.16	-0.11	0.11	0.16	0.19
	β	-0.03	0.03	0.08	0.15	-0.06	0.07	0.1	0.15
	θ	-0.03	0.17	0.26	0.12	-0.02	0.33	0.29	0.22
Generating Model M2PLMix N = 2000, J = 80 $\lambda = 0.6$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.01	0.03	0.07	0.06	-0.15	0.23	0.23	0.21
	β	-0.03	0.03	0.06	0.07	-0.16	0.21	0.18	0.23
	θ	-0.03	0.09	0.19	0.15	-0.02	0.41	0.3	0.28
Ordered	α	-0.02	0.04	0.07	0.05	-0.1	0.11	0.14	0.18
	β	-0.03	0.03	0.06	0.06	-0.06	0.06	0.08	0.07
	θ	-0.03	0.11	0.19	0.15	-0.02	0.31	0.24	0.23

M_1 in Table 4.10 was similar for both the M2PLMix and the 2PL models. However, the correct specification for when $\lambda = 0.6$ was greater than that of the other levels because the number of speeded examinees was smaller with this set. Similar patterns for the other measures were found as the other levels of λ.

Table 4.10: Location of Speededness: Generated Model M2PLMix; Estimated Models the M2PLMix and 2PL with $\lambda = 0.6$.

Generating Model M2PLMix N = 500, J = 20 $\lambda = 0.6$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.8	0.01	1	-0.19	-0.2	0.8	0	1	-0.2	-0.2
Ordered	δ	0.79	0	1	-0.19	-0.21	0.79	0	1	-0.19	-0.21
Generating Model M2PLMix N = 1000, J = 40 $\lambda = 0.6$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.81	0.27	1	-0.08	-0.14	0.8	0	1	-0.16	-0.2
Ordered	δ	0.79	0.1	1	-0.13	-0.18	0.8	0	1	-0.19	-0.2
Generating Model M2PLMix N = 2000, J = 80 $\lambda = 0.6$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.81	0.59	1	-0.03	-0.08	0.8	0	1	-0.21	-0.2
Ordered	δ	0.78	0.23	1	-0.1	-0.17	0.78	0	1	-0.2	-0.22

M_1: correct specification of the location, M_2: correct specification for speeded examinees, M_3: correct specification for nonspeeded examinees, M_4: relative bias, M_5: proportion difference of specified speeded between true and estimated.

Table 4.11 shows the model fit results in which the item ordering difficulty was random (ordered is found in Appendix B). The M2PLMix model does not fit the data well when the sample size and the number of items were small (i.e., $N = 500$ and $J = 20$) across all baselines of probability of not being speeded. However as the amount of parameters increase the model starts to fit the data well, according to the $-2LL$ fit. Even though the items were better estimated with the M2PLMix than the 2PL, when the sample size was small, the model fit of the 2PL was slightly better than the M2PLMix which was due to the fact that the M2PLMix model estimates 500 more parameters (δ) than the 2PL. This was not the case for the other 5 measures of model fit. The other measures indicated that the 2PL fitted the data better than the M2PLMix with a few exceptions which was caused by the number of parameters the M2PLMix model was estimating. The number of parameters does not have an impact on $-2LL$ as much as the others measures in terms of the number of parameters found in the M2PLMix model compared to 2PL model.

Table 4.11: Model Fit Data Generation under M2PLMix Model; $\lambda = 0.2$, $\lambda = 0.4$, $\lambda = 0.6$, Item Difficulty Ordering Random

Model Fit Data Generation under the M2PLMix Model						
		Estimating Model M2PLMix and 2PL				
M2PLMix Model						
		$-2LL$	AIC	AIC_c	BIC	$Adj.BIC$
$\lambda = 0.2$	$N = 500, J = 20$	8710.53	10790.53	6788.17	15173.73	11872.7
	$N = 1000, J = 40$	34186.4	38346.4	30338.11	48554.53	41948.33
	$N = 2000, J = 80$	130134.6	138454.6	122434.47	161754.36	148537.81
$\lambda = 0.4$	$N = 500, J = 20$	8815.63	10895.63	6893.26	15278.82	11977.8
	$N = 1000, J = 40$	34657.43	38817.43	30809.15	49025.57	42419.37
	$N = 2000, J = 80$	135331.38	143651.38	127631.25	166951.14	153734.59
$\lambda = 0.6$	$N = 500, J = 20$	8792.62	10872.62	6870.25	15255.81	11954.79
	$N = 1000, J = 40$	35447.63	39607.63	31599.34	49815.76	43209.56
	$N = 2000, J = 80$	138533.06	146853.06	130832.93	170152.82	156936.27
2PL Model						
		$-2LL$	AIC	AIC_c	BIC	$Adj.BIC$
$\lambda = 0.2$	$N = 500, J = 20$	8474.07	9554.07	4696.66	11829.96	10115.96
	$N = 1000, J = 40$	34745.77	36905.77	8079.1	42206.14	38776
	$N = 2000, J = 80$	140560.41	144880.41	86895.81	156978.36	150115.92
$\lambda = 0.4$	$N = 500, J = 20$	8646.19	9726.19	4524.54	12002.08	10288.09
	$N = 1000, J = 40$	35189.56	37349.56	8522.89	42649.94	39219.8
	$N = 2000, J = 80$	143365.79	147685.79	89701.19	159783.74	152921.3
$\lambda = 0.6$	$N = 500, J = 20$	8681.78	9761.78	4488.95	12037.67	10323.68
	$N = 1000, J = 40$	36137.36	38297.36	9470.69	43597.74	40167.6
	$N = 2000, J = 80$	144807.11	149127.11	91142.52	161225.06	154362.63

Data generated under the 2PL Model

No Speededness Present

This section provides the results when the data were generated under the 2PL model in which speededness was not present. Model parameters were estimated under the 2PL and M2PLMix models. In general, when the data were generated under the 2PL model and the sample size was small, the M2PLMix model did not recover the IRT model parameters as well as the 2PL (see Table 4.12). The M2PLMix model assumes that examinees with low ability were seen as speeded. The low RMSE and bias for the estimates of the parameters under the 2PL model also indicates that the proposed algorithm works well in estimating model parameters. On the other hand, the M2PLMix model only works well in the absence of speededness if the sample size and item length were large. Additionally, the model parameters were better estimated when the items were ordered based on difficulty, similar to when the data was generated under the M2PLMix model.

Table 4.12: IRT model parameters: Generated Model 2PL; Estimated Models the M2PLMix and 2PL, no speededness.

Generating Model 2PL N = 500, J = 20									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.14	0.13	0.2	0.13	-0.02	0.06	0.14	0.2
	β	0.11	0.25	0.23	0.42	-0.02	0.04	0.11	0.19
	θ	-0.02	0.23	0.34	0.1	-0.02	0.2	0.32	0.09
Ordered	α	0.09	0.06	0.17	0.17	-0.02	0.05	0.14	0.19
	β	0.03	0.08	0.13	0.21	-0.02	0.03	0.11	0.18
	θ	-0.02	0.22	0.33	0.1	-0.02	0.19	0.32	0.09
Generating Model 2PL N = 1000, J = 40									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.11	0.1	0.15	0.18	-0.02	0.03	0.09	0.17
	β	0.03	0.2	0.15	0.23	-0.03	0.02	0.08	0.11
	θ	-0.02	0.12	0.25	0.09	-0.02	0.12	0.23	0.08
Ordered	α	0.08	0.05	0.12	0.19	-0.02	0.04	0.1	0.17
	β	-0.02	0.07	0.09	0.16	-0.03	0.02	0.07	0.11
	θ	-0.02	0.12	0.24	0.1	-0.03	0.12	0.23	0.09
Generating Model 2PL N = 2000, J = 80									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.04	0.04	0.08	0.07	-0.01	0.02	0.07	0.06
	β	-0.01	0.06	0.06	0.11	-0.02	0.02	0.05	0.05
	θ	-0.03	0.06	0.17	0.13	-0.02	0.08	0.17	0.14
Ordered	α	0.04	0.04	0.07	0.06	-0.02	0.02	0.07	0.05
	β	-0.03	0.03	0.06	0.09	-0.02	0.02	0.05	0
	θ	-0.03	0.07	0.17	0.13	-0.02	0.08	0.17	0.14

The recovery for the item parameters estimated under the 2PL and M2PLMix models when data were generated under the 2PL models are found in figures 4.11 - 4.14. These figures show that the 2PL model was able to recover item parameters well when speededness is not present. In contrast, the discrimination and the difficulty parameters were not recovered well towards the end of an assessment for the M2PLMix model. The M2PLMix model improved in terms of estimation of the item parameters towards the end of the assessment when item difficulty was ordered. The estimation of item parameters are better estimated when item difficulty is ordered because

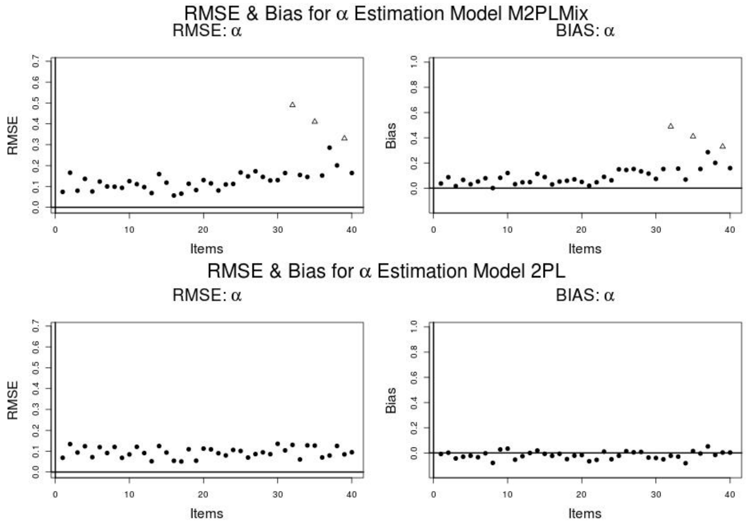

Figure 4.11: Bias and RMSE for α, data generation under 2PL model, estimation were the M2PLMix and 2PL models, N =1000, J = 40 , Item Ordering was Random

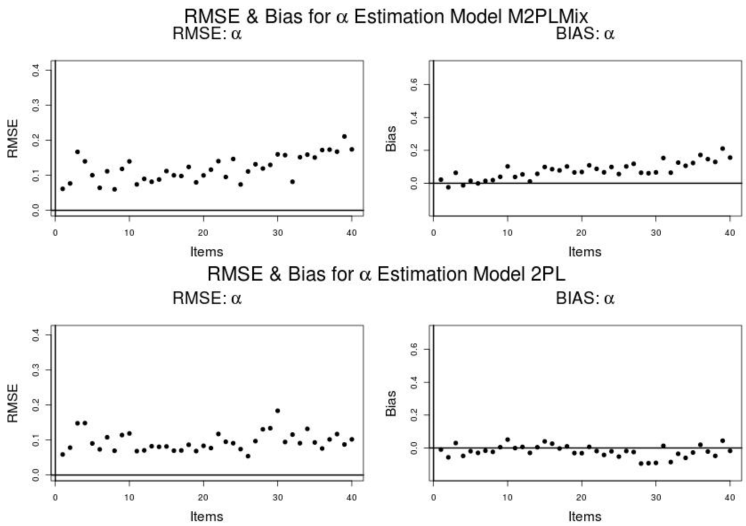

Figure 4.12: Bias and RMSE for α, data generation under 2PL model, estimation were the M2PLMix and 2PL models, N =1000, J = 40 , Item Ordering was Ordered

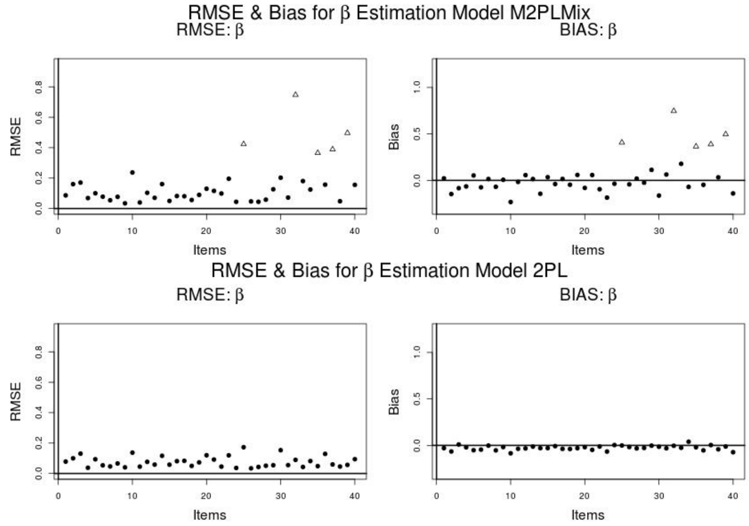

Figure 4.13: Bias and RMSE for β, data generation under 2PL model, estimation were the M2PLMix and 2PL models, N =1000, J = 40 , Item Ordering was Random

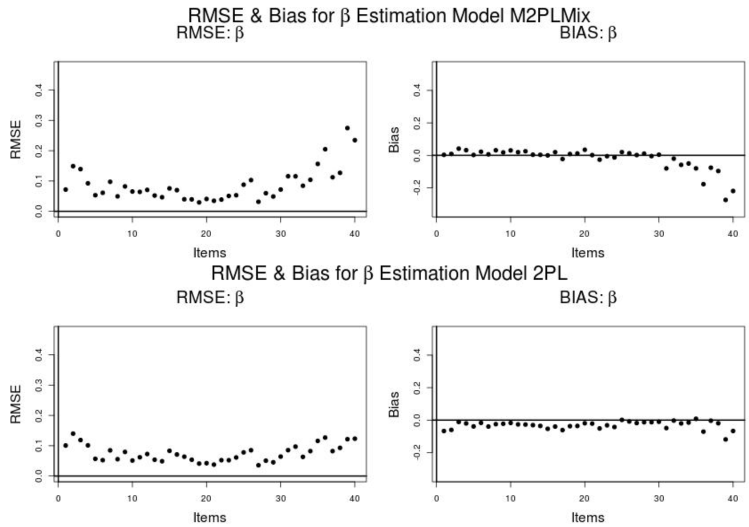

Figure 4.14: Bias and RMSE for β, data generation under 2PL model, estimation were the M2PLMix and 2PL models, N =1000, J = 40 , Item Ordering was Ordered

Table 4.13 shows the measures used to estimate the inception of speededness when the data were generated under the 2PL model. Since no speededness was present, M_1 and M_3, were 1 whereas M_4 and M_5 were 0. M_2 should not be applicable because there were no true speeded examinees in the data set. However, when the sample size was small and the item difficulty was ordered there were slight deviations based on the sample size and the length of the test.

Table 4.13: Location of Speededness: Generated Model 2PL; Estimated Models the M2PLMix and 2PL.

Generating Model 2PL N = 500, J = 20											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	1	NA	1	0	0	1	NA	1	0	0
Ordered	δ	0.99	NA	0.99	0.01	0.01	1	NA	1	0	0
Generating Model 2PL N = 1000, J = 40											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	1	NA	1	0	0	1	NA	1	0	0
Ordered	δ	1	NA	1	0	0	1	NA	1	0	0
Generating Model 2PL N = 2000, J = 80											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	1	NA	1	0	0	1	NA	1	0	0
Ordered	δ	1	NA	1	0	0	1	NA	1	0	0

M_1: correct specification of the location, M_2: correct specification for speeded examinees, M_3: correct specification for nonspeeded examinees, M_4: relative bias, M_5: proportion difference of specified speeded between true and estimated.

Table 4.14 shows the model fit when the data were generated under the 2PL model and estimated under the M2PLMix and 2PL models for randomly ordered items. The model fit indices indicate that the data fits the 2PL model better than M2PLMix model under all conditions. This was caused by the data generated under the 2PL model and N more parameters to be estimated in the M2PLMix. However, as the sample size and length of test increase the M2PLMix and 2PL model became comparable for the $-2LL$ and AIC due to the number of parameters. This result was also produced when the items were ordered by increasing difficulty.

Table 4.14: Model Fit Data Generation under 2PL Model

Model Fit Data Generation under the 2PL Model						
M2PLMix Model						
		$-2LL$	AIC	AIC_c	BIC	$Adj.BIC$
No Speededness	$N = 500, J = 20$	9128.81	11208.81	7206.45	15592.01	12290.98
	$N = 1000, J = 40$	36632.93	40792.93	32784.64	51001.06	44394.87
	$N = 2000, J = 80$	148022.74	156342.74	140322.61	179642.5	166425.95
2PL Model						
		$-2LL$	AIC	AIC_c	BIC	$Adj.BIC$
No Speededness	$N = 500, J = 20$	8906.36	9986.36	4264.37	12262.25	10548.26
	$N = 1000, J = 40$	36440.12	38600.12	9773.45	43900.49	40470.35
	$N = 2000, J = 80$	147952.81	152272.81	94288.21	164370.76	157508.32

Table 4.15 displays the convergence rate of α, β, and θ. Each numerical value represents the average of parameters that did not converge. This table shows that under most conditions when the M2PLMix model was used to estimate the model parameters, almost all parameters converged using the Gelman-Rubin criteria. However, when the sample size and number of items were $N = 2000$ and $J = 80$, respectively,all parameters converged. Under the 2PL model, all the model parameters were able to converge with respect to the Gelman-Rubin criteria.

Table 4.15: Percentage of Non-Convergence of IRT Parameters, α, β, and θ for M2PLMix and 2PL models, generated under 2PL model, using the Gelman-Rubin Criteria

$N = 500,\ J = 20,\ \lambda = 0.4$			
		M2PLMix R	2PL R
Random	α	0	0
	β	0.05	0
	θ	0.012	0
Ordered	α	0	0
	β	0	0
	θ	0.016	0
$N = 1000,\ J = 40,\ \lambda = 0.4$			
		M2PLMix R	2PL R
Random	α	0	0
	β	0	0
	θ	0.001	0
Ordered	α	0	0
	β	0	0
	θ	0.011	0
$N = 2000,\ J = 80,\ \lambda = 0.4$			
		M2PLMix R	2PL R
Random	α	0	0
	β	0	0
	θ	0	0
Ordered	α	0	0
	β	0	0
	θ	0	0

4.3.3 Study 3 Outcomes

Data generated under M2PLMix Model where $\lambda = 0.2$

The following section discusses the results for when the data were generated under the M2PLMix model with the baseline probability of not being speeded $\lambda = 0.2$, and model parameters were estimated under the 2PLMix and M2PLMix models. Under this simulation, the 2PLMix and M2PLMix models provided comparable results except for a few conditions (large sample size and number of items) found in Table 4.16. The RMSE for the difficulty parameter estimated under the M2PLMix model was smaller than that of the 2PLMix model when the sample size was $N = 1000$ and $N = 2000$ and the number of items was $J = 40$ and $J = 80$, respectively, independent of item ordering. In contrast, the bias for the difficulty parameter was consistently smaller under the 2PLMix model when compared to the M2PLMix model. The bias of for the discrimination parameter was smaller under the M2PLMix when the sample size was $N = 1000$ and $N = 2000$ and the number of items was $J = 40$ and $J = 80$, except when item difficulty was ordered with $N = 1000$ and $J = 40$. The M2PLMix model also recovered the difficulty parameter better than the 2PLMix model when the sample size and test length was $N = 500$ and $J = 20$, respectively, but only when the item difficulty was ordered.

When the sample size and length of test increased, the RMSE, RMSE SD, bias, and bias SD decreased for α, β, and θ parameters. This pattern occurred when these parameters were estimated by the M2PLMix, except for bias of θ and RMSE SD of θ and α. This trend also occurred when the model parameters were estimated under the 2PLMix, except for bias of θ. These exceptions were due to the randomness of the data and did not show on any clear pattern regarding simulation factors. .

Table 4.16: IRT model parameters: Generated Model M2PLMix; Estimated Models the M2PLMix and 2PLMix with $\lambda = 0.2$.

Generating Model M2PLMix N = 500, J = 20 $\lambda = 0.2$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.16	0.2	0.24	0.17	-0.01	0.07	0.15	0.19
	β	-0.16	0.12	0.19	0.19	0.09	0.15	0.17	0.18
	θ	-0.02	0.4	0.4	0.21	-0.03	0.35	0.4	0.18
Ordered	α	-0.14	0.12	0.2	0.16	-0.02	0.05	0.15	0.24
	β	-0.07	0.08	0.13	0.22	0.11	0.12	0.16	0.21
	θ	-0.03	0.37	0.38	0.19	-0.02	0.35	0.39	0.18
Generating Model M2PLMix N = 1000, J = 40 $\lambda = 0.2$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.03	0.08	0.12	0.19	0.05	0.04	0.11	0.18
	β	-0.07	0.08	0.11	0.17	0.04	0.13	0.13	0.19
	θ	-0.03	0.17	0.28	0.12	-0.03	0.17	0.28	0.12
Ordered	α	-0.09	0.09	0.13	0.19	0.04	0.04	0.11	0.18
	β	-0.05	0.03	0.08	0.15	0.03	0.05	0.09	0.16
	θ	-0.03	0.19	0.28	0.12	-0.03	0.17	0.28	0.11
Generating Model M2PLMix N = 2000, J = 80 $\lambda = 0.2$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0	0.03	0.07	0.06	0.04	0.02	0.07	0.03
	β	-0.03	0.04	0.06	0.06	0	0.07	0.07	0.1
	θ	-0.02	0.1	0.2	0.15	-0.03	0.09	0.2	0.14
Ordered	α	-0.03	0.07	0.08	0.11	0.05	0.03	0.08	0.06
	β	-0.01	0.03	0.05	0.06	0	0.04	0.06	0.09
	θ	-0.02	0.11	0.19	0.15	-0.03	0.1	0.2	0.14

Table 4.17 shows the accuracy of $\hat{\delta}$ with respect to the M2PLMix and 2PLMix models when the data were generated under the M2PLMix model. The 2PLMix model cannot identify unspeeded individuals (M_3) when the sample size and number of items was small, but the M2PLMix was able to identity unspeeded examinees well. The accuracy of determining the exact location of speededness, M_1, was better estimated under the M2PLMix when the sample size and number of items were $N = 500, 1000$ and $J = 20, 40$ respectively.

The M2PLMix required $\hat{\theta}$ to estimate $\hat{\delta}$ well, whereas the 2PLMix did not require any information about $\hat{\theta}$. Further, the probability of becoming speeded on any item under the M2PLMix model required the ability of the examinee.

However, since the ability parameter was being estimated simultaneously with the location of the inception of speededness, δ was not recovered well under the M2PLMix model. In contrast, the probability of becoming speeded on any item under the 2PLMix model was the same for all examinees and did not use other information. Therefore, the M2PLMix model was unable to label speeded examinees as well as the 2PLMix, which was seen in M_2. Also, as the number of observations increases, M_4 and M_5 favor the 2PLMix over the M2PLMix. The results showed that the M2PLMix was able to better identify unspeeded examinees but the 2PLMix correctly labeled speeded examinees better than the M2PLMix when the baseline probability of not being speeded was $\lambda = 0.2$.

Table 4.17: Location of Speededness: Generated Model M2PLMix; Estimated Models the M2PLMix and 2PLMix with $\lambda = 0.2$.

Generating Model M2PLMix N = 500, J = 20 $\lambda = 0.2$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.59	0.02	1	-0.41	-0.4	0.53	0.47	0.87	-0.12	-0.14
Ordered	δ	0.61	0.01	1	-0.39	-0.39	0.46	0.58	0.74	0.09	-0.01
Generating Model M2PLMix N = 1000, J = 40 $\lambda = 0.2$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.61	0.24	1	-0.22	-0.3	0.59	0.54	0.97	-0.08	-0.16
Ordered	δ	0.58	0.1	1	-0.25	-0.38	0.54	0.35	0.94	-0.07	-0.24
Generating Model M2PLMix N = 2000, J = 80 $\lambda = 0.2$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.6	0.57	1	-0.05	-0.17	0.6	0.66	0.99	-0.02	-0.13
Ordered	δ	0.6	0.24	1	-0.13	-0.3	0.6	0.33	0.99	-0.09	-0.27

M_1: correct specification of the location, M_2: correct specification for speeded examinees, M_3: correct specification for nonspeeded examinees, M_4: relative bias, M_5: proportion difference of specified speeded between true and estimated.

Data generated under M2PLMix Model where $\lambda = 0.4$

The following section discusses the results when the data were generated under the M2PLMix model with the baseline probability of not being speeded $\lambda = 0.4$, and model parameters were estimated under the 2PLMix and M2PLMix models. When the baseline probability of not being speeded increased, it was expected that fewer examinees were affected by speededness. This expectation was confirmed in Tables 4.18 and 4.19 regarding the recovery of model parameters. The M2PLMix model was unable to proficiently recover α, β, and θ when the sample size $N = 500$ and the test length $J = 20$, given the lower number of speeded examinees. As the sample size and the number of items increased, the M2PLMix model was able to better estimate the α, β, and θ parameters. When the sample size was $N = 1000, 2000$ and the length of test was $J = 40, 80$, the M2PLMix and 2PLMix models were comparable in terms of the bias and RMSE of α, β, and θ parameters. It was sufficient to conclude that the models' recovery of the IRT model parameters were comparable except when the sample size and the number of items were small.

Table 4.18: IRT model parameters: Generated Model M2PLMix; Estimated Models the M2PLMix and 2PLMix with $\lambda = 0.4$.

Generating Model M2PLMix N = 500, J = 20 $\lambda = 0.4$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.11	0.11	0.19	0.2	-0.03	0.07	0.16	0.18
	β	-0.13	0.12	0.18	0.2	0.01	0.07	0.13	0.21
	θ	-0.02	0.34	0.38	0.16	-0.03	0.3	0.37	0.13
Ordered	α	-0.1	0.09	0.16	0.21	-0.04	0.05	0.14	0.22
	β	-0.06	0.04	0.11	0.2	0.04	0.06	0.12	0.18
	θ	-0.02	0.3	0.35	0.14	-0.02	0.28	0.35	0.13
Generating Model M2PLMix N = 1000, J = 40 $\lambda = 0.4$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.04	0.07	0.11	0.18	0.03	0.04	0.1	0.15
	β	-0.06	0.07	0.1	0.15	0	0.07	0.09	0.17
	θ	-0.03	0.18	0.27	0.12	-0.03	0.16	0.27	0.11
Ordered	α	-0.06	0.06	0.11	0.17	0.01	0.03	0.1	0.14
	β	-0.03	0.03	0.08	0.1	0.01	0.04	0.08	0.13
	θ	-0.03	0.18	0.26	0.14	-0.03	0.18	0.27	0.13
Generating Model M2PLMix N = 2000, J = 80 $\lambda = 0.4$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.01	0.04	0.07	0.06	0.03	0.03	0.07	0.05
	β	-0.04	0.04	0.06	0.07	-0.02	0.04	0.06	0.09
	θ	-0.03	0.09	0.19	0.14	-0.03	0.08	0.19	0.14
Ordered	α	-0.04	0.05	0.08	0.1	0.01	0.02	0.07	0.05
	β	-0.02	0.02	0.06	0.04	-0.01	0.02	0.05	0.06
	θ	-0.03	0.12	0.2	0.16	-0.03	0.1	0.19	0.14

Within Figures 4.15 - 4.18, the bias and RMSE of the estimation for the discrimination and difficulty parameters under the M2PLMix and 2PLMix models are shown. Figures 4.15 and 4.16 show that the 2PLMix was able to recover the discrimination parameter regardless of item ordering, especially towards the end of the test. This was not the case for the M2PLMix model, wherein the discrimination parameter was overestimated towards the end of the test for both types of item ordering. Figure 4.17 reveals that the difficulty parameter for both models experienced a similar pattern when item ordering was random. Towards the end of the test, the difficulty parameter was overestimated for the M2PLMix and underestimated for the 2PLMix model. However, in Figure 4.18, the difficulty parameters were recovered well for both models, although the M2PLMix did not vary as much as the 2PLMix towards the end of the assessment.

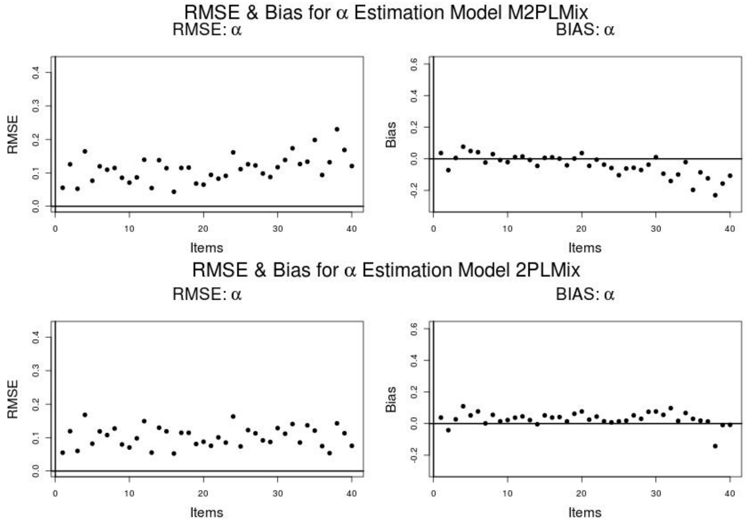

Figure 4.15: Bias and RMSE for α, data generation under M2PLMix model, estimation was M2PLMix (top) and 2PLMix models (bottom), N =1000, J = 40 λ = 0.4, Item Ordering was Random

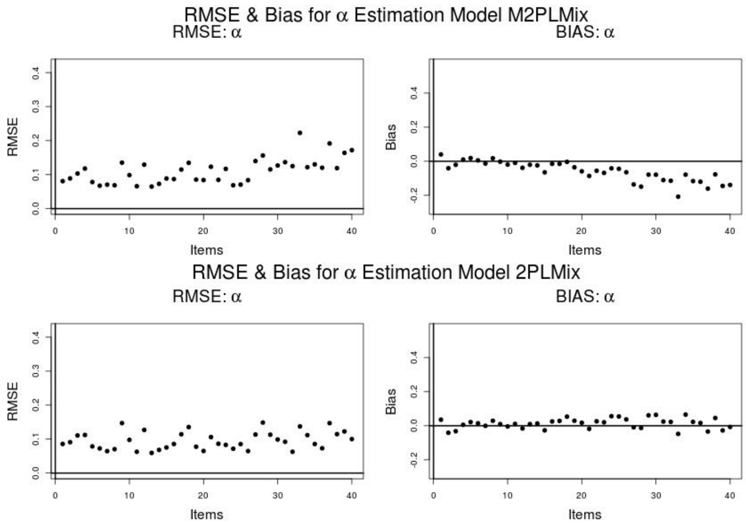

Figure 4.16: Bias and RMSE for α, data generation under M2PLMix model, estimation was M2PLMix and 2PLMix models, N =1000, J = 40 λ = 0.4, Item Ordering was Ordered

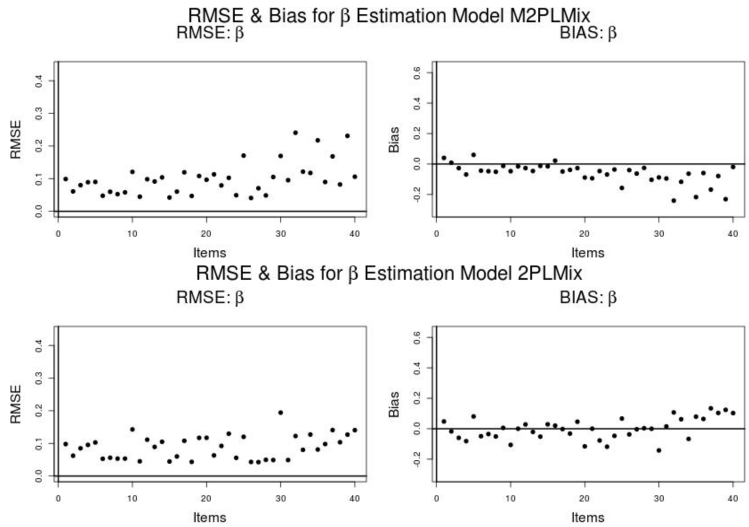

Figure 4.17: Bias and RMSE for β, data generation under M2PLMix model, estimation was M2PLMix and 2PLMix models, N =1000, J = 40 λ = 0.4, Item Ordering was Random

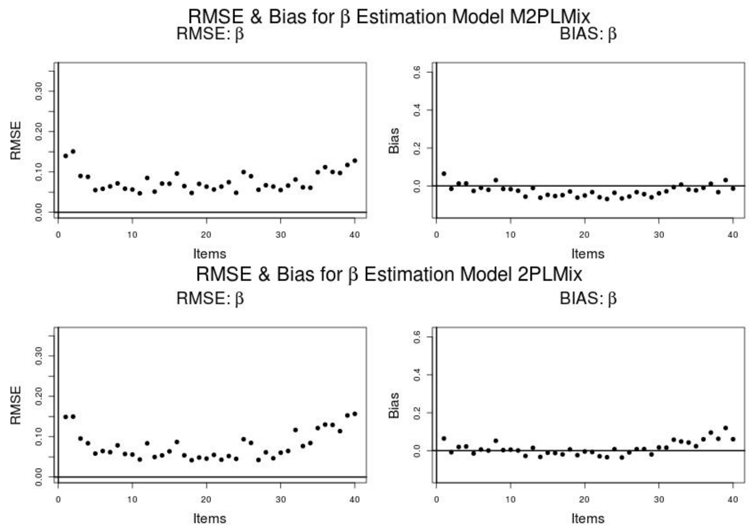

Figure 4.18: Bias and RMSE for β, data generation under M2PLMix model, estimation was M2PLMix and 2PLMix models, N =1000, J = 40 λ = 0.4, Item Ordering was Ordered

The recovery of the location of speededness parameter is found in Table 4.19. M_2 showed that the 2PLMix was able to correctly identify speeded examinees more accurately than the M2PLMix model with respect to all conditions shown on the table. When the sample size and test length were $N = 500$ and $J = 20$, respectively, the M2PLMix model could not sufficiently identify speeded examinees, especially when item difficulty was ordered. As the baseline probability of not being speeded increased, less examinees became speeded, meaning the M_1 increased as well. M_3, M_4, and M_5 indicate that the 2PLMix model did better than the M2PLMix in terms of measuring the speededness location. However, the 2PLMix was unable to recover unspeeded examinees as well as the M2PLMix when the sample size and the number of items were $N = 500$ and $J = 20$, respectively.

Table 4.19: Location of Speededness: Generated Model M2PLMix; Estimated Models the M2PLMix and 2PLMix with $\lambda = 0.4$.

Generating Model M2PLMix N = 500, J = 20 $\lambda = 0.4$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.7	0.01	1	-0.27	-0.29	0.69	0.16	0.98	-0.2	-0.24
Ordered	δ	0.67	0	1	-0.27	-0.33	0.65	0.12	0.97	-0.17	-0.28
Generating Model M2PLMix N = 1000, J = 40 $\lambda = 0.4$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.71	0.25	1	-0.17	-0.22	0.71	0.39	1	-0.11	-0.18
Ordered	δ	0.7	0.08	1	-0.22	-0.27	0.7	0.17	0.99	-0.17	-0.24
Generating Model M2PLMix N = 2000, J = 80 $\lambda = 0.4$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.72	0.55	1	-0.04	-0.13	0.72	0.6	1	-0.02	-0.11
Ordered	δ	0.69	0.24	1	-0.16	-0.24	0.69	0.29	1	-0.11	-0.22

M_1: correct specification of the location, M_2: correct specification for speeded examinees, M_3: correct specification for nonspeeded examinees, M_4: relative bias, M_5: proportion difference of specified speeded between true and estimated.

Table 4.20 denotes the non-convergence rate of α, β, and θ. Each numerical value represents the average percent of parameters that did not converge. This table shows that under most conditions nearly all parameters converged using the Gelman-Rubin criteria. Though some parameters did not converge under these speeded models, these occurrences were trivially significant.

Table 4.20: Percentage of Non-Convergence of IRT Parameters, α, β, and θ for M2PLMix and 2PLMix models, generated under M2PMixL model, using the Gelman-Rubin Criteria

$N = 500,\ J = 20,\ \lambda = 0.4$			
		M2PLMix R	2PLMix R
Random	α	0	0
	β	0	0
	θ	0.008	0.01
Ordered	α	0	0.05
	β	0.05	0.05
	θ	0.012	0.028
$N = 1000,\ J = 40,\ \lambda = 0.4$			
		M2PLMix R	2PLMix R
Random	α	0	0
	β	0.1	0
	θ	0.001	0.003
Ordered	α	0	0
	β	0	0
	θ	0.022	0.01
$N = 2000,\ J = 80,\ \lambda = 0.4$			
		M2PLMix R	2PLMix R
Random	α	0	0
	β	0	0
	θ	0.005	0
Ordered	α	0	0
	β	0	0
	θ	0.032	0.005

Data generated under M2PLMix Model where $\lambda = 0.6$

This section provides the results when the data were generated under the M2PLMix model with the baseline probability of not being speeded $\lambda = 0.6$, and model parameters were estimated under the 2PLMix and M2PLMix models. In terms of the recovery of α, β and θ, fewer individuals were affected by speededness when $\lambda = 0.6$ than the previous two sections. As a consequence, those that became speeded tended to have low ability. In Table 4.21, this effect was demonstrated by the small values for bias and RMSE under both models for all conditions compared to $\lambda = 0.2$ or $\lambda = 0.4$. When the sample size and number of items were $N = 2000$ and $J = 80$, respectively, the item ordering style did not impact the recovery of the α, β and θ parameters, with the exception of RMSE SD for the α parameter.

As noted previously, the M2PLMix did not recover the IRT model parameters well when the number of data points (sample size and number of items) were small, whereas the 2PLMix model was able to recover the parameters well. The M2PLMix model became comparable to the 2PLMix model as the sample size and number of items increased.

Table 4.21: IRT model parameters: Generated Model M2PLMix; Estimated Models the M2PLMix and 2PLMix with $\lambda = 0.6$.

Generating Model M2PLMix N = 500, J = 20 $\lambda = 0.6$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.11	0.12	0.18	0.17	-0.05	0.07	0.14	0.19
	β	-0.09	0.11	0.15	0.18	-0.01	0.06	0.11	0.21
	θ	-0.03	0.31	0.35	0.15	-0.02	0.29	0.35	0.14
Ordered	α	-0.04	0.06	0.14	0.19	-0.01	0.05	0.13	0.21
	β	-0.04	0.06	0.11	0.26	0.02	0.04	0.11	0.26
	θ	-0.02	0.29	0.34	0.15	-0.02	0.28	0.35	0.14
Generating Model M2PLMix N = 1000, J = 40 $\lambda = 0.6$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.01	0.05	0.1	0.16	0.02	0.03	0.1	0.16
	β	-0.05	0.04	0.09	0.13	-0.03	0.04	0.08	0.14
	θ	-0.02	0.16	0.26	0.1	-0.03	0.15	0.26	0.1
Ordered	α	-0.04	0.06	0.1	0.16	0	0.04	0.09	0.13
	β	-0.03	0.03	0.08	0.15	0	0.03	0.08	0.14
	θ	-0.03	0.17	0.26	0.12	-0.03	0.15	0.26	0.12
Generating Model M2PLMix N = 2000, J = 80 $\lambda = 0.6$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.01	0.03	0.07	0.06	0.01	0.02	0.07	0.07
	β	-0.03	0.03	0.06	0.07	-0.02	0.03	0.06	0.08
	θ	-0.03	0.09	0.19	0.15	-0.03	0.09	0.19	0.14
Ordered	α	-0.02	0.04	0.07	0.05	0.01	0.03	0.07	0
	β	-0.03	0.03	0.06	0.06	-0.02	0.02	0.05	0.06
	θ	-0.03	0.11	0.19	0.15	-0.03	0.09	0.19	0.14

The measures used to gauge the recovery of δ can be found in Table 4.22. Similar patterns were found within this table as the ones found in Table 4.19, especially as the sample size and number of items increased. M_1 was closer to 1, and M_4 and M_5 were close to 0, which was caused by the low number of speeded examinees. M_2 was small when the sample size and number of items were small but increased as the number of speeded examinees increased. Lastly, M_3 was consistently 1 for both models due to the number of examinees not being speeded. The 2PLMix was comparable to the M2PLMix for all measures except for M_2.

Table 4.22: Location of Speededness: Generated Model M2PLMix; Estimated Models the M2PLMix and 2PLMix with $\lambda = 0.6$.

Generating Model M2PLMix N = 500, J = 20 $\lambda = 0.6$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.8	0.01	1	-0.19	-0.2	0.8	0.03	1	-0.19	-0.2
Ordered	δ	0.79	0	1	-0.19	-0.21	0.79	0.01	1	-0.18	-0.2
Generating Model M2PLMix N = 1000, J = 40 $\lambda = 0.6$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.81	0.27	1	-0.08	-0.14	0.81	0.32	1	-0.07	-0.13
Ordered	δ	0.79	0.1	1	-0.13	-0.18	0.79	0.15	1	-0.1	-0.17
Generating Model M2PLMix N = 2000, J = 80 $\lambda = 0.6$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.81	0.59	1	-0.03	-0.08	0.81	0.62	1	-0.02	-0.08
Ordered	δ	0.78	0.23	1	-0.1	-0.17	0.78	0.29	1	-0.06	-0.16

M_1: correct specification of the location, M_2: correct specification for speeded examinees, M_3: correct specification for nonspeeded examinees, M_4: relative bias, M_5: proportion difference of specified speeded between true and estimated.

Table 4.23 displays model fit indices when the data were generated under the M2PLMix model and estimated by the M2PLMix and 2PLMix models for each level of λ when the item difficulty was random (item difficulty ordered in Appendix C). The model fit for these two models shows that the M2PLMix model did not fit the data well compared to the 2PLMix model under all measures except for when the sample size and number of items were $N = 500$ and $J = 20$ and the baseline probability of not being speeded was $\lambda = 0.4$ and $\lambda = 0.6$. However, as the sample size and number of items increased the difference between the models became smaller for all measures, regarding the $-2LL$. For the other four measures, the 2PLMix and M2PLMix models were comparable in terms of fitting the data. This result was probably caused by the poor calibration of M2PLMix model parameters, especially when the sample size and the number of items were small.

Table 4.23: Model Fit Data Generation under M2PLMix Model; $\lambda = 0.2$, $\lambda = 0.4$, $\lambda = 0.6$, Item Difficulty Ordering Random

Model Fit Data Generation under the M2PLMix Model						
M2PLMix Model						
		$-2LL$	AIC	AIC_c	BIC	$Adj.BIC$
$\lambda = 0.2$	$N = 500, J = 20$	8710.53	10790.53	6788.17	15173.73	11872.7
	$N = 1000, J = 40$	34186.4	38346.4	30338.11	48554.53	41948.33
	$N = 2000, J = 80$	130134.6	138454.6	122434.47	161754.36	148537.81
$\lambda = 0.4$	$N = 500, J = 20$	8815.63	10895.63	6893.26	15278.82	11977.8
	$N = 1000, J = 40$	34657.43	38817.43	30809.15	49025.57	42419.37
	$N = 2000, J = 80$	135331.38	143651.38	127631.25	166951.14	153734.59
$\lambda = 0.6$	$N = 500, J = 20$	8792.62	10872.62	6870.25	15255.81	11954.79
	$N = 1000, J = 40$	35447.63	39607.63	31599.34	49815.76	43209.56
	$N = 2000, J = 80$	138533.06	146853.06	130832.93	170152.82	156936.27
2PLMix Model						
		$-2LL$	AIC	AIC_c	BIC	$Adj.BIC$
$\lambda = 0.2$	$N = 500, J = 20$	8563.77	10643.77	6641.41	15026.96	11725.94
	$N = 1000, J = 40$	33096.74	37256.74	29248.46	47464.88	40858.68
	$N = 2000, J = 80$	129468.32	137788.32	121768.18	161088.08	147871.53
$\lambda = 0.4$	$N = 500, J = 20$	8902.46	10982.46	6980.09	15365.65	12064.62
	$N = 1000, J = 40$	34348.36	38508.36	30500.07	48716.49	42110.29
	$N = 2000, J = 80$	135160.33	143480.33	127460.2	166780.09	153563.54
$\lambda = 0.6$	$N = 500, J = 20$	8923.83	11003.83	7001.46	15387.02	12085.99
	$N = 1000, J = 40$	35431.06	39591.06	31582.77	49799.19	43193
	$N = 2000, J = 80$	138371.07	146691.07	130670.93	169990.82	156774.27

Data generated under 2PLMix Model where $\pi = 0.2$

The following section discusses the results when the data were generated under the 2PLMix model with the probability of not being speeded $\pi = 0.2$, and model parameters were estimated under the 2PLMix and M2PLMix models. In this simulation (assuming no association between ability and speededness), the 2PLMix was able to recover its model parameters well, except for the difficulty parameter when the sample size was $N = 500$ and the length of test was $J = 20$ (see Table 4.24). In contrast, the M2PLMix was unable to recover the parameters well except for a few conditions. The M2PLMix could not sufficiently recover the difficulty parameter well, regardless of item ordering. Also, the discrimination parameter under the M2PLMix tended to be overestimated when the item difficulty was ordered compared to the case in which the item difficulty was random for $N = 1000, 2000$ and $J = 40, 80$. The RMSE of θ was consistently larger than the RMSE for other parameters no matter model or condition. In contrast, the bias of θ was the smallest regardless of the model or condition. It can be concluded that the 2PLMix was able to recover α and β better than M2PLMix under all conditions. The M2PLMix model poor recover of α, β, and θ was due to the number of examinees that were speeded. As noted in section 3, the 2PLMix generates data in which more examinees are speeded compared to the M2PLMix.

Table 4.24: IRT model parameters: Generated Model 2PLMix; Estimated Models the M2PLMix and 2PLMix with $\pi = 0.2$.

Generating Model 2PLMix N = 500, J = 20 $\pi = 0.2$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.14	0.16	0.23	0.09	-0.06	0.09	0.18	0.16
	β	-0.34	0.25	0.37	0.21	-0.15	0.1	0.18	0.21
	θ	-0.02	0.46	0.45	0.22	-0.02	0.4	0.42	0.19
Ordered	α	-0.06	0.13	0.19	0.18	0	0.07	0.16	0.19
	β	-0.32	0.31	0.38	0.23	-0.15	0.18	0.21	0.3
	θ	-0.03	0.48	0.46	0.22	-0.02	0.44	0.44	0.2
Generating Model 2PLMix N = 1000, J = 40 $\pi = 0.2$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.03	0.06	0.11	0.18	0.04	0.04	0.11	0.18
	β	-0.22	0.24	0.24	0.27	-0.08	0.08	0.11	0.2
	θ	-0.02	0.24	0.33	0.14	-0.03	0.22	0.31	0.13
Ordered	α	-0.12	0.14	0.18	0.2	-0.01	0.06	0.12	0.21
	β	-0.2	0.19	0.23	0.23	-0.08	0.1	0.12	0.2
	θ	-0.03	0.36	0.36	0.22	-0.03	0.3	0.33	0.19
Generating Model 2PLMix N = 2000, J = 80 $\pi = 0.2$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.04	0.05	0.09	0.09	0.03	0.03	0.08	0.06
	β	-0.11	0.17	0.13	0.2	-0.05	0.06	0.08	0.12
	θ	-0.03	0.13	0.23	0.13	-0.02	0.13	0.22	0.12
Ordered	α	-0.09	0.11	0.13	0.19	-0.01	0.04	0.08	0.11
	β	-0.1	0.13	0.12	0.18	-0.04	0.05	0.07	0.17
	θ	-0.02	0.17	0.24	0.15	-0.03	0.14	0.23	0.14

The M2PLMix was able to identify nonspeededness, M_3, better than the 2PLMix model when the sample size and the number of items were small (see

Table 4.25). M_1 was consistent for both models regardless of the conditions. M_2 shows that the 2PLMix was able to correctly identify speeded examinees better than the M2PLMix. M_4 and M_5 decreased as the sample size and the number of items increased, and they performed better when items were randomly ordered compared to being ordered by difficulty. This occurred for both models.

Table 4.25: Location of Speededness: Generated Model 2PLMix; Estimated Models the M2PLMix and 2PLMix with $\pi = 0.2$.

Generating Model 2PLMix N = 500, J = 20 $\pi = 0.2$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.2	0.02	1	-0.7	-0.78	0.2	0.34	0.91	-0.36	-0.51
Ordered	δ	0.2	0	1	-0.61	-0.8	0.18	0.45	0.79	-0.07	-0.4
Generating Model 2PLMix N = 1000, J = 40 $\pi = 0.2$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.19	0.25	1	-0.37	-0.61	0.21	0.5	0.99	-0.15	-0.4
Ordered	δ	0.2	0.1	1	-0.5	-0.72	0.2	0.35	0.96	-0.18	-0.51
Generating Model 2PLMix N = 2000, J = 80 $\pi = 0.2$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.21	0.59	1	-0.07	-0.33	0.22	0.69	1	-0.03	-0.25
Ordered	δ	0.21	0.3	1	-0.28	-0.55	0.21	0.43	1	-0.14	-0.45

M_1: correct specification of the location, M_2: correct specification for speeded examinees, M_3: correct specification for nonspeeded examinees, M_4: relative bias, M_5: proportion difference of specified speeded between true and estimated.

Data generated under 2PLMix Model where $\pi = 0.4$

This section explains the results when the data were generated under the 2PLMix model with the probability of not being speeded $\pi = 0.4$, and model parameters were estimated under the 2PLMix and M2PLMix models. The difficulty parameters were affected the most by both models when the sample size and the number of items were small regardless of item difficulty ordering, as seen in Table 4.26. However, when the sample size increased, both models estimated the difficulty parameter better, with the 2PLMix outperforming the M2PLMix. The discrimination parameters were recovered well by the M2PLMix when the item difficulty was randomly ordered, but when the item difficulty was ordered, these estimates were distorted. The discrimination parameter under the 2PLMix model was generally well estimated, however, when the sample size and the length of test were $N = 500$ and $J = 20$, respectively, these parameters were overestimated. The recovery of the ability parameters for both models was comparable between the models and improved as the sample size and the number of items increased.

Table 4.26: IRT model parameters: Generated Model 2PLMix; Estimated Models the M2PLMix and 2PLMix with $\pi = 0.4$.

Generating Model 2PLMix N = 500, J = 20 $\pi = 0.4$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.07	0.12	0.18	0.18	-0.03	0.06	0.16	0.21
	β	-0.28	0.21	0.31	0.24	-0.15	0.11	0.18	0.21
	θ	-0.02	0.44	0.43	0.22	-0.02	0.38	0.41	0.19
Ordered	α	-0.15	0.19	0.24	0.13	-0.11	0.12	0.19	0.17
	β	-0.21	0.18	0.26	0.18	-0.11	0.1	0.16	0.23
	θ	-0.03	0.51	0.45	0.27	-0.03	0.46	0.43	0.24
Generating Model 2PLMix N = 1000, J = 40 $\pi = 0.4$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.02	0.07	0.11	0.21	0.03	0.05	0.1	0.18
	β	-0.16	0.16	0.19	0.21	-0.08	0.08	0.12	0.18
	θ	-0.03	0.22	0.31	0.14	-0.03	0.19	0.29	0.12
Ordered	α	-0.08	0.09	0.14	0.19	-0.03	0.05	0.11	0.19
	β	-0.17	0.15	0.19	0.22	-0.09	0.09	0.12	0.2
	θ	-0.03	0.31	0.34	0.19	-0.02	0.26	0.32	0.16
Generating Model 2PLMix N = 2000, J = 80 $\pi = 0.4$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.02	0.03	0.07	0.06	0.02	0.03	0.07	0.05
	β	-0.06	0.08	0.09	0.13	-0.04	0.04	0.07	0.11
	θ	-0.03	0.12	0.21	0.13	-0.03	0.12	0.21	0.13
Ordered	α	-0.07	0.08	0.1	0.15	-0.02	0.04	0.08	0.09
	β	-0.07	0.07	0.09	0.15	-0.03	0.03	0.06	0.1
	θ	-0.03	0.17	0.23	0.15	-0.03	0.14	0.21	0.14

Figures 4.19 and 4.20 shows that the 2PLMix estimates of α were comparable to the M2PLMix estimates except towards the end of the assessment. However, when the item difficulty was ordered, the α estimates for both models overesti-

mated towards the end of the test. The inability of the M2PLMix and 2PLmix models to recover the β can be seen in Figures 4.21 and 4.22. The 2PLMix model estimated the β parameter better than the M2PLMix, but both models seemed to overestimate the difficulty parameters towards the end of the assessment independent of item ordering. However, the M2PLMix model estimates of the difficulty parameter, regardless of item ordering, was worse compared to the 2PLMix model in which many items were beyond the threshold of 0.3 towards the end.

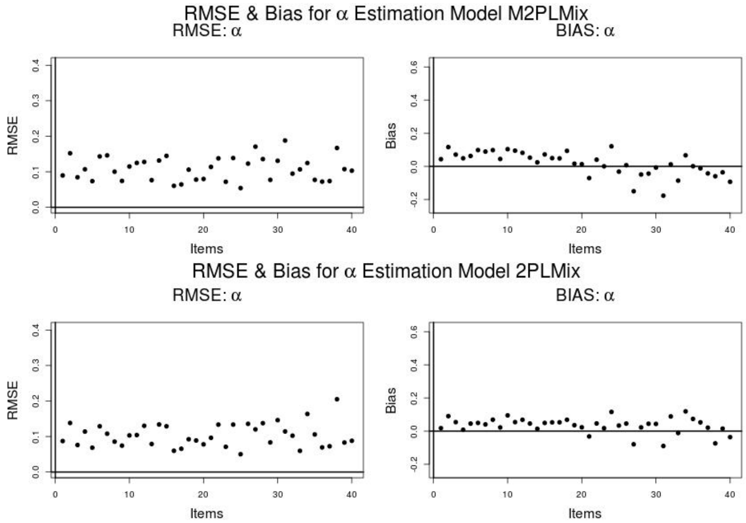

Figure 4.19: Bias and RMSE for α, data generation under 2PLMix model, estimation were the M2PLMix and 2PLMix models, N =1000, J = 40 $\pi = 0.4$, Item Ordering was Random

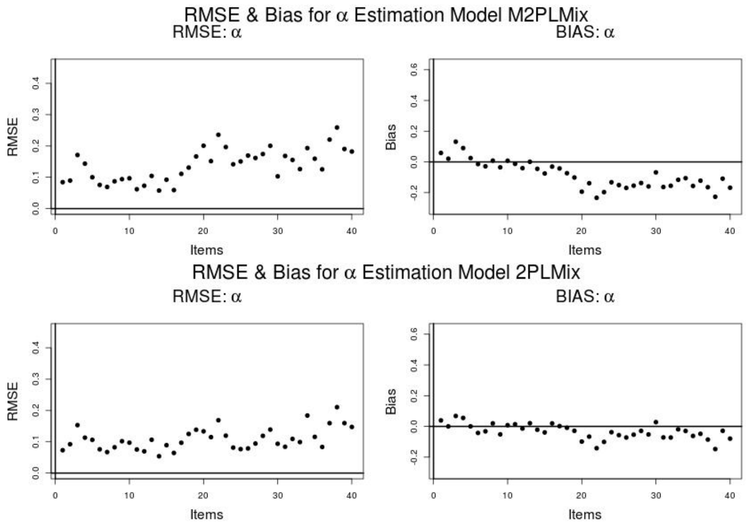

Figure 4.20: Bias and RMSE for α, data generation under 2PLMix model, estimation were the M2PLMix and 2PLMixmodels, N =1000, J = 40 $\pi = 0.4$, Item Ordering was Ordered

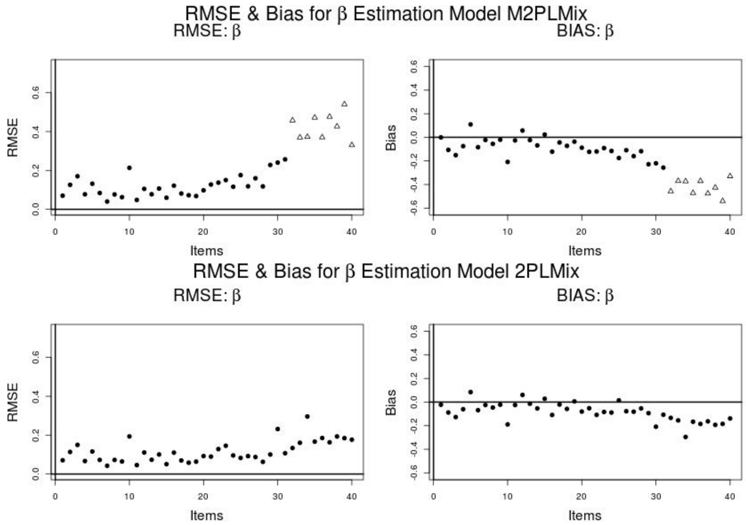

Figure 4.21: Bias and RMSE for β, data generation under 2PLMix model, estimation were the M2PLMix and 2PLMix models, N =1000, J = 40 $\pi = 0.4$, Item Ordering was Random

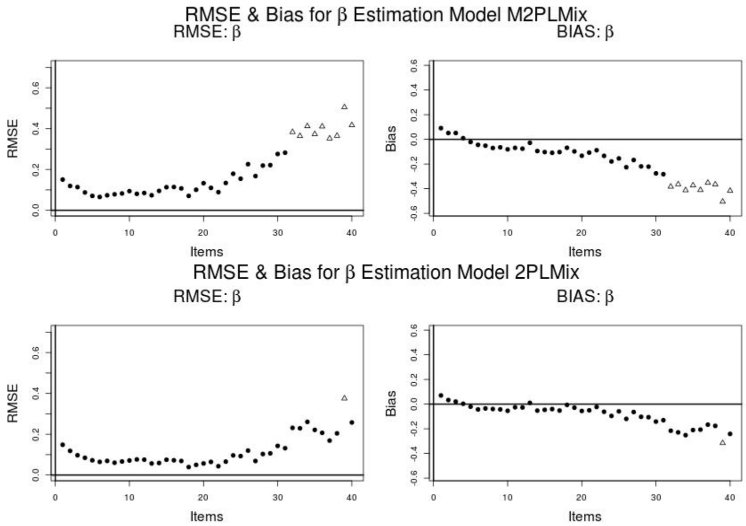

Figure 4.22: Bias and RMSE for β, data generation under 2PLMix model, estimation were the M2PLMix and 2PLMix models, N =1000, J = 40 $\pi = 0.4$, Item Ordering was Ordered.

Similar patterns in 4.25 were found in Table 4.27 in terms of the measures of accuracy of the inception of speededness. For M_3, the 2PLMix did not proficiently recover those that were not speeded compared to the M2PLMix when the sample size and number of items were small. Similarly, the M2PLMix did not recover those that were speeded as well as the 2PLMix. A slight difference between the $\pi = 0.2$ and $\pi = 0.4$ simulations was that the classification of speeded vs. nonspeeded examinees caused the M_1 statistic to increase for both models.

Table 4.27: Location of Speededness: Generated Model 2PLMix; Estimated Models the M2PLMix and 2PLMix with $\pi = 0.4$.

Generating Model 2PLMix N = 500, J = 20 $\pi = 0.4$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.4	0.01	1	-0.53	-0.6	0.4	0.1	0.99	-0.45	-0.54
Ordered	δ	0.4	0	1	-0.58	-0.6	0.4	0.11	0.98	-0.42	-0.52
Generating Model 2PLMix N = 1000, J = 40 $\pi = 0.4$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.4	0.26	1	-0.24	-0.45	0.4	0.38	1	-0.17	-0.38
Ordered	δ	0.38	0.1	1	-0.41	-0.55	0.38	0.22	0.99	-0.26	-0.48
Generating Model 2PLMix N = 2000, J = 80 $\pi = 0.4$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.42	0.63	1	-0.05	-0.22	0.42	0.67	1	-0.04	-0.2
Ordered	δ	0.4	0.27	1	-0.22	-0.44	0.4	0.34	1	-0.15	-0.4

M_1: correct specification of the location, M_2: correct specification for speeded examinees, M_3: correct specification for nonspeeded examinees, M_4: relative bias, M_5: proportion difference of specified speeded between true and estimated.

Table 4.28 denotes the convergence rate of α, β, and θ. This table shows that under most conditions when the M2PLMix or 2PLMix models were used to estimate the model parameters, almost all parameters converged using the Gelman-Rubin criteria. When the sample size and the number of items were large (i.e., $N = 2000$, $J = 80$), all the parameters converged under the 2PLMix model. As the sample size and number of items increased, fewer parameters converged under M2PLMix, but this difference was negligible.

Table 4.28: Percentage of Non-Convergence of IRT Parameters, α, β, and θ for M2PLMix and 2PLMix models, generated under 2PLMix model, using the Gelman-Rubin Criteria

$N = 500,\ J = 20,\ \lambda = 0.4$			
		M2PLMix R	2PLMix R
Random	α	0	0
	β	0	0
	θ	0.01	0.024
Ordered	α	0	0
	β	0	0
	θ	0.01	0.014
$N = 1000,\ J = 40,\ \lambda = 0.4$			
		M2PLMix R	2PLMix R
Random	α	0	0
	β	0	0
	θ	0.03	0.01
Ordered	α	0	0
	β	0	0.025
	θ	0.055	0.01
$N = 2000,\ J = 80,\ \lambda = 0.4$			
		M2PLMix R	2PLMix R
Random	α	0.013	0
	β	0.013	0
	θ	0.016	0
Ordered	α	0.013	0
	β	0	0
	θ	0.046	0

Data generated under 2PLMix Model where $\pi = 0.6$

The following section provides the results when the data were generated under the 2PLMix model with the probability of not being speeded $\pi = 0.6$, and model parameters were estimated under the 2PLMix and M2PLMix models. Similar results were produced when $\pi = 0.6$ (compared to other levels of π) in terms of the α, β, and θ as well as the measures for the location of speededness, as seen in Tables 4.29 and 4.30. The difficulty parameters were not well estimated under the M2PLMix model when the number of items and sample size were small. However, when the number of observations increased the M2PLMix began to recover the difficulty parameter at a rate close to that of the 2PLMix model. Since the proportion of speeded examinees was smaller than that in the other two conditions ($\pi = 0.2$ or $\pi = 0.4$), the bias and RMSE were smaller (see Table 4.29).

Table 4.29: IRT model parameters: Generated Model 2PLMix; Estimated Models the M2PLMix and 2PLMix with $\pi = 0.6$.

		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Generating Model 2PLMix N = 500, J = 20 $\pi = 0.6$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.16	0.17	0.24	0.12	-0.12	0.11	0.2	0.12
	β	-0.22	0.17	0.26	0.19	-0.14	0.1	0.18	0.2
	θ	-0.02	0.44	0.42	0.21	-0.02	0.38	0.4	0.18
Ordered	α	-0.04	0.1	0.16	0.2	-0.03	0.07	0.15	0.19
	β	-0.15	0.19	0.22	0.24	-0.1	0.14	0.17	0.26
	θ	-0.02	0.42	0.4	0.23	-0.02	0.4	0.39	0.21
Generating Model 2PLMix N = 1000, J = 40 $\pi = 0.6$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0	0.04	0.1	0.14	0.02	0.03	0.1	0.13
	β	-0.11	0.1	0.14	0.18	-0.07	0.06	0.1	0.16
	θ	-0.03	0.23	0.29	0.15	-0.03	0.2	0.28	0.13
Ordered	α	-0.06	0.07	0.12	0.2	-0.02	0.04	0.1	0.18
	β	-0.1	0.08	0.13	0.2	-0.06	0.05	0.11	0.19
	θ	-0.02	0.31	0.32	0.21	-0.03	0.26	0.3	0.17
Generating Model 2PLMix N = 2000, J = 80 $\pi = 0.6$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.01	0.02	0.07	0.04	0.02	0.02	0.07	0.04
	β	-0.05	0.05	0.07	0.1	-0.04	0.03	0.07	0.1
	θ	-0.02	0.1	0.2	0.14	-0.03	0.1	0.2	0.13
Ordered	α	-0.03	0.05	0.08	0.13	-0.01	0.03	0.07	0.08
	β	-0.06	0.07	0.08	0.14	-0.03	0.04	0.06	0.13
	θ	-0.02	0.15	0.21	0.16	-0.03	0.12	0.2	0.14

As shown in Table 4.30, the ability to correctly label examinees as speeded or not speeded (M_1) were lower compared to when the data was generated under the M2PLMix (see Table 4.22) because the number of individuals affected by speededness was greater under the 2PLMix model. Since all examinees have the same likelihood of being speeded under the 2PLMix model, more people were assumed to be speeded, making it difficult to correctly identify the inception of speededness. This trend was found in the other measures of the inception of speededness. As with the other levels of π, item ordering was found to have an impact on measures M_2 to M_5 for $\pi = 0.6$.

Table 4.30: Location of Speededness: Generated Model 2PLMix; Estimated Models the M2PLMix and 2PLMix with $\pi = 0.6$.

Generating Model 2PLMix N = 500, J = 20 $\pi = 0.6$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.6	0	1	-0.45	-0.4	0.6	0.03	1	-0.43	-0.39
Ordered	δ	0.61	0	1	-0.34	-0.39	0.61	0.01	1	-0.33	-0.38
Generating Model 2PLMix N = 1000, J = 40 $\pi = 0.6$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.59	0.28	1	-0.22	-0.3	0.59	0.33	1	-0.2	-0.27
Ordered	δ	0.59	0.11	1	-0.29	-0.36	0.59	0.17	1	-0.23	-0.34
Generating Model 2PLMix N = 2000, J = 80 $\pi = 0.6$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.62	0.63	1	-0.05	-0.14	0.62	0.66	1	-0.03	-0.13
Ordered	δ	0.62	0.24	1	-0.15	-0.29	0.62	0.33	1	-0.09	-0.25

M_1: correct specification of the location, M_2: correct specification for speeded examinees, M_3: correct specification for nonspeeded examinees, M_4: relative bias, M_5: proportion difference of specified speeded between true and estimated.

The model fit found in Table 4.31 indicates that the 2PLMix model fits the data better than M2PLMix model for all conditions, except when $\pi = 0.4$ or $\pi = 0.6$ and the sample size and number of items were $N = 500$ and $J = 20$, respectively, for all indices. As the sample size, test length, and probability of not being speeded increased, the difference between the fit of the two models became more comparable for all measures, with a few exceptions in which the M2PLMix model fits the data better.

Table 4.31: Model Fit Data Generation: 2PLMix Model and Parameter Estimation: M2PLMix (top 3) and 2PLMix Models (bottom 3)

Model Fit Data Generation under the 2PLMix Model						
M2PLMix Model						
		$-2LL$	AIC	AIC_c	BIC	$Adj.BIC$
$\pi = 0.2$	$N = 500, J = 20$	8539.99	10619.99	6617.62	15003.18	11702.15
	$N = 1000, J = 40$	31825.29	35985.29	27977.01	46193.43	39587.23
	$N = 2000, J = 80$	112551.16	120871.16	104851.02	144170.91	130954.37
$\pi = 0.4$	$N = 500, J = 20$	8866	10946	6943.63	15329.19	12028.17
	$N = 1000, J = 40$	33258.75	37418.75	29410.47	47626.88	41020.69
	$N = 2000, J = 80$	120792.62	129112.62	113092.48	152412.37	139195.83
$\pi = 0.6$	$N = 500, J = 20$	8690.92	10770.92	6768.55	15154.11	11853.08
	$N = 1000, J = 40$	34379.62	38539.62	30531.33	48747.75	42141.56
	$N = 2000, J = 80$	130492.98	138812.98	122792.84	162112.74	148896.19
2PLMix Model						
		$-2LL$	AIC	AIC_c	BIC	$Adj.BIC$
$\pi = 0.2$	$N = 500, J = 20$	8289.19	10369.19	6366.83	14752.39	11451.36
	$N = 1000, J = 40$	29717.21	33877.21	25868.92	44085.34	37479.15
	$N = 2000, J = 80$	110943.22	119263.22	103243.08	142562.97	129346.42
$\pi = 0.4$	$N = 500, J = 20$	8891.78	10971.78	6969.41	15354.97	12053.94
	$N = 1000, J = 40$	32645.4	36805.4	28797.11	47013.53	40407.34
	$N = 2000, J = 80$	120279.63	128599.63	112579.49	151899.38	138682.84
$\pi = 0.6$	$N = 500, J = 20$	8814.45	10894.45	6892.08	15277.64	11976.62
	$N = 1000, J = 40$	34286.08	38446.08	30437.8	48654.21	42048.02
	$N = 2000, J = 80$	130266.33	138586.33	122566.19	161886.09	148669.54

4.3.4 Study 4 Outcomes

Data generated under GPC Model

This section provides the results in which the data were generated under the GPC model, and model parameters were estimated under the 2PLMix and M2PLMix models. The sample size and the test length were $N = 1000$ and $J = 40$, respectively, for all conditions in this study. When the data were generated under the GPC model, a large discrepancy occurred that was not seen during the generation of data within the other two speededness models (M2PLMix and 2PLMix). Under all the data sets that were generated under the GPC model, on average three examinees were simulated as not speeded within data. This concept directly impacted the capabilities of the M2PLMix and 2PLMix models to recover model parameters well.

Figures 4.23 - 4.28 display the inability of the M2PLMix and 2PLMix models to estimate item parameters when the data were generated under the GPC model. These figures show that no matter how the data were generated under the GPC model, the model parameters were not recovered the well. Though these figures are not ideal, certain patterns can be captured in these graphs. For example, the bias and RMSE patterns found during the estimation of α and β were comparable between the M2PLMix and 2PLMix models. When the correlation between ability and speededness was the smallest (i.e., $\rho(\theta, \eta) = 0.5$) there were more items that were classified as below 0.3 for both models.

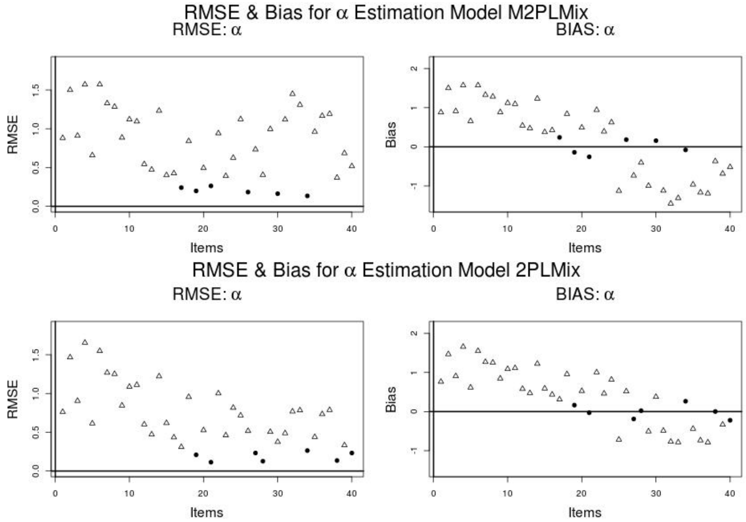

Figure 4.23: Bias and RMSE for α, data generation under the GPC Model ($\rho(\theta, \eta) = 0.5$) and parameter estimation under the M2PLMix (top) and 2PLMix (bottom) models, N =1000, J = 40.

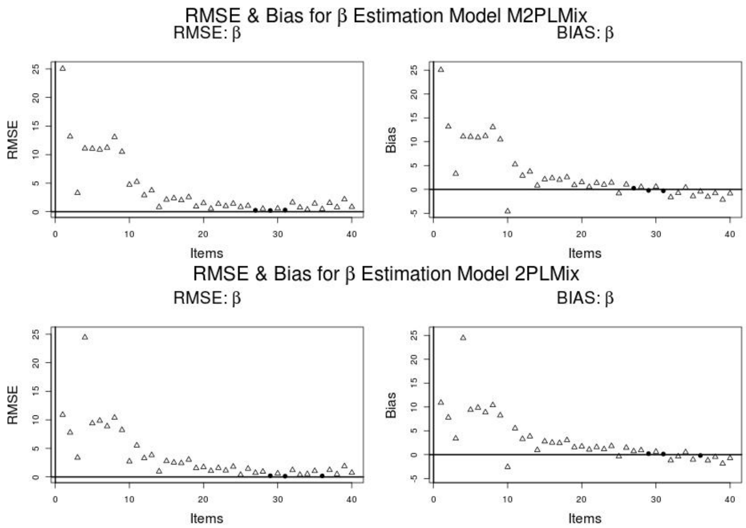

Figure 4.24: Bias and RMSE for β, data generation under the GPC Model ($\rho(\theta, \eta) = 0.5$) and parameter estimation under the M2PLMix (top) and 2PLMix (bottom) models, N =1000, J = 40.

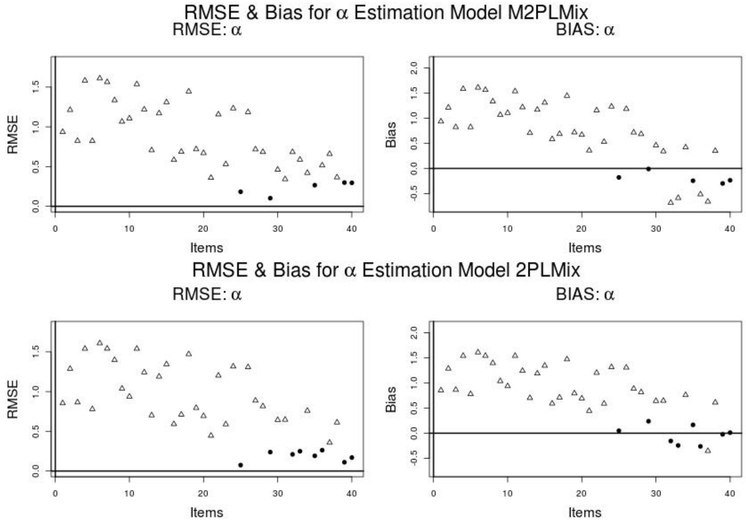

Figure 4.25: Bias and RMSE for α, data generation under the GPC Model ($\rho(\theta, \eta) = 0.6$) and parameter estimation under the M2PLMix (top) and 2PLMix (bottom) models, N =1000, J = 40.

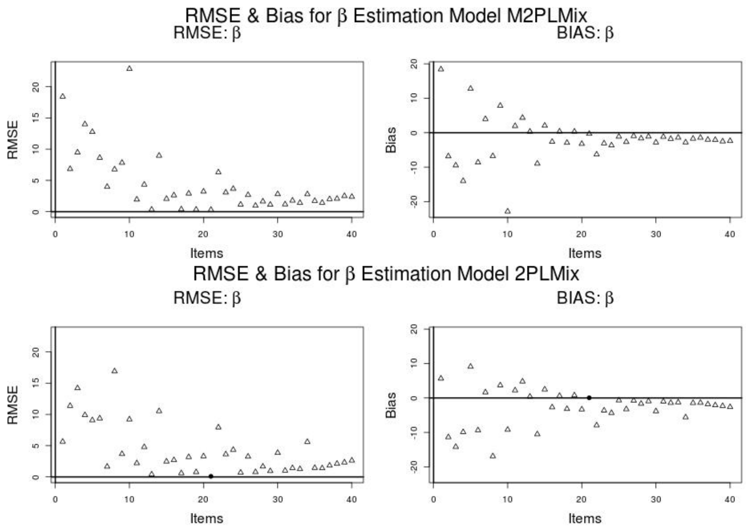

Figure 4.26: Bias and RMSE for β, data generation under the GPC Model ($\rho(\theta, \eta) = 0.6$) and parameter estimation under the M2PLMix(top) and 2PLMix (bottom) models, N =1000, J = 40.

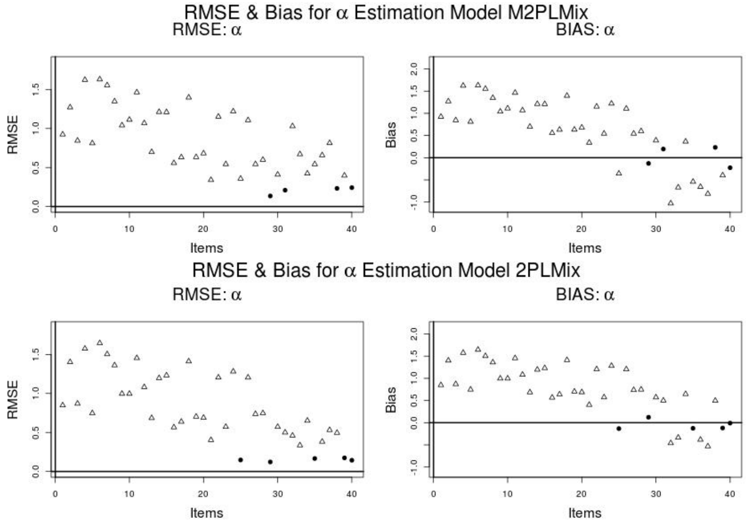

Figure 4.27: Bias and RMSE for α, data generation under the GPC Model ($\rho(\theta, \eta) = 0.7$) and parameter estimation under the M2PLMix(top) and 2PLMix (bottom) models, N =1000, J = 40.

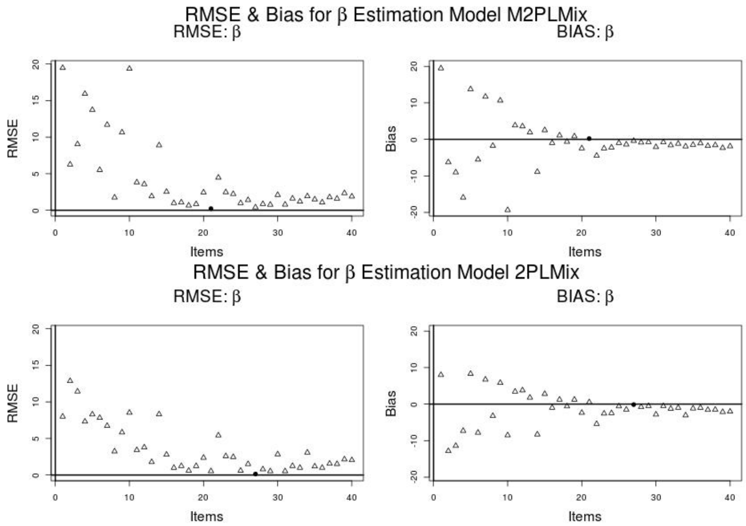

Figure 4.28: Bias and RMSE for β, data generation under the GPC Model ($\rho(\theta, \eta) = 0.7$) and parameter estimation under the M2PLMix(top) and 2PLMix (bottom) models, N =1000, J = 40.

Tables 4.32 and 4.33 display the recovery model parameters from both the M2PLMix and the 2PLMix. Again these results were not ideal, but they provide insight into the capabilities of these speededness models. For instance, the greater the correlation between inception of speededness and ability level, the worse the performance on the assessment, which makes the ability level easy to recover. However, the difficulty parameter had too large standard deviations of bias and RMSE to make any generalizations.

Table 4.32: IRT model parameters: Generation of data through Gradual Process Change Model under 3 different correlation models.

Estimating Under $\rho(\eta, \theta) = 0.5$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Low	α	0.18	0.91	0.81	0.38	0.38	0.71	0.69	0.37
	β	3.12	5.79	3.9	5.28	3.02	5.01	3.53	4.67
	θ	0.04	0.99	0.84	0.5	0.03	0.93	0.81	0.44
Estimating Under $\rho(\eta, \theta) = 0.6$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Low	α	0.65	0.67	0.83	0.42	0.77	0.56	0.83	0.43
	β	-1.92	6.53	4.54	5.02	-2.69	5.28	4.25	4.08
	θ	0.07	0.89	0.76	0.42	0.06	0.86	0.75	0.4
Estimating Under $\rho(\eta, \theta) = 0.7$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Low	α	0.59	0.73	0.83	0.4	0.7	0.61	0.81	0.41
	β	-0.82	6.73	4.29	5.2	-1.3	4.63	3.48	3.27
	θ	-0.01	0.82	0.71	0.38	-0.02	0.8	0.7	0.37

In terms of M_1 (see Table 4.33), it was impossible for both models to determine the exact location of speededness, since 99 % of the test-taking population was speeded. M_3 was consistently 0 because most examinees were speeded, which implies M_2 to be 1. M_5, the difference between the proportion of speeded examinees, was also very small due to the models identifying most examinees as speeded. M_4 on the other hand was difficult to interpret due to the instability of this statistic in other contexts discussed in the previous studies. These results show that both models have the capacity to identify speeded examinees generated under a different model.

Table 4.33: Location of Speededness: Generation of data through Gradual Process Change Model under 3 different correlation models.

Generated Under GPC Model											
		M2PLMix					2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
$\rho(\eta, \theta) = 0.5$	δ	0.01	1	0	-0.95	0.01	0.02	1	0	-0.8	0.01
$\rho(\eta, \theta) = 0.6$	δ	0	1	0	-0.98	0.01	0	1	0	-0.99	0.01
$\rho(\eta, \theta) = 0.7$	δ	0	1	0	-0.94	0.01	0.01	1	0	-0.87	0.01

M_1: correct specification, M_2: correct speeded specification, M_3: correct not speeded specification, M_4: relative bias, M_5: percentage difference of specified speeded

Chapter 5

Conclusion

5.1 Summary and Implications

In Study 1, an analysis was done to understand the implication of assuming the hyperparameters to be known within the M2PLMix model. When data were generated in which $\omega = 2.5$, the RMSE, Bias, RMSE SD and Bias SD, for all IRT parameters were similar to the normal condition except for the bias of the discrimination parameter. In terms of the inception of speededness, M_1, M_3, and M_5 were comparable between the normal condition and the condition with $\omega = 2.5$; however, M_2 and M_4 were lower when $\omega = 2.5$ than the normal condition. This implies that assuming a lower ω parameter during the estimation of the M2PLMix parameters causes speeded examinees to be classified as not speeded according to the M_2 measure. The relative bias indicated that though misclassification of a hyperparameter occurred, the estimated value was close to the true value.

The discrimination parameter for each item were always underestimated when the data was generated under the $\omega = 2.5$ and was always underestimated with respect to the normal condition except for seven items. On the other hand, the difficulty parameter tended to be around zero when $\omega = 2.5$; however, this parameter became more overestimated towards the end of the test under the normal condition. Increasing the ω parameters during the generations of data causes examinees to become speeded later during the assessment. According to the results, if ω is smaller during estimation of the M2PLMix model parameters compared to ω during the generation of data, the recovery of the item parameters will be the same as the normal condition throughout the entire test. However, individual items tended to lose their discrimination power, and it seems as though less people were classified as speeded.

For data generated under $\omega = 1.5$, the RMSE, Bias, RMSE SD, and Bias SD for IRT model parameters were comparable to the normal condition. In terms of the location of speededness, M_1, M_3, and M_5 were comparable between these two conditions; however M_2 and M_4 were greater when $\omega = 1.5$ compared to the normal condition. This implies that assuming a larger ω parameter during the estimation of the M2PLMix parameters causes speeded examinees to be correctly classified more frequently than the normal condition. The relative bias indicated that when misclassification occurred, examinees that were speeded earlier on the test were misclassified more often than those speeded later during the test. It is important to note that more people were correctly classified as speeded, but the penalty was greater for misclassification earlier on the assessment than later which caused the relative bias measure when $\omega = 1.5$ to be greater than the normal condition. The item parameters were graphically the same when $\omega = 1.5$ compared to the normal condition, seen in figures found in Appendix A. In conclusion, assuming ω to be larger during the estimation process only slightly impacted the recovery of the location of speededness.

When data were generated in which $\kappa = 0.25$, the RMSE, Bias, RMSE SD, and Bias SD for IRT model parameters were comparable to the normal condition. In terms of the location of speededness, $M_1 - M_5$ were comparable between $\kappa = 0.25$ and the normal condition. This implies that assuming a smaller κ parameter during estimation does not impact the recovery of M2PLMix model parameters. The item parameters were graphically the same in terms of the RMSE and bias with respect to the two conditions. In conclusion, assuming κ to be smaller during the estimation process does not have an impact on the estimation of M2PLMix model parameters. When data were generated in which $\kappa = 0.15$, the RMSE, Bias, RMSE SD, and Bias SD for all IRT parameters were similar to the normal condition except for the bias and RMSE of the discrimination parameter. In terms of the location of speededness, M_1, M_3, M_4, and M_5 were comparable between these two conditions, and M_2 was lower for $\kappa = 0.15$. This implies that assuming a larger κ parameter during the estimation of the M2PLMix parameters causes the discrimination parameter to lose its power and causes speeded examinees to be classified as not speeded. The discrimination parameter for each item was always underestimated when $\kappa = 0.15$. These figures can be found in Appendix A. In conclusion, assuming a larger value for the κ during the estimation of item parameters mainly affects the calibration of the discrimination parameter and the model's ability to recover those that are speeded.

For data generated under $\lambda = 0.45$, the majority of M2PLMix model parameters were not affected by estimating parameters under $\lambda = 0.4$ (same results as normal condition). The discrimination parameters for the individual items were underestimated when $\lambda = 0.45$ for all items except for one but this result is comparable to the normal condition. Assuming the baseline probability of not being speeded to be $\lambda = 0.45$ did not impact the test in terms of bias and RMSE compared to $\lambda = 0.4$. The discrimination parameters for each individual item were underestimated for all items except for one item when $\lambda = 0.35$. The difficulty parameters were overestimated when data were generated under either $\lambda = 0.35$ or $\lambda = 0.4$. These figures can be found in Appendix A. Assuming the baseline probability of not being speeded to be $\lambda = 0.35$ did not impact the entire test in terms of bias and RMSE compared to $\lambda = 0.4$. However, individual items for the discrimination and difficulty parameters were impacted when $\lambda = 0.35$ but this was found in the normal condition as well.

In sum, the first study showed that generating data with certain values for the hyperparameters and then assuming that hyperparameter to be another value during the estimation of model parameters were not detrimental in all cases. This research showed that variations to ω and κ may negatively impact the discrimination parameters as a whole. This is important for the implementation of estimation model parameters from real data, as knowing how certain hyperparameter values may cause biases would lead to better estimation of IRT model parameters as well as speededness location. An important note to make when comparing the six different conditions to the normal conditions was that the normal condition did not estimate the M2PLMix model parameters perfectly.

Modifying more than one hyperparameter to determine the impact on the estimation of IRT and location of speededness parameters would contribute to the research. Estimating the hyperparameters would also be valuable as a future study by potentially providing a better estimation of the IRT and speededness model parameters. Lastly, using real data to determine the implications of assuming these hyperparameters to be known or unknown will provide insight into better estimation of the other model parameters.

The second study showed that the 2PL model did not perform well when speededness was present and associated with ability. The M2PLMix model outperformed the 2PL model under all conditions in terms of RMSE, bias, RMSE SD, bias SD, and the inception of speededness when data were generated under the M2PLMix model. As the amount of data and the baseline probability of not being speeded increased both models estimated the examinee and item parameters better, with the M2PLMix model consistently outperforming the 2PL model. This result was expected because the 2PL model does not account for speededness. These results implied that ignoring speededness when it was associated with ability negatively impacts the recovery of model parameters.

On other hand, the recovery of the 2PL model parameters when data were generated under the 2PL model verified the proposed MCMC algorithm's capability of estimation. Moreover, if speededness was not present, the M2PLMix model was only able to recover its parameters well when the number of items and the sample size were large. These conclusions were also found in Appendix B in terms of other conditions. In summary, if speededness was not present, the M2PLMix was able to recover the IRT model parameters as long as the sample size and number of items were large. On the other hand, the 2PL model was consistently unable to recover its parameters when speededness was present and associated with ability.

The third study showed the 2PLMix model can recover its model parameters well when speededness was present and associated with ability, regardless of simulation conditions. This study also determined the M2PlMix model was able to recover its model parameters well when speededness was present but not associated with ability only when the sample size and the number of items were large. When speededness and ability were associated, both models were generally able to recover the IRT model parameters at comparable rates when the sample size and the number of items were large. Only when the sample size and the number of items were small, was the 2PLMix model able to recover the IRT model parameters better than the M2PLMix model. The 2PLMix model was able to recover its model parameters better than the M2PLMix model under the generation of data from both models because the 2PLMix model did not take into account the estimation of θ when estimating δ. This allowed the 2PLMix model to estimate its model parameters better than the M2PLMix model. These patterns were also found in Appendix C.

Ignoring the association between ability and speededness was not detrimental to the recovery of IRT model parameters for the 2PLMix model. These results imply that if there was an association between ability and speededness that occurred in the mechanism of the M2PLMix model, the 2PLMix model can will be able to recover its parameters. But there are other ways of associating ability and speededness that need to be evaluated before discounting the ramifications of the association between ability and speededness. This result does not imply that all speededness models that do not account for this association will recover model parameters well.

On the other hand, when there was no association between speededness and ability, the M2PLMix model was able to recover the IRT model parameters as well as the 2PLMix model only when π, N, and J were large. This implies that the 2PLMix model was the optimal model outside of these circumstances. Essentially, if there was an association between speededness and ability and the sample size and the number of items were large ($N = 2000$ and $J = 80$), the M2PLMix model can be used, otherwise, it would be better to implement the 2PLMix model. These generalizations were also found in the results of the conditions shown within the Appendix C.

When data were generated under the GPC model, both the 2PLMix and M2PLMix models were unable to recover their model parameters well, respectively. This was due to the fact that all examinees were affected by speededness. This implies a scenario at which the 2PLMix model was not able to recover model parameters when the association between ability and speededness was present. This study showed an alternative way to associate ability and speededness, but also portrayed the unrealism of the GPC model, which identified 99% of examinees were speeded.

5.2 Conclusion and Further Studies

The purpose of this dissertation was to propose a model that associated ability with speededness and propose an algorithm to estimate not only IRT model parameters but also the inception of speededness parameter within this speededness model. This model was used to determine the ramifications of assuming hyperparameters to be known during the estimation of its model parameters. This dissertation also investigated the implications when not only the association between speededness and ability was ignored, but also when speededness was ignored entirely. Furthermore, the studies provided insight in the assumption of associating ability with speededness under various conditions.

The results suggest that assuming κ and ω to be known when estimating M2PLMix model parameters can have an impact on the recovery of the discrimination parameters and the location at which speededness begins. The findings also verify that the 2PL model cannot estimate model parameters well when speededness is present and associated with ability. If speededness is not present, the M2PLMix model only estimates model parameters well when the sample size and the number of items are large. These findings also suggest that the M2PLMix model can recover its parameters well when speededness is not associated with ability when the sample size and the number if items are large. In addition, the 2PLMix model is able to recover model parameters well when ability and speededness are associated using the M2PLMix model to generate data. Though the M2PLMix model only estimates its parameters well when the sample size is large within size, the idea of a large sample is relative in terms of the implementation of this model on real data in which the sample size could be 100,000 examinees depending on the assessment.

The M2PLmix and 2PLMix models could not recover their model parameters well, respectively, when ability and speededness are associated under the GPC model. One of the conclusions of this dissertation is not that any speededness model that ignores the association between ability and speededness is appropriate. Therefore, the assumption regarding ability and speededness needs to be tested within previous and future speededness models.

There are many facets of this study that contribute to the research of education measurement, more specifically the topic of speededness. The proposed M2PLMix model provides insights about speededness that has an impact on the calibration of item parameters and the estimation of the ability parameter. According to the results, implementing the M2PLMix model when speededness and ability are associated allows for a better calibration of item parameters and ability estimates compared to when the speededness is not accounted for. This research also shows that item ordering has a large impact on the recovery of item parameters when speededness is present. This information is useful in that test administrations may want to implement this design to reduce speededness effects. This research also enhances the understanding of speededness because it considers various sample sizes and number of items. This is insightful because most studies ignore how the sample size and the number of items on an assessment play a role in how speededness impacts the recovery of model parameters.

In education measurement, the notion of speededness may have an impact on many applications that have not been investigated. One such application is the use of cut scores. If speededness can be account for during an assessment (reducing speededness effects), some individuals that are near the cut point(s) of proficient skill(s) could be better estimated. This could allow for better tests and items that accurately measure one's ability. Another facet that could be analyzed is the baseline probability of not being speeded. In other words, creating multiple baseline probabilities of not being speeded based on a latent variable (not ability). For example, it is possible that different global cultures have different approaches on guessing behavior on an exam, which would impact the baseline probability of not being speeded. This could be modeled on an international test such as the Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS), which can use such models that account for different test-taking strategies.

The inspiration to evaluate test speededness was from the possibility of speededness impacting not only item parameters within IRT but also linking coefficients and equating functions. No research has shown how speededness-reducing models have an influence on the estimation of linking coefficients and equating functions. This research is important because the speededness models do not always remove the entire presence of speededness, especially when the sample size and the number of items are small.

Lastly, other approaches that can be studied based on this dissertation would be to change more than one hyperparameter at a time within the first study.

As mentioned before, changing more than one hyperparameter at a time would give researchers more insight into the ramifications of assuming these parameters known. Another approach would be to implement other speeded IRT models such as the GPC or the IRT-TG to estimate speededness and recovery of model parameters to determine if ignoring this association has negligible effects. Furthermore, creating and implementing other techniques to measure the estimation of the location of speededness would be worthwhile. Most research do not evaluate the effectiveness of recovering the location of speededness. A more robust technique for determining if the location parameter is properly recovered would add to this line of work as well.

REFERENCES

Bejar, I. I. (1985). *Test speededness under number-right scoring: An analysis of the Test of English as a Foreign Language.* (Report No. ETS-RR-85-11). Princeton, NJ: Educational Testing Services.

Bolt, D. M., Cohen, A. S., & Wollack, J. A. (2001). A mixture item response model for multiple-choice data. *Journal of Educational and Behavioral Statistics*, 26(4), 381-409.

Bolt, D. M., Cohen, A. S., & Wollack, J. A. (2002). Item parameter estimation under conditions of test speededness: Application of a mixture Rasch model with ordinal constraints. *Journal of Educational Measurement, 39*, 331-348.

Bridgeman, B., & Cline, F. (2004). Effects of Differentially Time-Consuming Tests on Computer-Adaptive Test Scores. *Journal of Educational Measurement, 41*, 137-148.

Brown, T. L., Li, S., & Yang, Z. (2013). *The Impact of Speededness on Scaling and Equating for Tests with Dichotomous ly and Polytomously Scored Items.* Paper presented at the annual meeting of the American Education Research Association, San Francisco, CA.

Cao, J., & Stokes, S. L. (2008). Bayesian IRT guessing models for partial guessing behaviors. *Psychometrika, 73*, 209-230.

Chang, Y. W., Tsai, R. C., & Hsu, N. J. (2014). A speeded item response model: Leave the harder till later. *Psychometrika, 79*, 255-274.

Chib, S., & Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The American statistician, 49*, 327-335.

De Boeck, P., Cho, S. J., & Wilson, M. (2011). Explanatory secondary dimension modeling of latent differential item functioning. *Applied Psychological Measurement, 35*, 583-603.

De La Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika, 69*, 333-353.

de la Torre, J. (2009). Improving the quality of ability estimates through multidimensional scoring and incorporation of ancillary variables. *Applied Psychological Measurement, 33*, 465-485.

Evans, F. R., & Reilly, R. R. (1972). A Study of Speededness as a Source of Test Bias. *Journal of Educational Measurement, 9*, 123-131.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 457-472.

Goegebeur, Y., De Boeck, P., Wollack, J. A., & Cohen, A. S. (2008). A speeded item response model with gradual process change. *Psychometrika, 73*, 65-87.

Goegebeur, Y., De Boeck, P., & Molenberghs, G. (2010). Person Fit for Test Speededness. *Methodology.*

Gulliksen, H. (1950). *Theory of mental tests*. New York: John Wiley.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika, 57*, 97-109.

Jin, K. Y., & Wang, W. C. (2014). Item response theory models for performance decline during testing. *Journal of Educational Measurement, 51*, 178-200.

Kingston, N. M., & Dorans, N. J. (1984). Item location effects and their implications for IRT equating and adaptive testing. *Applied Psychological Measurement, 8*, 147-154.

Klein Entink, R. H., Fox, J. P., & Van Der Linden, W. J. (2009). A multivariate multilevel approach to the modeling of accuracy and speed of test takers. *Psychometrika, 74*, 21-48.

Lawrence, Ida M. (1993) *The effect of test speededness on subgroup performance.* (Report No. ETS-RR-93-49). Princeton, NJ: Educational Testing Services.

Lee, G., Kolen, M. J., Frisbie, D. A., & Ankenmann, R. D. (2001). Comparison of dichotomous and polytomous item response models in equating scores from tests composed of testlets.*Applied Psychological Measurement, 25*, 357-372.

Lee, Y. H., & Ying, Z. (2015). A Mixture Cure-Rate Model for Responses and Response Times in Time-Limit Tests. *Psychometrika, 80*, 748-775.

Lord, F. M. (1956). A study of speed factors in tests and academic grades. *Psychometrika, 21*, 31-50.

Lord, F. M. (1980).*Applications of item response theory to practical testing problems*. Routledge.

Lu, Y., & Sireci, S. G. (2007). Validity issues in test speededness. *Educational Measurement: Issues and Practice, 26*, 29-37.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics, 21*, 1087-109

Meyers, J. L., Murphy, S., Goodman, J., & Turhan, A. (2012). *The impact of item position change on item parameters and common equating results under the 3PL model.* Paper presented at annual meeting of the National Council on Measurement in Education, Vancouver, BC.

Musch, J., & Broder, A. (1999). Test anxiety versus academic skills: A comparison of two alternative models for predicting performance in a statistics exam. *British Journal of Educational Psychology, 69*, 105-116.

Oshima, T. C. (1994). The effect of speededness on parameter estimation in item response theory.*Journal of Educational Measurement, 31*, 200-219.

Patelis, T. (2000). An Overview of Computer-Based Testing. Research Notes. RN-09. *College Entrance Examination Board* .

Rindler, S. E. (1979). Pitfalls in assessing test speededness. *Journal of Educational Measurement, 16*, 261-270.

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement, 14*, 271-282.

Sanderson, C., & Curtin, R. (2016). Armadillo: a template-based C++ library for linear algebra. *Journal of Open Source Software.*

Schmitt, T. A., Sass, D. A., Sullivan, J. R., & Walker, C. M. (2010). A Monte Carlo simulation investigating the validity and reliability of ability estimation in item response theory with speeded computer adaptive tests. *International Journal of Testing, 10*, 230-261.

Sireci, S. G., Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement, 28*, 237-247.

Spiegelhalter, D., Thomas, A., Best, N., & Gilks, W. (1996). BUGS 0.5* Examples Volume 2 (version ii). *MRC Biostatistics Unit.*

Suh, Y., & Bolt, D. M. (2010). Nested logit models for multiple-choice item response data. *Psychometrika, 75*, 454-473.

Suh, Y., Cho, S. J., & Wollack, J. A. (2012). A comparison of item calibration procedures in the presence of test speededness. *Journal of Educational Measurement, 49*, 285-311.

Swineford, F. (1974). *The test analysis manual.* (ETS SR 74-06). Princeton, NJ: Educational Testing Service.

van der Linden, W. J. (2006). A lognormal model for response times on test items. *Journal of Educational and Behavioral Statistics, 31*, 181-204.

van der Linden, W. J. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika, 72*, 287-308.

van der Linden, W. J., Breithaupt, K., Chuah, S. C., & Zhang, Y. (2007). Detecting differential speededness in multistage testing. *Journal of Educational Measurement, 44*, 117-130.

van der Linden, W. J. (2009). Conceptual issues in response—time modeling. *Journal of Educational Measurement, 46*, 247-272.

van der Linden, W. J. (2011). Setting time limits on tests. *Applied Psychological Measurement, 35*, 183-199.

Wang, C., & Xu, G. (2015). A mixture hierarchical model for response times and response accuracy. *British Journal of Mathematical and Statistical Psychology.*

Wise, S. L., & DeMars, C. E. (2006). An application of item response time: The effort-moderated IRT model. *Journal of Educational Measurement, 43*, 19-38.

Wise, S. L., Kingsbury, G. G., & Webb, N. L. (2015). Evaluating Content Alignment in Computerized Adaptive Testing. *Educational Measurement: Issues and Practice, 34*, 41-48.

Wollack, J. A., Cohen, A. S., & Wells, C. S. (2003). A method for maintaining scale stability in the presence of test speededness. *Journal of Educational Measurement, 40*, 307-330.

Wollack, J. A., Suh, Y., Bolt, D. M. (2007). *Using the Testlet Model to Mitigate Test Speededness Effects.* Paper presented at annual meeting of the National Council on Measurement in Education, Chicago, IL.

Yamamoto, K. (1995). *Estimating the effects of test length and test time on parameter estimation using the HYBRID model.* (Report No. ETS-RR-95-02). Princeton, NJ: Educational Testing Services.

Yamamoto, K., & Everson, H. (1997). Modeling the effects of test length and test time on parameter estimation using the HYBRID model. *Applications of latent trait and latent class models in the social sciences*, 89-98.

Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement, 30*, 187-213.

Appendix A

Appendix A

Study 1 Results Figures

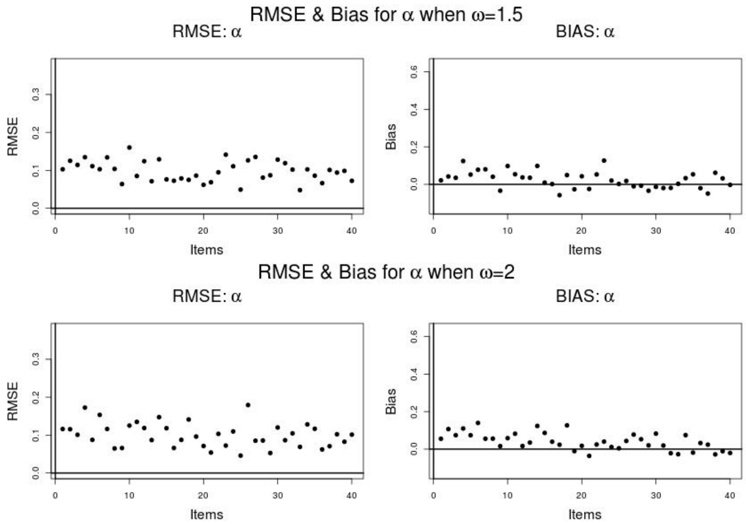

Figure A.1: Bias and RMSE for α, the estimation model with $\lambda = 0.4$, $\omega = 2$, and $\kappa = 0.2$, data generation with $\omega = 1.5$ (top) and normal (bottom) models, $N = 1000$, $J = 40$

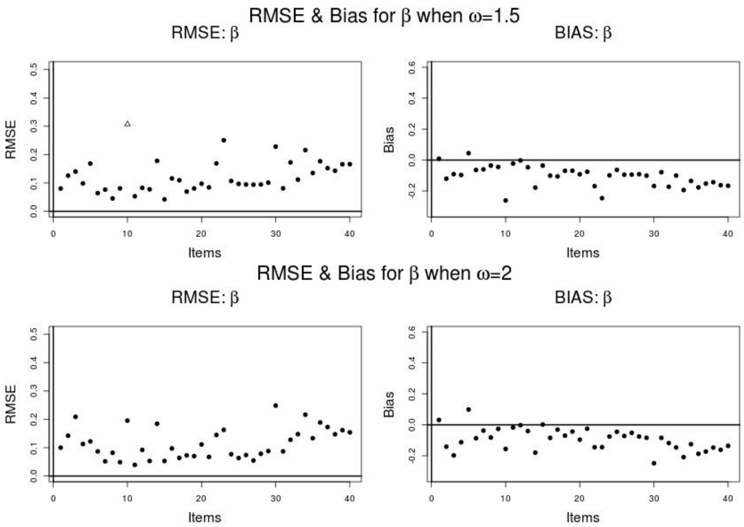

Figure A.2: Bias and RMSE for β, the estimation model with $\lambda = 0.4$, $\omega = 2$, and $\kappa = 0.2$, data generation with $\omega = 1.5$ (top) and normal (bottom) models, $N = 1000$, $J = 40$

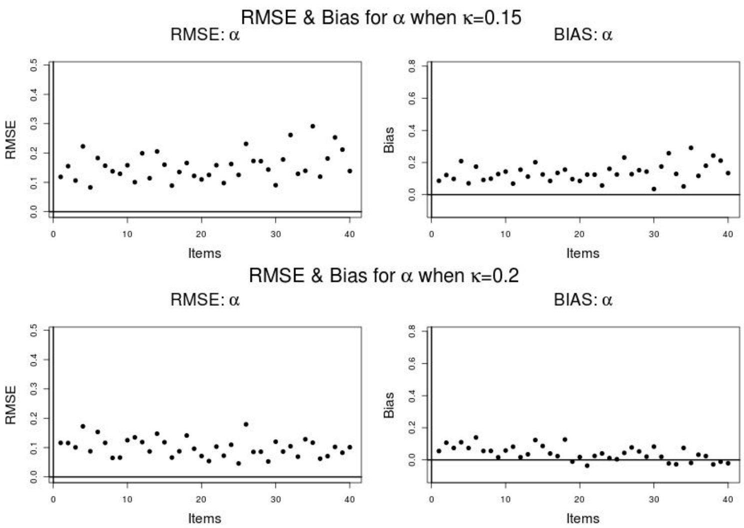

Figure A.3: Bias and RMSE for α, the estimation model with $\lambda = 0.4$, $\omega = 2$, and $\kappa = 0.2$, data generation with $\kappa = 0.15$ (top) and normal (bottom) models $N = 1000$, $J = 40$

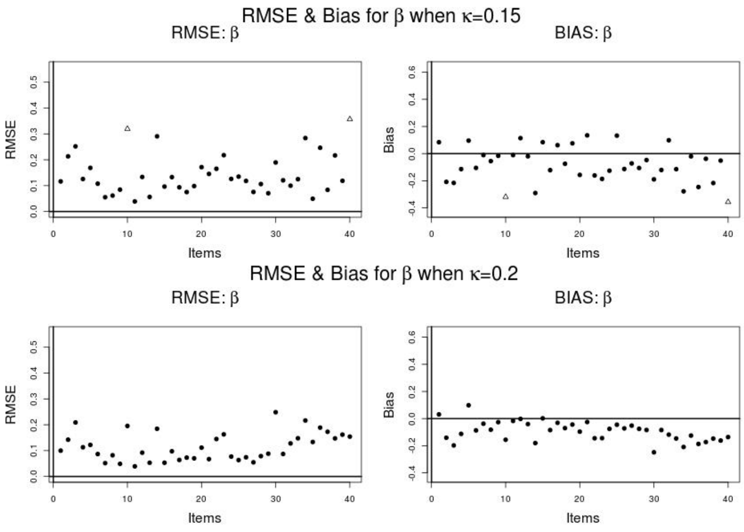

Figure A.4: Bias and RMSE for β, the estimation model with $\lambda = 0.4$, $\omega = 2$, and $\kappa = 0.2$, data generation with $\kappa = 0.15$ (top) and normal (bottom) models, $N = 1000$, $J = 40$

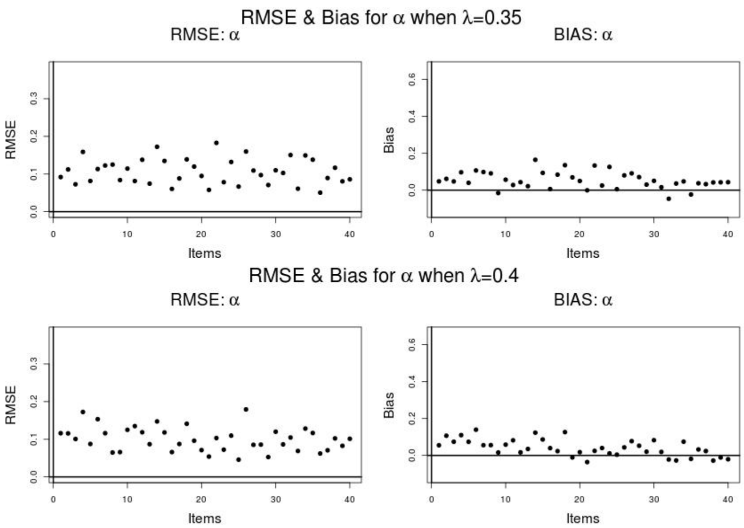

Figure A.5: Bias and RMSE for α, the estimation model with $\lambda = 0.4$, $\omega = 2$, and $\kappa = 0.2$, data generation with $\lambda = 0.35$ (top) and normal (bottom) models, $N = 1000$, $J = 40$

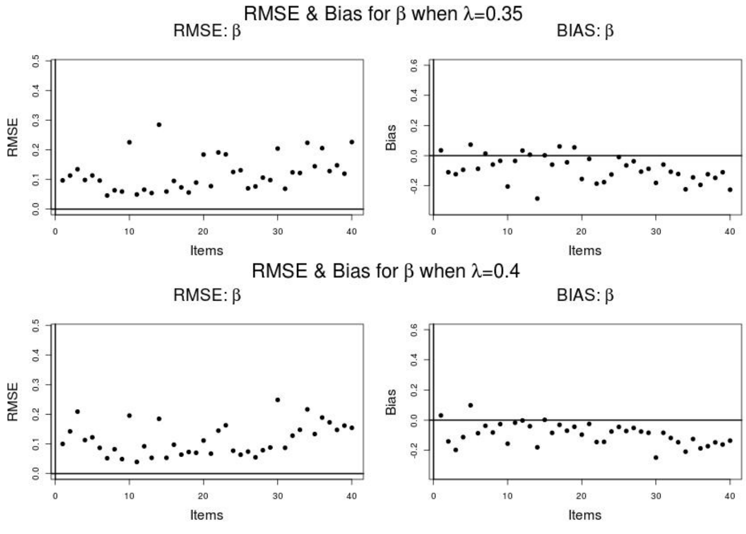

Figure A.6: Bias and RMSE for β, the estimation model with $\lambda = 0.4$, $\omega = 2$, and $\kappa = 0.2$, data generation with $\lambda = 0.35$ (top) and normal (bottom) models, $N = 1000$, $J = 40$

Appendix B

Study 2 Results Tables & Figures

Table B.1: IRT model parameters: Generated Model M2PLMix Estimated Models the M2PLMix and 2PL, $N = 500$, $J = 40$, $\lambda = 0.2$.

Generating Model M2PLMix $N = 500$, $J = 40$, $\lambda = 0.2$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.03	0.11	0.15	0.2	-0.32	0.46	0.44	0.34
	β	-0.08	0.1	0.14	0.21	-0.25	0.31	0.29	0.3
	θ	-0.03	0.17	0.28	0.12	-0.01	0.51	0.41	0.28
Ordered	α	-0.05	0.1	0.16	0.22	-0.25	0.23	0.31	0.21
	β	-0.04	0.09	0.13	0.22	-0.11	0.14	0.18	0.18
	θ	-0.03	0.23	0.29	0.17	-0.02	0.48	0.37	0.31
Generating Model M2PLMix N = 500, J = 40, $\lambda = 0.4$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.02	0.1	0.15	0.2	-0.22	0.37	0.35	0.25
	β	-0.07	0.08	0.13	0.21	-0.2	0.24	0.24	0.25
	θ	-0.02	0.16	0.27	0.11	-0.01	0.46	0.37	0.28
Ordered	α	-0.04	0.07	0.16	0.19	-0.19	0.16	0.25	0.16
	β	-0.05	0.08	0.13	0.23	-0.1	0.09	0.15	0.2
	θ	-0.03	0.21	0.28	0.16	-0.02	0.4	0.32	0.28
Generating Model M2PLMix N = 500, J = 40, $\lambda = 0.6$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.01	0.07	0.14	0.18	-0.14	0.22	0.24	0.18
	β	-0.05	0.07	0.12	0.22	-0.14	0.18	0.19	0.21
	θ	-0.02	0.14	0.26	0.1	-0.02	0.37	0.32	0.21
Ordered	α	-0.02	0.07	0.14	0.19	-0.13	0.12	0.19	0.18
	β	-0.02	0.06	0.11	0.19	-0.05	0.06	0.11	0.16
	θ	-0.03	0.13	0.25	0.1	-0.03	0.24	0.27	0.13

Table B.2: Location of Speededness: Generated Model M2PLMix Estimated Models the M2PLMix and 2PL, $N = 500$, $J = 40$, $\lambda = 0.2$.

Generating Model M2PLMix $N = 500$, $J = 40$, $\lambda = 0.2$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.57	0.22	1	-0.24	-0.33	0.57	0	1	-0.39	-0.43
Ordered	δ	0.57	0.12	1	-0.24	-0.38	0.57	0	1	-0.42	-0.43
Generating Model M2PLMix N = 500, J = 40, $\lambda = 0.4$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.68	0.22	1	-0.13	-0.25	0.68	0	1	-0.25	-0.32
Ordered	δ	0.67	0.09	1	-0.27	-0.3	0.67	0	1	-0.35	-0.33
Generating Model M2PLMix N = 500, J = 40, $\lambda = 0.6$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.82	0.25	1	-0.11	-0.13	0.82	0	1	-0.18	-0.18
Ordered	δ	0.8	0.08	1	-0.14	-0.19	0.8	0	1	-0.19	-0.2

Table B.3: IRT model parameters: Generated Model M2PLMix Estimated Models the M2PLMix and 2PL, $N = 500$, $J = 80$, $\lambda = 0.2$.

Generating Model M2PLMix $N = 500$, $J = 80$, $\lambda = 0.2$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.08	0.06	0.14	0.18	-0.24	0.47	0.44	0.27
	β	-0.05	0.13	0.14	0.22	-0.27	0.33	0.32	0.32
	θ	-0.03	0.09	0.2	0.11	0.01	0.61	0.45	0.38
Ordered	α	0.06	0.06	0.15	0.17	-0.25	0.26	0.32	0.2
	β	-0.02	0.12	0.13	0.23	-0.09	0.13	0.16	0.19
	θ	-0.02	0.09	0.2	0.14	-0.01	0.45	0.32	0.3
Generating Model M2PLMix N = 500, J = 80, $\lambda = 0.4$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.08	0.05	0.14	0.18	-0.21	0.31	0.32	0.21
	β	-0.04	0.12	0.14	0.21	-0.2	0.26	0.25	0.27
	θ	-0.03	0.09	0.2	0.12	0	0.52	0.36	0.35
Ordered	α	0.06	0.06	0.14	0.19	-0.17	0.19	0.24	0.17
	β	-0.02	0.12	0.13	0.22	-0.07	0.1	0.14	0.18
	θ	-0.02	0.09	0.2	0.13	-0.01	0.37	0.28	0.25
Generating Model M2PLMix N = 500, J = 80, $\lambda = 0.6$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.07	0.05	0.14	0.17	-0.1	0.26	0.26	0.17
	β	-0.04	0.11	0.13	0.24	-0.16	0.21	0.21	0.22
	θ	-0.02	0.07	0.19	0.13	0	0.44	0.31	0.3
Ordered	α	0.08	0.06	0.14	0.2	-0.08	0.11	0.16	0.18
	β	-0.03	0.12	0.13	0.23	-0.04	0.06	0.1	0.18
	θ	-0.03	0.11	0.2	0.14	-0.02	0.28	0.22	0.23

Table B.4: Location of Speededness: Generated Model M2PLMix Estimated Models the M2PLMix and 2PL, $N = 500$, $J = 80$, $\lambda = 0.2$.

Generating Model M2PLMix $N = 500$, $J = 80$, $\lambda = 0.2$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.6	0.6	1	-0.02	-0.16	0.59	0	1	-0.38	-0.41
Ordered	δ	0.58	0.31	1	-0.15	-0.29	0.58	0	1	-0.45	-0.42
Generating Model M2PLMix N = 500, J = 80, $\lambda = 0.4$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.7	0.56	1	-0.02	-0.13	0.7	0	1	-0.34	-0.3
Ordered	δ	0.68	0.32	1	-0.09	-0.21	0.68	0	1	-0.29	-0.32
Generating Model M2PLMix N = 500, J = 80, $\lambda = 0.6$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.8	0.63	1	-0.03	-0.08	0.79	0	1	-0.19	-0.21
Ordered	δ	0.81	0.21	1	-0.08	-0.14	0.82	0	1	-0.16	-0.18

Table B.5: IRT model parameters: Generated Model M2PLMix Estimated Models the M2PLMix and 2PL, $N = 1000$, $J = 20$, $\lambda = 0.2$.

Generating Model M2PLMix $N = 1000$, $J = 20$, $\lambda = 0.2$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.15	0.15	0.19	0.2	-0.22	0.3	0.3	0.26
	β	-0.16	0.15	0.18	0.22	-0.25	0.23	0.27	0.26
	θ	-0.02	0.36	0.38	0.18	-0.02	0.46	0.42	0.22
Ordered	α	-0.11	0.09	0.15	0.2	-0.19	0.16	0.23	0.19
	β	-0.06	0.04	0.1	0.17	-0.09	0.08	0.13	0.19
	θ	-0.03	0.32	0.36	0.15	-0.02	0.35	0.36	0.18
Generating Model M2PLMix N = 1000, J = 20, $\lambda = 0.4$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.15	0.11	0.17	0.17	-0.23	0.23	0.27	0.21
	β	-0.1	0.11	0.13	0.16	-0.17	0.19	0.2	0.22
	θ	-0.03	0.31	0.36	0.15	-0.02	0.4	0.38	0.19
Ordered	α	-0.09	0.08	0.13	0.2	-0.15	0.13	0.18	0.18
	β	-0.04	0.05	0.09	0.1	-0.07	0.08	0.11	0.1
	θ	-0.03	0.31	0.35	0.16	-0.03	0.33	0.35	0.17
Generating Model M2PLMix N = 1000, J = 20, $\lambda = 0.6$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.09	0.09	0.14	0.22	-0.13	0.16	0.18	0.2
	β	-0.08	0.07	0.12	0.12	-0.13	0.13	0.16	0.18
	θ	-0.03	0.3	0.35	0.15	-0.02	0.37	0.37	0.2
Ordered	α	-0.07	0.06	0.13	0.18	-0.11	0.09	0.15	0.19
	β	-0.05	0.04	0.09	0.13	-0.06	0.05	0.1	0.08
	θ	-0.02	0.31	0.35	0.16	-0.02	0.33	0.36	0.17

Table B.6: Location of Speededness: Generated Model M2PLMix Estimated Models the M2PLMix and 2PL, $N = 1000$, $J = 20$, $\lambda = 0.2$.

Generating Model M2PLMix $N = 1000$, $J = 20$, $\lambda = 0.2$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.57	0.02	1	-0.34	-0.42	0.57	0	1	-0.35	-0.43
Ordered	δ	0.61	0.01	1	-0.32	-0.39	0.61	0	1	-0.33	-0.39
Generating Model M2PLMix N = 1000, J = 20, $\lambda = 0.4$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.72	0.01	1	-0.25	-0.28	0.72	0	1	-0.26	-0.28
Ordered	δ	0.7	0	1	-0.25	-0.3	0.7	0	1	-0.26	-0.3
Generating Model M2PLMix N = 1000, J = 20, $\lambda = 0.6$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.81	0.01	1	-0.15	-0.19	0.81	0	1	-0.16	-0.19
Ordered	δ	0.79	0	1	-0.21	-0.21	0.79	0	1	-0.21	-0.21

Table B.7: IRT model parameters: Generated Model M2PLMix Estimated Models the M2PLMix and 2PL, $N = 1000$, $J = 80$, $\lambda = 0.2$.

Generating Model M2PLMix $N = 1000$, $J = 80$, $\lambda = 0.2$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.02	0.05	0.1	0.15	-0.36	0.5	0.47	0.4
	β	-0.03	0.05	0.08	0.15	-0.23	0.34	0.28	0.33
	θ	-0.03	0.1	0.2	0.15	0	0.51	0.39	0.31
Ordered	α	-0.01	0.07	0.11	0.17	-0.23	0.23	0.29	0.21
	β	-0.02	0.06	0.08	0.16	-0.08	0.11	0.14	0.16
	θ	-0.02	0.13	0.2	0.16	-0.02	0.4	0.29	0.28
Generating Model M2PLMix N = 1000, J = 80, $\lambda = 0.4$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.02	0.04	0.1	0.13	-0.19	0.33	0.31	0.25
	β	-0.04	0.06	0.09	0.16	-0.2	0.26	0.23	0.28
	θ	-0.03	0.08	0.19	0.13	-0.01	0.48	0.35	0.3
Ordered	α	0	0.06	0.1	0.17	-0.16	0.17	0.21	0.18
	β	-0.01	0.05	0.08	0.15	-0.06	0.09	0.11	0.14
	θ	-0.02	0.08	0.19	0.14	-0.02	0.33	0.25	0.23
Generating Model M2PLMix N = 1000, J = 80, $\lambda = 0.6$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.02	0.04	0.09	0.11	-0.16	0.26	0.26	0.21
	β	-0.04	0.04	0.08	0.14	-0.18	0.23	0.21	0.24
	θ	-0.02	0.08	0.19	0.14	-0.01	0.46	0.33	0.31
Ordered	α	0.01	0.04	0.09	0.14	-0.12	0.12	0.16	0.2
	β	-0.03	0.05	0.08	0.16	-0.05	0.06	0.09	0.1
	θ	-0.03	0.09	0.19	0.15	-0.02	0.3	0.23	0.23

Table B.8: Location of Speededness: Generated Model M2PLMix Estimated Models the M2PLMix and 2PL, $N = 1000$, $J = 80$, $\lambda = 0.2$.

Generating Model M2PLMix $N = 1000$, $J = 80$, $\lambda = 0.2$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.61	0.59	1	-0.06	-0.16	0.6	0	1	-0.49	-0.4
Ordered	δ	0.57	0.25	1	-0.2	-0.32	0.57	0	1	-0.45	-0.43
Generating Model M2PLMix N = 1000, J = 80, $\lambda = 0.4$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.7	0.56	1	-0.03	-0.13	0.69	0	1	-0.28	-0.31
Ordered	δ	0.71	0.25	1	-0.11	-0.22	0.71	0	1	-0.26	-0.29
Generating Model M2PLMix N = 1000, J = 80, $\lambda = 0.6$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.79	0.59	1	-0.05	-0.09	0.78	0	1	-0.27	-0.22
Ordered	δ	0.79	0.28	1	-0.08	-0.15	0.8	0	1	-0.2	-0.2

Table B.9: IRT model parameters: Generated Model M2PLMix Estimated Models the M2PLMix and 2PL, $N = 2000$, $J = 20$, $\lambda = 0.2$.

Generating Model M2PLMix $N = 2000$, $J = 20$, $\lambda = 0.2$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.18	0.12	0.19	0.18	-0.26	0.27	0.3	0.26
	β	-0.13	0.14	0.15	0.16	-0.23	0.24	0.24	0.26
	θ	-0.03	0.35	0.37	0.16	-0.02	0.44	0.41	0.21
Ordered	α	-0.13	0.09	0.15	0.18	-0.19	0.16	0.22	0.19
	β	-0.05	0.06	0.09	0	-0.09	0.09	0.13	0.07
	θ	-0.03	0.34	0.36	0.16	-0.03	0.37	0.37	0.19
Generating Model M2PLMix N = 2000, J = 20, $\lambda = 0.4$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.14	0.1	0.15	0.19	-0.18	0.21	0.22	0.22
	β	-0.11	0.12	0.13	0.15	-0.18	0.19	0.2	0.22
	θ	-0.03	0.33	0.36	0.16	-0.03	0.41	0.39	0.2
Ordered	α	-0.1	0.07	0.13	0.18	-0.15	0.12	0.17	0.18
	β	-0.04	0.04	0.07	0	-0.07	0.07	0.1	0
	θ	-0.03	0.31	0.35	0.15	-0.03	0.34	0.36	0.17
Generating Model M2PLMix N = 2000, J = 20, $\lambda = 0.6$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.13	0.11	0.15	0.18	-0.17	0.19	0.21	0.21
	β	-0.08	0.11	0.11	0.13	-0.14	0.16	0.16	0.17
	θ	-0.03	0.31	0.35	0.15	-0.03	0.37	0.37	0.19
Ordered	α	-0.07	0.06	0.1	0.17	-0.1	0.09	0.13	0.19
	β	-0.04	0.03	0.07	0	-0.06	0.05	0.09	0.07
	θ	-0.03	0.28	0.34	0.13	-0.03	0.29	0.34	0.14

Table B.10: Location of Speededness: Generated Model M2PLMix Estimated Models the M2PLMix and 2PL, $N = 2000$, $J = 20$, $\lambda = 0.2$.

Generating Model M2PLMix $N = 2000$, $J = 20$, $\lambda = 0.2$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.58	0.02	1	-0.37	-0.41	0.58	0	1	-0.39	-0.42
Ordered	δ	0.61	0.01	1	-0.33	-0.39	0.61	0	1	-0.34	-0.39
Generating Model M2PLMix N = 2000, J = 20, $\lambda = 0.4$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.71	0.01	1	-0.27	-0.28	0.71	0	1	-0.27	-0.29
Ordered	δ	0.7	0	1	-0.25	-0.3	0.7	0	1	-0.25	-0.3
Generating Model M2PLMix N = 2000, J = 20, $\lambda = 0.6$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.79	0.01	1	-0.19	-0.2	0.79	0	1	-0.19	-0.21
Ordered	δ	0.8	0	1	-0.18	-0.2	0.8	0	1	-0.18	-0.2

Table B.11: IRT model parameters: Generated Model M2PLMix Estimated Models the M2PLMix and 2PL, $N = 2000$, $J = 40$, $\lambda = 0.2$.

Generating Model M2PLMix $N = 2000$, $J = 40$, $\lambda = 0.2$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.06	0.09	0.1	0.14	-0.32	0.47	0.44	0.38
	β	-0.07	0.09	0.09	0.12	-0.26	0.31	0.28	0.33
	θ	-0.03	0.19	0.28	0.12	-0.02	0.53	0.42	0.29
Ordered	α	-0.09	0.08	0.11	0.16	-0.21	0.21	0.26	0.21
	β	-0.02	0.04	0.06	0	-0.09	0.11	0.13	0.09
	θ	-0.03	0.2	0.28	0.13	-0.02	0.41	0.33	0.25
Generating Model M2PLMix N = 2000, J = 40, $\lambda = 0.4$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.06	0.08	0.1	0.14	-0.23	0.34	0.32	0.27
	β	-0.06	0.08	0.09	0.15	-0.21	0.26	0.24	0.28
	θ	-0.02	0.18	0.27	0.12	-0.02	0.47	0.38	0.27
Ordered	α	-0.09	0.07	0.11	0.17	-0.17	0.16	0.21	0.18
	β	-0.02	0.03	0.06	0	-0.07	0.08	0.1	0.05
	θ	-0.02	0.18	0.27	0.12	-0.02	0.33	0.3	0.2
Generating Model M2PLMix N = 2000, J = 40, $\lambda = 0.6$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.03	0.04	0.08	0.1	-0.14	0.2	0.2	0.21
	β	-0.04	0.04	0.06	0.08	-0.16	0.18	0.17	0.21
	θ	-0.02	0.16	0.26	0.11	-0.02	0.39	0.33	0.23
Ordered	α	-0.05	0.05	0.09	0.15	-0.11	0.1	0.14	0.19
	β	-0.03	0.02	0.06	0	-0.05	0.06	0.08	0.05
	θ	-0.03	0.17	0.26	0.12	-0.02	0.28	0.28	0.17

Table B.12: Location of Speededness: Generated Model M2PLMix Estimated Models the M2PLMix and 2PL, $N = 2000$, $J = 40$, $\lambda = 0.2$.

Generating Model M2PLMix $N = 2000$, $J = 40$, $\lambda = 0.2$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.6	0.25	1	-0.23	-0.31	0.59	0	1	-0.4	-0.41
Ordered	δ	0.61	0.11	1	-0.25	-0.35	0.61	0	1	-0.37	-0.39
Generating Model M2PLMix N = 2000, J = 40, $\lambda = 0.4$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.69	0.24	1	-0.16	-0.24	0.69	0	1	-0.28	-0.31
Ordered	δ	0.7	0.09	1	-0.23	-0.28	0.7	0	1	-0.3	-0.3
Generating Model M2PLMix N = 2000, J = 40, $\lambda = 0.6$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.8	0.26	1	-0.09	-0.15	0.8	0	1	-0.17	-0.2
Ordered	δ	0.8	0.07	1	-0.15	-0.19	0.8	0	1	-0.19	-0.2

Table B.13: Model Fit Data Generation Model M2PLMix Model, $N = 500$, Item Difficulty Random

Model Fit Data Generation under the M2PLMix Model, $N = 500$, Item Difficulty Random					
	Estimating Model M2PLMix and M2PLMix				
	-2LL	AIC	AICA	BIC	Adj. BIC
E=M2PLMix, $J = 40$, $\lambda = 0.2$	17009.14	19169.14	15150.28	23720.92	20292.93
E=2PL, $J = 40$, $\lambda = 0.2$	17207.68	18367.68	10047.18	20812.15	18971.2
E=M2PLMix, $J = 80$, $\lambda = 0.2$	32304.12	34624.12	30549.21	39513.07	35831.16
E=2PL, $J = 80$, $\lambda = 0.2$	35374.3	36694.3	31274.92	39475.94	37381.06
E=M2PLMix, $J = 40$, $\lambda = 0.4$	17234.69	19394.69	15375.83	23946.47	20518.48
E=2PL, $J = 40$, $\lambda = 0.4$	17538.92	18698.92	10378.43	21143.4	19302.44
E=M2PLMix, $J = 80$, $\lambda = 0.4$	33523.7	35843.7	31768.79	40732.65	37050.74
E=2PL, $J = 80$, $\lambda = 0.4$	35551.81	36871.81	31452.43	39653.45	37558.57
E=M2PLMix, $J = 40$, $\lambda = 0.6$	17686.9	19846.9	15828.03	24398.67	20970.69
E=2PL, $J = 40$, $\lambda = 0.6$	17855.49	19015.49	10695	21459.96	19619.01
E=M2PLMix, $J = 80$, $\lambda = 0.6$	34494.91	36814.91	32739.99	41703.85	38021.94
E=2PL, $J = 80$, $\lambda = 0.6$	36354.87	37674.87	32255.5	40456.52	38361.64

Table B.14: Model Fit Data Generation Model M2PLMix Model, N = 500, Item Difficulty Ordered

Model Fit Data Generation under the M2PLMix Model, $N = 500$, Item Difficulty Ordered					
	Estimating Model M2PLMix and M2PLMix				
	-2LL	AIC	AICA	BIC	Adj. BIC
E=M2PLMix, $J = 40$, $\lambda = 0.2$	16565.95	18725.95	14707.09	23277.73	19849.74
E=2PL, $J = 40$, $\lambda = 0.2$	16688.72	17848.72	9528.23	20293.19	18452.24
E=M2PLMix, $J = 80$, $\lambda = 0.2$	32548.54	34868.54	30793.62	39757.49	36075.57
E=2PL, $J = 80$, $\lambda = 0.2$	33745.83	35065.83	29646.45	37847.47	35752.59
E=M2PLMix, $J = 40$, $\lambda = 0.4$	17093.52	19253.52	15234.66	23805.3	20377.31
E=2PL, $J = 40$, $\lambda = 0.4$	17093.49	18253.49	9932.99	20697.96	18857
E=M2PLMix, $J = 80$, $\lambda = 0.4$	33617.74	35937.74	31862.83	40826.69	37144.78
E=2PL, $J = 80$, $\lambda = 0.4$	34545.15	35865.15	30445.77	38646.79	36551.91
E=M2PLMix, $J = 40$, $\lambda = 0.6$	17492.95	19652.95	15634.09	24204.73	20776.74
E=2PL, $J = 40$, $\lambda = 0.6$	17503.1	18663.1	10342.61	21107.58	19266.62
E=M2PLMix, $J = 80$, $\lambda = 0.6$	35393.97	37713.97	33639.05	42602.92	38921
E=2PL, $J = 80$, $\lambda = 0.6$	35740.44	37060.44	31641.06	39842.08	37747.2

Table B.15: Model Fit Data Generation Model M2PLMix Model, N = 1000, Item Difficulty Random

Model Fit Data Generation under the M2PLMix Model, $N = 1000$, Item Difficulty Random					
	Estimating Model M2PLMix and M2PLMix				
	-2LL	AIC	AICA	BIC	Adj. BIC
E=M2PLMix, $J = 20$, $\lambda = 0.2$	17623.46	21703.46	13704.15	31715.28	25236.12
E=2PL, $J = 20$, $\lambda = 0.2$	17239.91	19319.91	-33491.8	24423.98	21120.88
E=M2PLMix, $J = 80$, $\lambda = 0.2$	64373.04	68693.04	60652.11	79293.79	72433.51
E=2PL, $J = 80$, $\lambda = 0.2$	69071.77	71391.77	54661.83	77084.76	73400.54
E=M2PLMix, $J = 20$, $\lambda = 0.4$	17551.91	21631.91	13632.6	31643.73	25164.58
E=2PL, $J = 20$, $\lambda = 0.4$	17212.13	19292.13	-33519.58	24396.19	21093.1
E=M2PLMix, $J = 80$, $\lambda = 0.4$	67313.64	71633.64	63592.71	82234.39	75374.11
E=2PL, $J = 80$, $\lambda = 0.4$	71690.14	74010.14	57280.2	79703.13	76018.91
E=M2PLMix, $J = 20$, $\lambda = 0.6$	17736.95	21816.95	13817.64	31828.77	25349.62
E=2PL, $J = 20$, $\lambda = 0.6$	17541.85	19621.85	-33189.85	24725.92	21422.82
E=M2PLMix, $J = 80$, $\lambda = 0.6$	68759.44	73079.44	65038.51	83680.19	76819.91
E=2PL, $J = 80$, $\lambda = 0.6$	72255	74575	57845.06	80267.99	76583.77

Table B.16: Model Fit Data Generation Model M2PLMix Model, N = 1000, Item Difficulty Ordered

Model Fit Data Generation under the M2PLMix Model, $N = 1000$, Item Difficulty Ordered					
	Estimating Model M2PLMix and M2PLMix				
	-2LL	AIC	AICA	BIC	Adj. BIC
E=M2PLMix, $J = 20$, $\lambda = 0.2$	16889.35	20969.35	12970.04	30981.17	24502.02
E=2PL, $J = 20$, $\lambda = 0.2$	16695.11	18775.11	-34036.59	23879.18	20576.08
E=M2PLMix, $J = 80$, $\lambda = 0.2$	66267.82	70587.82	62546.89	81188.57	74328.29
E=2PL, $J = 80$, $\lambda = 0.2$	68125.58	70445.58	53715.64	76138.58	72454.35
E=M2PLMix, $J = 20$, $\lambda = 0.4$	17199.68	21279.68	13280.37	31291.5	24812.34
E=2PL, $J = 20$, $\lambda = 0.4$	17054.87	19134.87	-33676.84	24238.93	20935.84
E=M2PLMix, $J = 80$, $\lambda = 0.4$	68601.22	72921.22	64880.29	83521.97	76661.69
E=2PL, $J = 80$, $\lambda = 0.4$	69999.42	72319.42	55589.48	78012.41	74328.19
E=M2PLMix, $J = 20$, $\lambda = 0.6$	17315	21395	13395.7	31406.83	24927.67
E=2PL, $J = 20$, $\lambda = 0.6$	17224.53	19304.53	-33507.18	24408.6	21105.5
E=M2PLMix, $J = 80$, $\lambda = 0.6$	69781.2	74101.2	66060.27	84701.96	77841.68
E=2PL, $J = 80$, $\lambda = 0.6$	70844.45	73164.45	56434.51	78857.45	75173.22

Table B.17: Model Fit Data Generation Model M2PLMix Model, N = 2000, Item Difficulty Random

Model Fit Data Generation under the M2PLMix Model, $N = 2000$, Item Difficulty Random					
	Estimating Model M2PLMix and M2PLMix				
	-2LL	AIC	AICA	BIC	Adj. BIC
E=M2PLMix, $J = 20$, $\lambda = 0.2$	34969.05	43049.05	27051.36	65676.7	52841.4
E=2PL, $J = 20$, $\lambda = 0.2$	34152.07	38232.07	-164872.32	49657.91	43176.72
E=M2PLMix, $J = 40$, $\lambda = 0.2$	67861.33	76021.33	60018.94	98873.01	85910.63
E=2PL, $J = 40$, $\lambda = 0.2$	68956.06	73116.06	-33759.99	84765.94	78157.66
E=M2PLMix, $J = 20$, $\lambda = 0.4$	35316.39	43396.39	27398.7	66024.04	53188.74
E=2PL, $J = 20$, $\lambda = 0.4$	34741.24	38821.24	-164283.15	50247.08	43765.89
E=M2PLMix, $J = 40$, $\lambda = 0.4$	69219.95	77379.95	61377.56	100231.63	87269.25
E=2PL, $J = 40$, $\lambda = 0.4$	70409.19	74569.19	-32306.86	86219.07	79610.8
E=M2PLMix, $J = 20$, $\lambda = 0.6$	35310.27	43390.27	27392.59	66017.92	53182.62
E=2PL, $J = 20$, $\lambda = 0.6$	34862.19	38942.19	-164162.2	50368.03	43886.84
E=M2PLMix, $J = 40$, $\lambda = 0.6$	70730.02	78890.02	62887.63	101741.7	88779.32
E=2PL, $J = 40$, $\lambda = 0.6$	71836.04	75996.04	-30880.01	87645.91	81037.64

Table B.18: Model Fit Data Generation Model M2PLMix Model, $N = 2000$, Item Difficulty Ordered

Model Fit Data Generation under the M2PLMix Model, $N = 2000$, Item Difficulty Ordered						
		Estimating Model M2PLMix and M2PLMix				
		-2LL	AIC	AICA	BIC	Adj. BIC
	E=M2PLMix, $J = 20$, $\lambda = 0.2$	33922.59	42002.59	26004.91	64630.24	51794.94
	E=2PL, $J = 20$, $\lambda = 0.2$	33542.08	37622.08	-165482.31	49047.92	42566.73
	E=M2PLMix, $J = 40$, $\lambda = 0.2$	67337.42	75497.42	59495.04	98349.1	85386.72
	E=2PL, $J = 40$, $\lambda = 0.2$	67726.44	71886.44	-34989.61	83536.32	76928.05
	E=M2PLMix, $J = 20$, $\lambda = 0.4$	34324.85	42404.85	26407.16	65032.49	52197.19
	E=2PL, $J = 20$, $\lambda = 0.4$	34038.09	38118.09	-164986.3	49543.93	43062.74
	E=M2PLMix, $J = 40$, $\lambda = 0.4$	68737.08	76897.08	60894.7	99748.76	86786.38
	E=2PL, $J = 40$, $\lambda = 0.4$	68833.12	72993.12	-33882.93	84642.99	78034.72
	E=M2PLMix, $J = 20$, $\lambda = 0.6$	34765.6	42845.6	26847.91	65473.25	52637.95
	E=2PL, $J = 20$, $\lambda = 0.6$	34594.02	38674.02	-164430.37	50099.86	43618.67
	E=M2PLMix, $J = 40$, $\lambda = 0.6$	70404.72	78564.72	62562.34	101416.41	88454.02
	E=2PL, $J = 40$, $\lambda = 0.6$	70409.92	74569.92	-32306.13	86219.79	79611.52

Table B.19: IRT model parameters: Generated Model 2PL Estimated Models the M2PLMix and 2PL, $N = 500$, $J = 40$.

Generating Model 2PL $N = 500$, $J = 40$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.14	0.11	0.19	0.18	-0.03	0.05	0.14	0.22
	β	0.05	0.24	0.19	0.25	-0.01	0.03	0.09	0.18
	θ	-0.03	0.12	0.26	0.1	-0.02	0.12	0.24	0.08
Ordered	α	0.11	0.07	0.17	0.17	-0.01	0.04	0.13	0.19
	β	-0.02	0.11	0.13	0.21	-0.02	0.03	0.09	0.18
	θ	-0.03	0.12	0.24	0.09	-0.02	0.12	0.23	0.07
Generating Model 2PL N = 1000, J = 20									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.11	0.14	0.17	0.19	-0.03	0.04	0.1	0.18
	β	0.12	0.19	0.17	0.22	-0.01	0.03	0.07	0
	θ	-0.03	0.24	0.34	0.11	-0.03	0.19	0.32	0.08
Ordered	α	0.06	0.07	0.13	0.19	-0.03	0.03	0.11	0.18
	β	0.02	0.06	0.09	0.17	-0.03	0.02	0.07	0.16
	θ	-0.02	0.22	0.33	0.1	-0.02	0.19	0.32	0.09
Generating Model 2PL N = 2000, J = 20									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.12	0.13	0.15	0.18	-0.02	0.02	0.07	0.08
	β	0.1	0.2	0.15	0.21	-0.03	0.01	0.05	0
	θ	-0.03	0.24	0.34	0.11	-0.03	0.19	0.32	0.09
Ordered	α	0.07	0.06	0.11	0.12	-0.01	0.02	0.07	0
	β	0.03	0.04	0.07	0	-0.02	0.02	0.06	0
	θ	-0.03	0.23	0.33	0.1	-0.03	0.2	0.32	0.09

Table B.20: Location of Speededness: Generated Model 2PL Estimated Models the M2PLMix and 2PL, $N = 500$, $J = 40$.

Generating Model 2PL $N = 500$, $J = 40$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	1	NA	1	0	0	1	NA	1	0	0
Ordered	δ	1	NA	1	0	0	1	NA	1	0	0
Generating Model 2PL N = 1000, J = 20											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	1	NA	1	0	0	1	NA	1	0	0
Ordered	δ	0.99	NA	0.99	0.01	0.01	1	NA	1	0	0
Generating Model 2PL N = 2000, J = 20											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	1	NA	1	0	0	1	NA	1	0	0
Ordered	δ	0.99	NA	0.99	0.01	0.01	1	NA	1	0	0

Table B.21: IRT model parameters: Generated Model 2PL Estimated Models the M2PLMix and 2PL, $N = 500$, $J = 80$.

Generating Model 2PL $N = 500$, $J = 80$									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.13	0.06	0.17	0.16	0	0.04	0.12	0.19
	β	-0.02	0.17	0.17	0.23	-0.02	0.04	0.1	0.19
	θ	-0.02	0.08	0.19	0.11	-0.02	0.07	0.17	0.14
Ordered	α	0.12	0.05	0.16	0.16	-0.02	0.04	0.13	0.18
	β	-0.04	0.14	0.15	0.23	-0.02	0.04	0.1	0.17
	θ	-0.03	0.07	0.19	0.12	-0.02	0.07	0.17	0.13
Generating Model 2PL N = 1000, J = 80									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.08	0.04	0.11	0.16	-0.01	0.03	0.09	0.14
	β	-0.02	0.1	0.1	0.18	-0.02	0.02	0.07	0.1
	θ	-0.02	0.06	0.18	0.13	-0.02	0.07	0.17	0.14
Ordered	α	0.06	0.04	0.1	0.13	-0.01	0.03	0.09	0.14
	β	-0.03	0.06	0.08	0.17	-0.02	0.02	0.07	0.09
	θ	-0.03	0.06	0.17	0.12	-0.02	0.07	0.17	0.12
Generating Model 2PL N = 2000, J = 40									
		Estimating Model M2PLMix				Estimating Model 2PL			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.09	0.09	0.11	0.11	-0.02	0.02	0.07	0
	β	0.02	0.14	0.1	0.16	-0.02	0.02	0.05	0
	θ	-0.03	0.12	0.24	0.08	-0.02	0.12	0.23	0.08
Ordered	α	0.05	0.05	0.08	0.09	-0.02	0.03	0.07	0
	β	-0.01	0.03	0.05	0.09	-0.02	0.02	0.05	0.05
	θ	-0.03	0.12	0.24	0.09	-0.03	0.12	0.23	0.08

Table B.22: Location of Speededness: Generated Model 2PL Estimated Models the M2PLMix and 2PL, $N = 500$, $J = 80$.

Generating Model 2PL $N = 500$, $J = 80$											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	1	NA	1	0	0	1	NA	1	0	0
Ordered	δ	1	NA	1	0	0	1	NA	1	0	0
Generating Model 2PL N = 1000, J = 80											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	1	NA	1	0	0	1	NA	1	0	0
Ordered	δ	1	NA	1	0	0	1	NA	1	0	0
Generating Model 2PL N = 2000, J = 40											
		Estimating Model M2PLMix					Estimating Model 2PL				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	1	NA	1	0	0	1	NA	1	0	0
Ordered	δ	1	NA	1	0	0	1	NA	1	0	0

Table B.23: Model Fit Data Generation Model 2PL Model, $N = 500$, Item Difficulty Random

Model Fit Data Generation under the 2PL Model, $N = 500$, Item Difficulty Random					
	Estimating Model M2PLMix and M2PLMix				
	-2LL	AIC	AICA	BIC	Adj. BIC
E=M2PLMix, $J = 40$	18291.19	20451.19	16432.33	25002.97	21574.98
E=2PL, $J = 40$	18162.58	19322.58	11002.09	21767.05	19926.1
E=M2PLMix, $J = 80$	36918.02	39238.02	35163.11	44126.97	40445.06
E=2PL, $J = 80$	36883.01	38203.01	32783.63	40984.65	38889.77

Table B.24: Model Fit Data Generation Model 2PL Model, $N = 500$, Item Difficulty Ordered

Model Fit Data Generation under the 2PL Model, $N = 500$, Item Difficulty Ordered					
	Estimating Model M2PLMix and M2PLMix				
	-2LL	AIC	AICA	BIC	Adj. BIC
E=M2PLMix, $J = 40$	18287.73	20447.73	16428.87	24999.51	21571.52
E=2PL, $J = 40$	18222.02	19382.02	11061.53	21826.5	19985.54
E=M2PLMix, $J = 80$	36777.94	39097.94	35023.03	43986.89	40304.98
E=2PL, $J = 80$	36755.26	38075.26	32655.88	40856.9	38762.02

Table B.25: Model Fit Data Generation Model 2PL Model, $N = 1000$, Item Difficulty Random

Model Fit Data Generation under the 2PL Model, $N = 1000$, Item Difficulty Random					
	Estimating Model M2PLMix and M2PLMix				
	-2LL	AIC	AICA	BIC	Adj. BIC
E=M2PLMix, $J = 20$	18264.26	22344.26	14344.95	32356.08	25876.93
E=2PL, $J = 20$	17850.09	19930.09	-32881.61	25034.16	21731.06
E=M2PLMix, $J = 80$	73909.15	78229.15	70188.22	88829.9	81969.62
E=2PL, $J = 80$	73860.54	76180.54	59450.6	81873.54	78189.31

Table B.26: Model Fit Data Generation Model 2PL Model, $N = 1000$, Item Difficulty Ordered

Model Fit Data Generation under the 2PL Model, $N = 1000$, Item Difficulty Ordered					
	Estimating Model M2PLMix and M2PLMix				
	-2LL	AIC	AICA	BIC	Adj. BIC
E=M2PLMix, $J = 20$	17998.06	22078.06	14078.76	32089.89	25610.73
E=2PL, $J = 20$	17799.66	19879.66	-32932.05	24983.72	21680.63
E=M2PLMix, $J = 80$	73805.03	78125.03	70084.1	88725.78	81865.5
E=2PL, $J = 80$	73779.29	76099.29	59369.35	81792.29	78108.06

Table B.27: Model Fit Data Generation Model 2PL Model, $N = 2000$, Item Difficulty Random

Model Fit Data Generation under the 2PL Model, $N = 2000$, Item Difficulty Random					
	Estimating Model M2PLMix and M2PLMix				
	-2LL	AIC	AICA	BIC	Adj. BIC
E=M2PLMix, $J = 20$	36541.57	44621.57	28623.89	67249.22	54413.92
E=2PL, $J = 20$	35727.7	39807.7	-163296.69	51233.54	44752.35
E=M2PLMix, $J = 40$	73378.69	81538.69	65536.31	104390.37	91427.99
E=2PL, $J = 40$	73022.62	77182.62	-29693.43	88832.5	82224.22

Table B.28: Model Fit Data Generation Model 2PL Model, $N = 2000$, Item Difficulty Ordered

Model Fit Data Generation under the 2PL Model, $N = 2000$, Item Difficulty Ordered					
	Estimating Model M2PLMix and M2PLMix				
	-2LL	AIC	AICA	BIC	Adj. BIC
E=M2PLMix, $J = 20$	36237	44317	28319.31	66944.65	54109.35
E=2PL, $J = 20$	35860.89	39940.89	-163163.5	51366.73	44885.54
E=M2PLMix, $J = 40$	73201.03	81361.03	65358.64	104212.71	91250.33
E=2PL, $J = 40$	72977.98	77137.98	-29738.07	88787.85	82179.58

Appendix C

Study 3 Results Tables & Figures

Table C.1: IRT model parameters: Generated Model M2PLMix Estimated Models the M2PLMix and 2PLMix,$N = 500,\ J = 40,\ \lambda = 0.2$.

Generating Model M2PLMix $N = 500,\ J = 40,\ \lambda = 0.2$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.03	0.11	0.15	0.2	0.07	0.05	0.14	0.2
	β	-0.08	0.1	0.14	0.21	0.04	0.15	0.16	0.21
	θ	-0.03	0.17	0.28	0.12	-0.02	0.17	0.28	0.12
Ordered	α	-0.05	0.1	0.16	0.22	0.07	0.06	0.16	0.19
	β	-0.04	0.09	0.13	0.22	0.03	0.07	0.13	0.24
	θ	-0.03	0.23	0.29	0.17	-0.03	0.2	0.29	0.14
Generating Model M2PLMix N = 500, J = 40, $\lambda = 0.4$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.02	0.1	0.15	0.2	0.05	0.05	0.15	0.19
	β	-0.07	0.08	0.13	0.21	0	0.11	0.13	0.22
	θ	-0.02	0.16	0.27	0.11	-0.03	0.15	0.27	0.11
Ordered	α	-0.04	0.07	0.16	0.19	0.04	0.05	0.16	0.16
	β	-0.05	0.08	0.13	0.23	0	0.07	0.13	0.24
	θ	-0.03	0.21	0.28	0.16	-0.02	0.19	0.28	0.14
Generating Model M2PLMix N = 500, J = 40, $\lambda = 0.6$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.01	0.07	0.14	0.18	0.05	0.05	0.14	0.17
	β	-0.05	0.07	0.12	0.22	-0.01	0.09	0.12	0.21
	θ	-0.02	0.14	0.26	0.1	-0.02	0.13	0.26	0.09
Ordered	α	-0.02	0.07	0.14	0.19	0.02	0.05	0.13	0.18
	β	-0.02	0.06	0.11	0.19	0.01	0.05	0.11	0.19
	θ	-0.03	0.13	0.25	0.1	-0.03	0.13	0.26	0.1

Table C.2: Location of Speededness: Generated Model M2PLMix Estimated Models the M2PLMix and 2PLMix, $N = 500$, $J = 40$, $\lambda = 0.2$.

Generating Model M2PLMix $N = 500$, $J = 40$, $\lambda = 0.2$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.57	0.22	1	-0.24	-0.33	0.56	0.53	0.96	-0.07	-0.18
Ordered	δ	0.57	0.12	1	-0.24	-0.38	0.54	0.36	0.95	-0.07	-0.24
Generating Model M2PLMix N = 500, J = 40, $\lambda = 0.4$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.68	0.22	1	-0.13	-0.25	0.68	0.33	0.99	-0.09	-0.21
Ordered	δ	0.67	0.09	1	-0.27	-0.3	0.67	0.18	0.99	-0.18	-0.26
Generating Model M2PLMix N = 500, J = 40, $\lambda = 0.6$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.82	0.25	1	-0.11	-0.13	0.82	0.32	1	-0.09	-0.12
Ordered	δ	0.8	0.08	1	-0.14	-0.19	0.8	0.12	1	-0.13	-0.18

Table C.3: IRT model parameters: Generated Model M2PLMix Estimated Models the M2PLMix and 2PLMix, $N = 500$, $J = 80$, $\lambda = 0.2$.

Generating Model M2PLMix $N = 500$, $J = 80$, $\lambda = 0.2$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.08	0.06	0.14	0.18	0.11	0.05	0.15	0.18
	β	-0.05	0.13	0.14	0.22	-0.01	0.15	0.16	0.23
	θ	-0.03	0.09	0.2	0.11	-0.02	0.09	0.21	0.1
Ordered	α	0.06	0.06	0.15	0.17	0.13	0.07	0.18	0.16
	β	-0.02	0.12	0.13	0.23	-0.01	0.12	0.14	0.22
	θ	-0.02	0.09	0.2	0.14	-0.03	0.09	0.21	0.13
Generating Model M2PLMix N = 500, J = 80, $\lambda = 0.4$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.08	0.05	0.14	0.18	0.09	0.05	0.15	0.17
	β	-0.04	0.12	0.14	0.21	-0.03	0.13	0.14	0.22
	θ	-0.03	0.09	0.2	0.12	-0.03	0.09	0.2	0.13
Ordered	α	0.06	0.06	0.14	0.19	0.1	0.04	0.15	0.18
	β	-0.02	0.12	0.13	0.22	-0.02	0.11	0.13	0.23
	θ	-0.02	0.09	0.2	0.13	-0.03	0.08	0.2	0.12
Generating Model M2PLMix N = 500, J = 80, $\lambda = 0.6$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.07	0.05	0.14	0.17	0.09	0.05	0.14	0.17
	β	-0.04	0.11	0.13	0.24	-0.04	0.12	0.14	0.23
	θ	-0.02	0.07	0.19	0.13	-0.03	0.07	0.19	0.12
Ordered	α	0.08	0.06	0.14	0.2	0.1	0.05	0.15	0.2
	β	-0.03	0.12	0.13	0.23	-0.02	0.12	0.14	0.23
	θ	-0.03	0.11	0.2	0.14	-0.03	0.09	0.19	0.12

Table C.4: Location of Speededness: Generated Model M2PLMix Estimated Models the M2PLMix and 2PLMix, $N = 500$, $J = 80$, $\lambda = 0.2$.

Generating Model M2PLMix $N = 500$, $J = 80$, $\lambda = 0.2$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.6	0.6	1	-0.02	-0.16	0.6	0.68	0.99	0	-0.12
Ordered	δ	0.58	0.31	1	-0.15	-0.29	0.58	0.41	0.99	-0.09	-0.24
Generating Model M2PLMix N = 500, J = 80, $\lambda = 0.4$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.7	0.56	1	-0.02	-0.13	0.7	0.61	1	-0.02	-0.12
Ordered	δ	0.68	0.32	1	-0.09	-0.21	0.68	0.38	1	-0.06	-0.2
Generating Model M2PLMix N = 500, J = 80, $\lambda = 0.6$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.8	0.63	1	-0.03	-0.08	0.79	0.66	1	-0.01	-0.07
Ordered	δ	0.81	0.21	1	-0.08	-0.14	0.81	0.27	1	-0.06	-0.13

Table C.5: IRT model parameters: Generated Model M2PLMix Estimated Models the M2PLMix and 2PLMix, $N = 1000$, $J = 20$, $\lambda = 0.2$.

Generating Model M2PLMix $N = 1000$, $J = 20$, $\lambda = 0.2$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.15	0.15	0.19	0.2	0	0.05	0.11	0.17
	β	-0.16	0.15	0.18	0.22	0.08	0.13	0.14	0.2
	θ	-0.02	0.36	0.38	0.18	-0.03	0.32	0.38	0.16
Ordered	α	-0.11	0.09	0.15	0.2	0.01	0.04	0.11	0.19
	β	-0.06	0.04	0.1	0.17	0.12	0.11	0.15	0.2
	θ	-0.03	0.32	0.36	0.15	-0.03	0.31	0.38	0.14
Generating Model M2PLMix N = 1000, J = 20, $\lambda = 0.4$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.15	0.11	0.17	0.17	-0.04	0.05	0.1	0.17
	β	-0.1	0.11	0.13	0.16	0.04	0.08	0.1	0.14
	θ	-0.03	0.31	0.36	0.15	-0.03	0.29	0.36	0.14
Ordered	α	-0.09	0.08	0.13	0.2	-0.03	0.04	0.1	0.21
	β	-0.04	0.05	0.09	0.1	0.06	0.08	0.1	0.15
	θ	-0.03	0.31	0.35	0.16	-0.03	0.31	0.36	0.15
Generating Model M2PLMix N = 1000, J = 20, $\lambda = 0.6$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.09	0.09	0.14	0.22	-0.03	0.05	0.11	0.19
	β	-0.08	0.07	0.12	0.12	0	0.03	0.08	0.09
	θ	-0.03	0.3	0.35	0.15	-0.02	0.28	0.35	0.14
Ordered	α	-0.07	0.06	0.13	0.18	-0.05	0.04	0.11	0.18
	β	-0.05	0.04	0.09	0.13	0	0.04	0.08	0.14
	θ	-0.02	0.31	0.35	0.16	-0.02	0.3	0.35	0.14

Table C.6: Location of Speededness: Generated Model M2PLMix Estimated Models the M2PLMix and 2PLMix, $N = 1000$, $J = 20$, $\lambda = 0.2$.

Generating Model M2PLMix $N = 1000$, $J = 20$, $\lambda = 0.2$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.57	0.02	1	-0.34	-0.42	0.52	0.46	0.89	-0.07	-0.16
Ordered	δ	0.61	0.01	1	-0.32	-0.39	0.46	0.56	0.74	0.11	-0.01
Generating Model M2PLMix N = 1000, J = 20, $\lambda = 0.4$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.72	0.01	1	-0.25	-0.28	0.71	0.13	0.98	-0.2	-0.23
Ordered	δ	0.7	0	1	-0.25	-0.3	0.68	0.12	0.97	-0.16	-0.24
Generating Model M2PLMix N = 1000, J = 20, $\lambda = 0.6$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.81	0.01	1	-0.15	-0.19	0.81	0.05	1	-0.14	-0.18
Ordered	δ	0.79	0	1	-0.21	-0.21	0.79	0.01	1	-0.2	-0.21

Table C.7: IRT model parameters: Generated Model M2PLMix Estimated Models the M2PLMix and 2PLMix, $N = 1000$, $J = 80$, $\lambda = 0.2$.

Generating Model M2PLMix $N = 1000$, $J = 80$, $\lambda = 0.2$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.02	0.05	0.1	0.15	0.06	0.04	0.11	0.15
	β	-0.03	0.05	0.08	0.15	0	0.1	0.1	0.18
	θ	-0.03	0.1	0.2	0.15	-0.03	0.1	0.2	0.14
Ordered	α	-0.01	0.07	0.11	0.17	0.07	0.04	0.11	0.17
	β	-0.02	0.06	0.08	0.16	0	0.06	0.09	0.17
	θ	-0.02	0.13	0.2	0.16	-0.02	0.11	0.2	0.14
Generating Model M2PLMix N = 1000, J = 80, $\lambda = 0.4$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.02	0.04	0.1	0.13	0.05	0.03	0.1	0.13
	β	-0.04	0.06	0.09	0.16	-0.02	0.07	0.09	0.17
	θ	-0.03	0.08	0.19	0.13	-0.03	0.08	0.19	0.12
Ordered	α	0	0.06	0.1	0.17	0.05	0.04	0.1	0.16
	β	-0.01	0.05	0.08	0.15	-0.01	0.05	0.08	0.17
	θ	-0.02	0.08	0.19	0.14	-0.02	0.07	0.19	0.14
Generating Model M2PLMix N = 1000, J = 80, $\lambda = 0.6$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.02	0.04	0.09	0.11	0.04	0.03	0.1	0.13
	β	-0.04	0.04	0.08	0.14	-0.03	0.05	0.08	0.16
	θ	-0.02	0.08	0.19	0.14	-0.03	0.08	0.19	0.13
Ordered	α	0.01	0.04	0.09	0.14	0.03	0.03	0.1	0.13
	β	-0.03	0.05	0.08	0.16	-0.02	0.05	0.08	0.18
	θ	-0.03	0.09	0.19	0.15	-0.02	0.08	0.19	0.14

Table C.8: Location of Speededness: Generated Model M2PLMix Estimated Models the M2PLMix and 2PLMix, $N = 1000$, $J = 80$, $\lambda = 0.2$.

Generating Model M2PLMix $N = 1000$, $J = 80$, $\lambda = 0.2$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.61	0.59	1	-0.06	-0.16	0.61	0.69	0.99	-0.06	-0.12
Ordered	δ	0.57	0.25	1	-0.2	-0.32	0.57	0.33	0.99	-0.14	-0.28
Generating Model M2PLMix N = 1000, J = 80, $\lambda = 0.4$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.7	0.56	1	-0.03	-0.13	0.7	0.61	1	-0.03	-0.12
Ordered	δ	0.71	0.25	1	-0.11	-0.22	0.71	0.31	1	-0.08	-0.2
Generating Model M2PLMix N = 1000, J = 80, $\lambda = 0.6$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.79	0.59	1	-0.05	-0.09	0.79	0.63	1	-0.02	-0.08
Ordered	δ	0.79	0.28	1	-0.08	-0.15	0.79	0.32	1	-0.06	-0.14

Table C.9: IRT model parameters: Generated Model M2PLMix Estimated Models the M2PLMix and 2PLMix, $N = 2000$, $J = 20$, $\lambda = 0.2$.

Generating Model M2PLMix $N = 2000$, $J = 20$, $\lambda = 0.2$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.18	0.12	0.19	0.18	-0.02	0.05	0.08	0.11
	β	-0.13	0.14	0.15	0.16	0.11	0.13	0.14	0.15
	θ	-0.03	0.35	0.37	0.16	-0.03	0.32	0.38	0.14
Ordered	α	-0.13	0.09	0.15	0.18	-0.01	0.04	0.08	0
	β	-0.05	0.06	0.09	0	0.12	0.15	0.15	0.18
	θ	-0.03	0.34	0.36	0.16	-0.03	0.32	0.38	0.14
Generating Model M2PLMix N = 2000, J = 20, $\lambda = 0.4$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.14	0.1	0.15	0.19	-0.03	0.04	0.07	0.15
	β	-0.11	0.12	0.13	0.15	0.03	0.06	0.07	0.07
	θ	-0.03	0.33	0.36	0.16	-0.03	0.3	0.36	0.14
Ordered	α	-0.1	0.07	0.13	0.18	-0.04	0.03	0.08	0.12
	β	-0.04	0.04	0.07	0	0.06	0.09	0.09	0.15
	θ	-0.03	0.31	0.35	0.15	-0.03	0.3	0.36	0.14
Generating Model M2PLMix N = 2000, J = 20, $\lambda = 0.6$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.13	0.11	0.15	0.18	-0.07	0.05	0.1	0.12
	β	-0.08	0.11	0.11	0.13	-0.01	0.05	0.06	0
	θ	-0.03	0.31	0.35	0.15	-0.02	0.29	0.35	0.14
Ordered	α	-0.07	0.06	0.1	0.17	-0.05	0.04	0.09	0.11
	β	-0.04	0.03	0.07	0	0.01	0.05	0.06	0.1
	θ	-0.03	0.28	0.34	0.13	-0.03	0.27	0.34	0.12

Table C.10: Location of Speededness: Generated Model M2PLMix Estimated Models the M2PLMix and 2PLMix, $N = 2000$, $J = 20$, $\lambda = 0.2$.

Generating Model M2PLMix $N = 2000$, $J = 20$, $\lambda = 0.2$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.58	0.02	1	-0.37	-0.41	0.52	0.44	0.88	-0.11	-0.17
Ordered	δ	0.61	0.01	1	-0.33	-0.39	0.45	0.56	0.74	0.11	-0.01
Generating Model M2PLMix N = 2000, J = 20, $\lambda = 0.4$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.71	0.01	1	-0.27	-0.28	0.7	0.14	0.98	-0.21	-0.24
Ordered	δ	0.7	0	1	-0.25	-0.3	0.68	0.14	0.97	-0.14	-0.23
Generating Model M2PLMix N = 2000, J = 20, $\lambda = 0.6$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.79	0.01	1	-0.19	-0.2	0.79	0.04	1	-0.18	-0.2
Ordered	δ	0.8	0	1	-0.18	-0.2	0.8	0.01	1	-0.17	-0.2

Table C.11: IRT model parameters: Generated Model M2PLMix Estimated Models the M2PLMix and 2PLMix, $N = 2000$, $J = 40$, $\lambda = 0.2$.

Generating Model M2PLMix $N = 2000$, $J = 40$, $\lambda = 0.2$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.06	0.09	0.1	0.14	0.04	0.04	0.08	0.1
	β	-0.07	0.09	0.09	0.12	0.04	0.12	0.1	0.16
	θ	-0.03	0.19	0.28	0.12	-0.03	0.18	0.28	0.11
Ordered	α	-0.09	0.08	0.11	0.16	0.05	0.03	0.08	0.05
	β	-0.02	0.04	0.06	0	0.05	0.07	0.07	0.11
	θ	-0.03	0.2	0.28	0.13	-0.03	0.18	0.28	0.12
Generating Model M2PLMix N = 2000, J = 40, $\lambda = 0.4$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.06	0.08	0.1	0.14	0.01	0.03	0.07	0.09
	β	-0.06	0.08	0.09	0.15	0	0.05	0.06	0.11
	θ	-0.02	0.18	0.27	0.12	-0.03	0.16	0.27	0.11
Ordered	α	-0.09	0.07	0.11	0.17	-0.01	0.02	0.07	0.05
	β	-0.02	0.03	0.06	0	0.01	0.04	0.06	0.08
	θ	-0.02	0.18	0.27	0.12	-0.03	0.16	0.27	0.12
Generating Model M2PLMix N = 2000, J = 40, $\lambda = 0.6$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.03	0.04	0.08	0.1	0	0.03	0.07	0.08
	β	-0.04	0.04	0.06	0.08	-0.02	0.02	0.05	0.09
	θ	-0.02	0.16	0.26	0.11	-0.03	0.15	0.26	0.1
Ordered	α	-0.05	0.05	0.09	0.15	-0.01	0.02	0.07	0.07
	β	-0.03	0.02	0.06	0	0	0.03	0.05	0.05
	θ	-0.03	0.17	0.26	0.12	-0.03	0.15	0.26	0.11

Table C.12: Location of Speededness: Generated Model M2PLMix Estimated Models the M2PLMix and 2PLMix, $N = 2000$, $J = 40$, $\lambda = 0.2$.

Generating Model M2PLMix $N = 2000$, $J = 40$, $\lambda = 0.2$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.6	0.25	1	-0.23	-0.31	0.58	0.56	0.97	-0.07	-0.16
Ordered	δ	0.61	0.11	1	-0.25	-0.35	0.57	0.36	0.94	-0.08	-0.22
Generating Model M2PLMix N = 2000, J = 40, $\lambda = 0.4$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.69	0.24	1	-0.16	-0.24	0.69	0.36	1	-0.1	-0.2
Ordered	δ	0.7	0.09	1	-0.23	-0.28	0.69	0.18	0.99	-0.17	-0.24
Generating Model M2PLMix N = 2000, J = 40, $\lambda = 0.6$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.8	0.26	1	-0.09	-0.15	0.8	0.29	1	-0.08	-0.14
Ordered	δ	0.8	0.07	1	-0.15	-0.19	0.8	0.11	1	-0.13	-0.18

Table C.13: Model Fit Data Generation Model M2PLMix Model, $N = 500$, Item Difficulty Random

Model Fit Data Generation under the M2PLMix Model, $N = 500$, Item Difficulty Random					
	Estimating Model M2PLMix and M2PLMix				
	-2LL	AIC	AICA	BIC	Adj. BIC
E=M2PLMix, $J = 40$, $\lambda = 0.2$	17009.14	19169.14	15150.28	23720.92	20292.93
E=2PLMix, $J = 40$, $\lambda = 0.2$	16394.68	18554.68	14535.82	23106.46	19678.47
E=M2PLMix, $J = 80$, $\lambda = 0.2$	32304.12	34624.12	30549.21	39513.07	35831.16
E=2PLMix, $J = 80$, $\lambda = 0.2$	32165.83	34485.83	30410.91	39374.77	35692.86
E=M2PLMix, $J = 40$, $\lambda = 0.4$	17234.69	19394.69	15375.83	23946.47	20518.48
E=2PLMix, $J = 40$, $\lambda = 0.4$	17177.22	19337.22	15318.36	23889	20461.01
E=M2PLMix, $J = 80$, $\lambda = 0.4$	33523.7	35843.7	31768.79	40732.65	37050.74
E=2PLMix, $J = 80$, $\lambda = 0.4$	33483.88	35803.88	31728.97	40692.83	37010.92
E=M2PLMix, $J = 40$, $\lambda = 0.6$	17686.9	19846.9	15828.03	24398.67	20970.69
E=2PLMix, $J = 40$, $\lambda = 0.6$	17681.24	19841.24	15822.38	24393.02	20965.03
E=M2PLMix, $J = 80$, $\lambda = 0.6$	34494.91	36814.91	32739.99	41703.85	38021.94
E=2PLMix, $J = 80$, $\lambda = 0.6$	34451.09	36771.09	32696.17	41660.04	37978.12

Table C.14: Model Fit Data Generation Model M2PLMix Model, $N = 500$, Item Difficulty Ordered

Model Fit Data Generation under the M2PLMix Model, $N = 500$, Item Difficulty Ordered					
	Estimating Model M2PLMix and M2PLMix				
	-2LL	AIC	AICA	BIC	Adj. BIC
E=M2PLMix, $J = 40$, $\lambda = 0.2$	16565.95	18725.95	14707.09	23277.73	19849.74
E=2PLMix, $J = 40$, $\lambda = 0.2$	16109.8	18269.8	14250.94	22821.58	19393.59
E=M2PLMix, $J = 80$, $\lambda = 0.2$	32548.54	34868.54	30793.62	39757.49	36075.57
E=2PLMix, $J = 80$, $\lambda = 0.2$	32407.56	34727.56	30652.65	39616.51	35934.6
E=M2PLMix, $J = 40$, $\lambda = 0.4$	17093.52	19253.52	15234.66	23805.3	20377.31
E=2PLMix, $J = 40$, $\lambda = 0.4$	16940.3	19100.3	15081.44	23652.08	20224.09
E=M2PLMix, $J = 80$, $\lambda = 0.4$	33617.74	35937.74	31862.83	40826.69	37144.78
E=2PLMix, $J = 80$, $\lambda = 0.4$	33517.52	35837.52	31762.6	40726.46	37044.55
E=M2PLMix, $J = 40$, $\lambda = 0.6$	17492.95	19652.95	15634.09	24204.73	20776.74
E=2PLMix, $J = 40$, $\lambda = 0.6$	17500.4	19660.4	15641.54	24212.18	20784.19
E=M2PLMix, $J = 80$, $\lambda = 0.6$	35393.97	37713.97	33639.05	42602.92	38921
E=2PLMix, $J = 80$, $\lambda = 0.6$	35300.06	37620.06	33545.14	42509.01	38827.09

Table C.15: Model Fit Data Generation Model M2PLMix Model, $N = 1000$, Item Difficulty Random

Model Fit Data Generation under the M2PLMix Model, $N = 1000$, Item Difficulty Random					
	Estimating Model M2PLMix and M2PLMix				
	-2LL	AIC	AICA	BIC	Adj. BIC
E=M2PLMix, $J = 20$, $\lambda = 0.2$	17623.46	21703.46	13704.15	31715.28	25236.12
E=2PLMix, $J = 20$, $\lambda = 0.2$	17312.8	21392.8	13393.49	31404.62	24925.47
E=M2PLMix, $J = 80$, $\lambda = 0.2$	64373.04	68693.04	60652.11	79293.79	72433.51
E=2PLMix, $J = 80$, $\lambda = 0.2$	64063.02	68383.02	60342.09	78983.77	72123.49
E=M2PLMix, $J = 20$, $\lambda = 0.4$	17551.91	21631.91	13632.6	31643.73	25164.58
E=2PLMix, $J = 20$, $\lambda = 0.4$	17827.54	21907.54	13908.23	31919.36	25440.21
E=M2PLMix, $J = 80$, $\lambda = 0.4$	67313.64	71633.64	63592.71	82234.39	75374.11
E=2PLMix, $J = 80$, $\lambda = 0.4$	67267	71587	63546.07	82187.75	75327.47
E=M2PLMix, $J = 20$, $\lambda = 0.6$	17736.95	21816.95	13817.64	31828.77	25349.62
E=2PLMix, $J = 20$, $\lambda = 0.6$	17952.77	22032.77	14033.46	32044.59	25565.43
E=M2PLMix, $J = 80$, $\lambda = 0.6$	68759.44	73079.44	65038.51	83680.19	76819.91
E=2PLMix, $J = 80$, $\lambda = 0.6$	68664.82	72984.82	64943.89	83585.57	76725.29

Table C.16: Model Fit Data Generation Model M2PLMix Model, $N = 1000$, Item Difficulty Ordered

Model Fit Data Generation under the M2PLMix Model, $N = 1000$, Item Difficulty Ordered					
	Estimating Model M2PLMix and M2PLMix				
	-2LL	AIC	AICA	BIC	Adj. BIC
E=M2PLMix, $J = 20$, $\lambda = 0.2$	16889.35	20969.35	12970.04	30981.17	24502.02
E=2PLMix, $J = 20$, $\lambda = 0.2$	16268.24	20348.24	12348.93	30360.06	23880.91
E=M2PLMix, $J = 80$, $\lambda = 0.2$	66267.82	70587.82	62546.89	81188.57	74328.29
E=2PLMix, $J = 80$, $\lambda = 0.2$	66015.38	70335.38	62294.45	80936.14	74075.85
E=M2PLMix, $J = 20$, $\lambda = 0.4$	17199.68	21279.68	13280.37	31291.5	24812.34
E=2PLMix, $J = 20$, $\lambda = 0.4$	17198.68	21278.68	13279.37	31290.5	24811.34
E=M2PLMix, $J = 80$, $\lambda = 0.4$	68601.22	72921.22	64880.29	83521.97	76661.69
E=2PLMix, $J = 80$, $\lambda = 0.4$	68433.69	72753.69	64712.76	83354.45	76494.17
E=M2PLMix, $J = 20$, $\lambda = 0.6$	17315	21395	13395.7	31406.83	24927.67
E=2PLMix, $J = 20$, $\lambda = 0.6$	17474.91	21554.91	13555.6	31566.73	25087.57
E=M2PLMix, $J = 80$, $\lambda = 0.6$	69781.2	74101.2	66060.27	84701.96	77841.68
E=2PLMix, $J = 80$, $\lambda = 0.6$	69622.7	73942.7	65901.77	84543.45	77683.17

Table C.17: Model Fit Data Generation Model M2PLMix Model, $N = 2000$, Item Difficulty Random

Model Fit Data Generation under the M2PLMix Model, $N = 2000$, Item Difficulty Random					
	Estimating Model M2PLMix and M2PLMix				
	-2LL	AIC	AICA	BIC	Adj. BIC
E=M2PLMix, $J = 20$, $\lambda = 0.2$	34969.05	43049.05	27051.36	65676.7	52841.4
E=2PLMix, $J = 20$, $\lambda = 0.2$	34480.08	42560.08	26562.4	65187.73	52352.43
E=M2PLMix, $J = 40$, $\lambda = 0.2$	67861.33	76021.33	60018.94	98873.01	85910.63
E=2PLMix, $J = 40$, $\lambda = 0.2$	65471.79	73631.79	57629.41	96483.47	83521.09
E=M2PLMix, $J = 20$, $\lambda = 0.4$	35316.39	43396.39	27398.7	66024.04	53188.74
E=2PLMix, $J = 20$, $\lambda = 0.4$	35799.39	43879.39	27881.7	66507.03	53671.73
E=M2PLMix, $J = 40$, $\lambda = 0.4$	69219.95	77379.95	61377.56	100231.63	87269.25
E=2PLMix, $J = 40$, $\lambda = 0.4$	68702.53	76862.53	60860.15	99714.22	86751.83
E=M2PLMix, $J = 20$, $\lambda = 0.6$	35310.27	43390.27	27392.59	66017.92	53182.62
E=2PLMix, $J = 20$, $\lambda = 0.6$	35729.52	43809.52	27811.84	66437.17	53601.87
E=M2PLMix, $J = 40$, $\lambda = 0.6$	70730.02	78890.02	62887.63	101741.7	88779.32
E=2PLMix, $J = 40$, $\lambda = 0.6$	70715.18	78875.18	62872.8	101726.86	88764.48

Table C.18: Model Fit Data Generation Model M2PLMix Model, $N = 2000$, Item Difficulty Ordered

Model Fit Data Generation under the M2PLMix Model, $N = 2000$, Item Difficulty Ordered					
	Estimating Model M2PLMix and M2PLMix				
	-2LL	AIC	AICA	BIC	Adj. BIC
E=M2PLMix, $J = 20$, $\lambda = 0.2$	33922.59	42002.59	26004.91	64630.24	51794.94
E=2PLMix, $J = 20$, $\lambda = 0.2$	32713.92	40793.92	24796.23	63421.57	50586.27
E=M2PLMix, $J = 40$, $\lambda = 0.2$	67337.42	75497.42	59495.04	98349.1	85386.72
E=2PLMix, $J = 40$, $\lambda = 0.2$	65522.15	73682.15	57679.76	96533.83	83571.45
E=M2PLMix, $J = 20$, $\lambda = 0.4$	34324.85	42404.85	26407.16	65032.49	52197.19
E=2PLMix, $J = 20$, $\lambda = 0.4$	34263.41	42343.41	26345.72	64971.06	52135.76
E=M2PLMix, $J = 40$, $\lambda = 0.4$	68737.08	76897.08	60894.7	99748.76	86786.38
E=2PLMix, $J = 40$, $\lambda = 0.4$	68208.66	76368.66	60366.28	99220.34	86257.96
E=M2PLMix, $J = 20$, $\lambda = 0.6$	34765.6	42845.6	26847.91	65473.25	52637.95
E=2PLMix, $J = 20$, $\lambda = 0.6$	35074.59	43154.59	27156.9	65782.23	52946.94
E=M2PLMix, $J = 40$, $\lambda = 0.6$	70404.72	78564.72	62562.34	101416.41	88454.02
E=2PLMix, $J = 40$, $\lambda = 0.6$	70399.55	78559.55	62557.17	101411.24	88448.85

Table C.19: IRT model parameters: Generated Model 2PLMix Estimated Models the M2PLMix and 2PLMix, $N = 500$, $J = 40$, $\pi = 0.2$.

Generating Model 2PLMix $N = 500$, $J = 40$, $\pi = 0.2$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.01	0.14	0.17	0.2	0.02	0.08	0.15	0.2
	β	-0.24	0.28	0.29	0.28	-0.08	0.1	0.15	0.23
	θ	-0.02	0.24	0.34	0.14	-0.03	0.21	0.31	0.14
Ordered	α	-0.11	0.14	0.21	0.2	-0.01	0.07	0.17	0.17
	β	-0.21	0.21	0.26	0.24	-0.08	0.11	0.16	0.25
	θ	-0.03	0.32	0.36	0.18	-0.03	0.25	0.33	0.15
Generating Model 2PLMix N = 500, J = 40, π == 0.4									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.03	0.06	0.13	0.2	0.05	0.04	0.14	0.2
	β	-0.18	0.18	0.21	0.23	-0.09	0.11	0.14	0.24
	θ	-0.03	0.24	0.32	0.15	-0.02	0.21	0.31	0.14
Ordered	α	-0.01	0.07	0.15	0.17	0.02	0.05	0.15	0.19
	β	-0.2	0.25	0.26	0.29	-0.12	0.18	0.19	0.27
	θ	-0.03	0.34	0.35	0.21	-0.02	0.29	0.32	0.19
Generating Model 2PLMix N = 500, J = 40, $\pi = 0.6$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.04	0.1	0.16	0.18	0	0.06	0.14	0.19
	β	-0.14	0.13	0.19	0.22	-0.08	0.09	0.14	0.22
	θ	-0.03	0.24	0.31	0.16	-0.03	0.21	0.29	0.14
Ordered	α	-0.04	0.1	0.16	0.17	-0.01	0.07	0.15	0.19
	β	-0.11	0.15	0.17	0.23	-0.07	0.11	0.14	0.22
	θ	-0.03	0.25	0.31	0.15	-0.03	0.21	0.3	0.13

Table C.20: Location of Speededness: Generated Model 2PLMix Estimated Models the M2PLMix and 2PLMix, $N = 500$, $J = 40$, $\pi = 0.2$.

Generating Model 2PLMix $N = 500$, $J = 40$, $\pi = 0.2$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.19	0.2	1	-0.43	-0.66	0.21	0.48	0.99	-0.16	-0.42
Ordered	δ	0.19	0.11	1	-0.59	-0.72	0.19	0.38	0.97	-0.21	-0.49
Generating Model 2PLMix N = 500, J = 40, $\pi = 0.4$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.4	0.25	1	-0.33	-0.46	0.4	0.37	1	-0.25	-0.38
Ordered	δ	0.37	0.11	1	-0.43	-0.56	0.37	0.23	0.99	-0.29	-0.48
Generating Model 2PLMix N = 500, J = 40, $\pi = 0.6$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.57	0.25	1	-0.25	-0.33	0.57	0.33	1	-0.2	-0.29
Ordered	δ	0.56	0.1	1	-0.28	-0.4	0.56	0.16	1	-0.22	-0.37

Table C.21: IRT model parameters: Generated Model 2PLMix Estimated Models the M2PLMix and 2PLMix, $N = 500$, $J = 80$, $\pi = 0.2$.

Generating Model 2PLMix $N = 500$, $J = 80$, $\pi = 0.2$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.08	0.06	0.16	0.18	0.07	0.05	0.15	0.17
	β	-0.11	0.18	0.19	0.22	-0.06	0.12	0.15	0.22
	θ	-0.02	0.12	0.23	0.12	-0.02	0.11	0.22	0.13
Ordered	α	0.01	0.1	0.16	0.19	0.06	0.05	0.16	0.18
	β	-0.16	0.34	0.25	0.33	-0.09	0.22	0.19	0.29
	θ	-0.02	0.2	0.25	0.16	-0.02	0.17	0.24	0.15
Generating Model 2PLMix N = 500, J = 80, $\pi = 0.4$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.06	0.05	0.15	0.19	0.06	0.05	0.15	0.2
	β	-0.09	0.14	0.17	0.23	-0.06	0.13	0.15	0.23
	θ	-0.03	0.12	0.22	0.12	-0.02	0.12	0.22	0.13
Ordered	α	0.04	0.07	0.15	0.2	0.08	0.05	0.16	0.18
	β	-0.1	0.21	0.19	0.26	-0.06	0.16	0.17	0.26
	θ	-0.02	0.21	0.24	0.19	-0.02	0.16	0.23	0.15
Generating Model 2PLMix N = 500, J = 80, $\pi = 0.6$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.06	0.05	0.14	0.18	0.07	0.05	0.14	0.18
	β	-0.06	0.11	0.14	0.22	-0.04	0.11	0.13	0.23
	θ	-0.03	0.09	0.2	0.13	-0.03	0.08	0.2	0.13
Ordered	α	0.06	0.06	0.15	0.19	0.08	0.05	0.15	0.18
	β	-0.09	0.19	0.18	0.26	-0.06	0.15	0.16	0.26
	θ	-0.02	0.19	0.23	0.18	-0.02	0.13	0.21	0.13

Table C.22: Location of Speededness: Generated Model 2PLMix Estimated Models the M2PLMix and 2PLMix, $N = 500$, $J = 80$, $\pi = 0.2$.

Generating Model 2PLMix $N = 500$, $J = 80$, $\pi = 0.2$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.21	0.58	1	-0.04	-0.34	0.22	0.66	1	-0.02	-0.27
Ordered	δ	0.19	0.26	1	-0.28	-0.6	0.19	0.37	0.99	-0.15	-0.51
Generating Model 2PLMix N = 500, J = 80, $\pi = 0.4$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.37	0.64	1	-0.04	-0.23	0.38	0.68	1	-0.02	-0.2
Ordered	δ	0.42	0.32	1	-0.19	-0.39	0.42	0.4	1	-0.12	-0.35
Generating Model 2PLMix N = 500, J = 80, $\pi = 0.6$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.59	0.63	1	-0.03	-0.15	0.6	0.66	1	-0.01	-0.14
Ordered	δ	0.59	0.26	1	-0.19	-0.3	0.59	0.34	1	-0.13	-0.27

Table C.23: IRT model parameters: Generated Model 2PLMix Estimated Models the M2PLMix and 2PLMix, $N = 1000$, $J = 20$, $\pi = 0.2$.

Generating Model 2PLMix $N = 1000$, $J = 20$, $\pi = 0.2$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.15	0.18	0.22	0.17	-0.05	0.05	0.11	0.18
	β	-0.38	0.3	0.4	0.28	-0.14	0.09	0.16	0.19
	θ	-0.03	0.53	0.49	0.26	-0.03	0.44	0.44	0.22
Ordered	α	-0.15	0.16	0.22	0.18	-0.07	0.08	0.15	0.19
	β	-0.31	0.22	0.35	0.19	-0.15	0.1	0.18	0.22
	θ	-0.03	0.55	0.49	0.27	-0.03	0.49	0.46	0.24
Generating Model 2PLMix N = 1000, J = 20, $\pi = 0.4$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.08	0.12	0.16	0.18	-0.02	0.05	0.11	0.16
	β	-0.24	0.21	0.27	0.22	-0.09	0.08	0.13	0.21
	θ	-0.03	0.44	0.43	0.21	-0.03	0.37	0.4	0.18
Ordered	α	-0.11	0.12	0.16	0.21	-0.08	0.07	0.13	0.2
	β	-0.23	0.15	0.25	0.17	-0.13	0.08	0.16	0.2
	θ	-0.03	0.5	0.45	0.26	-0.03	0.46	0.43	0.25
Generating Model 2PLMix N = 1000, J = 20, $\pi = 0.6$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.11	0.13	0.17	0.19	-0.08	0.08	0.14	0.18
	β	-0.19	0.15	0.21	0.19	-0.12	0.08	0.15	0.19
	θ	-0.02	0.43	0.41	0.22	-0.03	0.39	0.4	0.2
Ordered	α	-0.09	0.11	0.15	0.2	-0.07	0.07	0.13	0.2
	β	-0.14	0.11	0.18	0.21	-0.09	0.06	0.13	0.19
	θ	-0.02	0.44	0.41	0.25	-0.02	0.42	0.4	0.24

Table C.24: Location of Speededness: Generated Model 2PLMix Estimated Models the M2PLMix and 2PLMix, $N = 1000$, $J = 20$, $\pi = 0.2$.

Generating Model 2PLMix $N = 1000$, $J = 20$, $\pi = 0.2$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.2	0.01	1	-0.78	-0.8	0.2	0.34	0.91	-0.4	-0.52
Ordered	δ	0.17	0	1	-0.73	-0.83	0.15	0.46	0.79	-0.09	-0.42
Generating Model 2PLMix N = 1000, J = 20, $\pi = 0.4$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.42	0.01	1	-0.42	-0.58	0.42	0.1	0.99	-0.34	-0.52
Ordered	δ	0.41	0	1	-0.54	-0.59	0.41	0.1	0.97	-0.39	-0.51
Generating Model 2PLMix N = 1000, J = 20, $\pi = 0.6$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.6	0.01	1	-0.38	-0.4	0.6	0.04	1	-0.36	-0.39
Ordered	δ	0.58	0	1	-0.37	-0.42	0.58	0.01	1	-0.36	-0.42

Table C.25: IRT model parameters: Generated Model 2PLMix Estimated Models the M2PLMix and 2PLMix, $N = 1000$, $J = 80$, $\pi = 0.2$.

Generating Model 2PLMix $N = 1000$, $J = 80$, $\pi = 0.2$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.05	0.05	0.11	0.18	0.04	0.04	0.11	0.17
	β	-0.11	0.17	0.16	0.22	-0.05	0.09	0.11	0.2
	θ	-0.03	0.12	0.22	0.12	-0.02	0.11	0.22	0.12
Ordered	α	-0.07	0.13	0.15	0.21	0	0.06	0.11	0.2
	β	-0.1	0.14	0.14	0.21	-0.04	0.07	0.1	0.19
	θ	-0.02	0.15	0.24	0.14	-0.02	0.14	0.23	0.14
Generating Model 2PLMix N = 1000, J = 80, $\pi = 0.4$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.04	0.04	0.1	0.16	0.04	0.03	0.1	0.16
	β	-0.08	0.09	0.12	0.18	-0.05	0.06	0.09	0.17
	θ	-0.03	0.12	0.21	0.13	-0.02	0.11	0.21	0.13
Ordered	α	-0.02	0.07	0.11	0.18	0.01	0.04	0.1	0.16
	β	-0.09	0.13	0.13	0.2	-0.04	0.08	0.1	0.18
	θ	-0.03	0.14	0.23	0.14	-0.02	0.12	0.21	0.14
Generating Model 2PLMix N = 1000, J = 80, $\pi = 0.6$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.03	0.04	0.1	0.15	0.04	0.04	0.1	0.15
	β	-0.06	0.07	0.09	0.17	-0.04	0.06	0.09	0.16
	θ	-0.03	0.09	0.2	0.13	-0.03	0.09	0.2	0.12
Ordered	α	-0.03	0.07	0.11	0.19	0.01	0.04	0.1	0.17
	β	-0.07	0.09	0.11	0.19	-0.05	0.07	0.09	0.18
	θ	-0.02	0.19	0.23	0.19	-0.03	0.14	0.21	0.15

Table C.26: Location of Speededness: Generated Model 2PLMix Estimated Models the M2PLMix and 2PLMix, $N = 1000$, $J = 80$, $\pi = 0.2$.

Generating Model 2PLMix $N = 1000$, $J = 80$, $\pi = 0.2$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.24	0.57	1	-0.05	-0.33	0.25	0.67	1	-0.02	-0.26
Ordered	δ	0.21	0.28	1	-0.26	-0.57	0.21	0.38	1	-0.17	-0.49
Generating Model 2PLMix N = 1000, J = 80, $\pi = 0.4$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.45	0.63	1	-0.05	-0.21	0.45	0.68	1	-0.03	-0.18
Ordered	δ	0.41	0.29	1	-0.19	-0.42	0.41	0.37	1	-0.12	-0.37
Generating Model 2PLMix N = 1000, J = 80, $\pi = 0.6$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.61	0.62	1	-0.02	-0.15	0.61	0.65	1	-0.01	-0.14
Ordered	δ	0.57	0.26	1	-0.23	-0.32	0.57	0.36	1	-0.15	-0.27

Table C.27: IRT model parameters: Generated Model 2PLMix Estimated Models the M2PLMix and 2PLMix, $N = 2000$, $J = 20$, $\pi = 0.2$.

Generating Model 2PLMix $N = 2000$, $J = 20$, $\pi = 0.2$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.11	0.14	0.17	0.18	-0.03	0.05	0.08	0.13
	β	-0.36	0.29	0.37	0.29	-0.13	0.08	0.14	0.17
	θ	-0.03	0.51	0.47	0.24	-0.03	0.42	0.43	0.2
Ordered	α	-0.15	0.16	0.2	0.16	-0.08	0.06	0.11	0.18
	β	-0.29	0.19	0.32	0.18	-0.12	0.06	0.14	0.2
	θ	-0.03	0.54	0.48	0.28	-0.02	0.49	0.46	0.25
Generating Model 2PLMix N = 2000, J = 20, $\pi = 0.4$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.11	0.12	0.15	0.19	-0.07	0.05	0.1	0.14
	β	-0.25	0.21	0.26	0.23	-0.12	0.09	0.14	0.15
	θ	-0.03	0.45	0.43	0.22	-0.03	0.4	0.41	0.19
Ordered	α	-0.09	0.12	0.15	0.2	-0.06	0.06	0.1	0.18
	β	-0.22	0.16	0.25	0.2	-0.12	0.08	0.14	0.18
	θ	-0.03	0.5	0.46	0.26	-0.02	0.46	0.43	0.24
Generating Model 2PLMix N = 2000, J = 20, $\pi = 0.6$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.12	0.14	0.16	0.19	-0.09	0.08	0.12	0.18
	β	-0.22	0.16	0.22	0.2	-0.13	0.09	0.15	0.15
	θ	-0.03	0.46	0.43	0.24	-0.02	0.41	0.4	0.22
Ordered	α	-0.1	0.1	0.14	0.19	-0.08	0.07	0.11	0.18
	β	-0.15	0.08	0.16	0.17	-0.09	0.05	0.11	0.1
	θ	-0.03	0.45	0.41	0.25	-0.03	0.43	0.4	0.23

Table C.28: Location of Speededness: Generated Model 2PLMix Estimated Models the M2PLMix and 2PLMix, $N = 2000$, $J = 20$, $\pi = 0.2$.

Generating Model 2PLMix $N = 2000$, $J = 20$, $\pi = 0.2$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.2	0.01	1	-0.68	-0.8	0.2	0.34	0.92	-0.32	-0.52
Ordered	δ	0.19	0	1	-0.75	-0.8	0.17	0.47	0.79	-0.09	-0.39
Generating Model 2PLMix N = 2000, J = 20, $\pi = 0.4$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.41	0.01	1	-0.52	-0.59	0.41	0.11	0.99	-0.44	-0.52
Ordered	δ	0.42	0	1	-0.52	-0.58	0.41	0.09	0.97	-0.39	-0.52
Generating Model 2PLMix N = 2000, J = 20, $\pi = 0.6$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.6	0.01	1	-0.4	-0.4	0.6	0.05	1	-0.38	-0.38
Ordered	δ	0.6	0	1	-0.4	-0.4	0.6	0.01	1	-0.39	-0.39

Table C.29: IRT model parameters: Generated Model 2PLMix Estimated Models the M2PLMix and 2PLMix, $N = 2000$, $J = 40$, $\pi = 0.2$.

Generating Model 2PLMix $N = 2000$, $J = 40$, $\pi = 0.2$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	0.02	0.05	0.08	0.11	0.03	0.03	0.08	0.11
	β	-0.19	0.22	0.21	0.25	-0.08	0.07	0.1	0.15
	θ	-0.02	0.26	0.33	0.16	-0.03	0.24	0.31	0.16
Ordered	α	-0.13	0.11	0.16	0.18	-0.04	0.04	0.09	0.13
	β	-0.19	0.15	0.2	0.21	-0.07	0.06	0.09	0.18
	θ	-0.03	0.34	0.35	0.21	-0.02	0.29	0.33	0.19
Generating Model 2PLMix N = 2000, J = 40, $\pi = 0.4$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.02	0.06	0.09	0.11	0.01	0.03	0.07	0.05
	β	-0.15	0.14	0.16	0.19	-0.07	0.04	0.08	0.1
	θ	-0.03	0.22	0.31	0.15	-0.03	0.2	0.29	0.13
Ordered	α	-0.11	0.1	0.14	0.2	-0.05	0.05	0.09	0.15
	β	-0.14	0.1	0.15	0.18	-0.07	0.04	0.09	0.16
	θ	-0.02	0.32	0.33	0.2	-0.02	0.27	0.31	0.17
Generating Model 2PLMix N = 2000, J = 40, $\pi = 0.6$									
		Estimating Model M2PLMix				Estimating Model 2PLMix			
		Bias	Bias SD	RMSE	RMSE SD	Bias	Bias SD	RMSE	RMSE SD
Random	α	-0.02	0.03	0.07	0.11	0	0.02	0.07	0.08
	β	-0.09	0.08	0.1	0.13	-0.06	0.03	0.07	0.1
	θ	-0.03	0.19	0.29	0.13	-0.03	0.18	0.28	0.12
Ordered	α	-0.06	0.06	0.1	0.15	-0.03	0.03	0.07	0.11
	β	-0.09	0.06	0.1	0.15	-0.05	0.04	0.07	0.09
	θ	-0.03	0.25	0.29	0.15	-0.03	0.22	0.28	0.14

Table C.30: Location of Speededness: Generated Model 2PLMix Estimated Models the M2PLMix and 2PLMix, $N = 2000$, $J = 40$, $\pi = 0.2$.

Generating Model 2PLMix $N = 2000$, $J = 40$, $\pi = 0.2$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.22	0.26	1	-0.34	-0.59	0.23	0.49	0.99	-0.16	-0.4
Ordered	δ	0.21	0.11	1	-0.5	-0.71	0.2	0.36	0.97	-0.18	-0.5
Generating Model 2PLMix N = 2000, J = 40, $\pi = 0.4$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.4	0.26	1	-0.3	-0.45	0.41	0.38	1	-0.2	-0.38
Ordered	δ	0.4	0.1	1	-0.4	-0.54	0.4	0.22	1	-0.26	-0.47
Generating Model 2PLMix N = 2000, J = 40, $\pi = 0.6$											
		Estimating Model M2PLMix					Estimating Model 2PLMix				
		M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
Random	δ	0.61	0.29	1	-0.18	-0.28	0.61	0.33	1	-0.15	-0.26
Ordered	δ	0.62	0.09	1	-0.23	-0.35	0.62	0.14	1	-0.19	-0.33

Table C.31: Model Fit Data Generation Model 2PLMix Model, $N = 500$, Item Difficulty Random

Model Fit Data Generation under the 2PLMix Model, $N = 500$, Item Difficulty Random					
	Estimating Model M2PLMix and M2PLMix				
	-2LL	AIC	AICA	BIC	Adj. BIC
E=M2PLMix, $J = 40$, $\pi = 0.2$	16012.11	18172.11	14153.25	22723.89	19295.9
E=2PLMix, $J = 40$, $\pi = 0.2$	14895.72	17055.72	13036.85	21607.49	18179.51
E=M2PLMix, $J = 80$, $\pi = 0.2$	28277.31	30597.31	26522.4	35486.26	31804.35
E=2PLMix, $J = 80$, $\pi = 0.2$	27936.37	30256.37	26181.46	35145.32	31463.41
E=M2PLMix, $J = 40$, $\pi = 0.4$	16755.06	18915.06	14896.2	23466.84	20038.85
E=2PLMix, $J = 40$, $\pi = 0.4$	16446.95	18606.95	14588.09	23158.73	19730.74
E=M2PLMix, $J = 80$, $\pi = 0.4$	29600.17	31920.17	27845.25	36809.11	33127.2
E=2PLMix, $J = 80$, $\pi = 0.4$	29416.08	31736.08	27661.17	36625.03	32943.12
E=M2PLMix, $J = 40$, $\pi = 0.6$	16818.1	18978.1	14959.23	23529.87	20101.89
E=2PLMix, $J = 40$, $\pi = 0.6$	16690	18850	14831.13	23401.77	19973.79
E=M2PLMix, $J = 80$, $\pi = 0.6$	32204.04	34524.04	30449.12	39412.98	35731.07
E=2PLMix, $J = 80$, $\pi = 0.6$	32108.87	34428.87	30353.96	39317.82	35635.91

Table C.32: Model Fit Data Generation Model 2PLMix Model, $N = 500$, Item Difficulty Ordered

Model Fit Data Generation under the 2PLMix Model, $N = 500$, Item Difficulty Ordered					
	Estimating Model M2PLMix and M2PLMix				
	-2LL	AIC	AICA	BIC	Adj. BIC
E=M2PLMix, $J = 40$, $\pi = 0.2$	15080.35	17240.35	13221.49	21792.13	18364.14
E=2PLMix, $J = 40$, $\pi = 0.2$	14006.26	16166.26	12147.4	20718.04	17290.05
E=M2PLMix, $J = 80$, $\pi = 0.2$	29085.01	31405.01	27330.09	36293.95	32612.04
E=2PLMix, $J = 80$, $\pi = 0.2$	28586.3	30906.3	26831.39	35795.25	32113.34
E=M2PLMix, $J = 40$, $\pi = 0.4$	15908.55	18068.55	14049.68	22620.32	19192.34
E=2PLMix, $J = 40$, $\pi = 0.4$	15460.02	17620.02	13601.16	22171.8	18743.81
E=M2PLMix, $J = 80$, $\pi = 0.4$	30785.02	33105.02	29030.1	37993.97	34312.05
E=2PLMix, $J = 80$, $\pi = 0.4$	30382.79	32702.79	28627.88	37591.74	33909.83
E=M2PLMix, $J = 40$, $\pi = 0.6$	16555.73	18715.73	14696.86	23267.5	19839.52
E=2PLMix, $J = 40$, $\pi = 0.6$	16436.85	18596.85	14577.98	23148.63	19720.64
E=M2PLMix, $J = 80$, $\pi = 0.6$	33240.42	35560.42	31485.5	40449.37	36767.45
E=2PLMix, $J = 80$, $\pi = 0.6$	32840.87	35160.87	31085.95	40049.82	36367.9

Table C.33: Model Fit Data Generation Model 2PLMix Model, $N = 1000$, Item Difficulty Random

Model Fit Data Generation under the 2PLMix Model, $N = 1000$, Item Difficulty Random					
	Estimating Model M2PLMix and M2PLMix				
	-2LL	AIC	AICA	BIC	Adj. BIC
E=M2PLMix, $J = 20$, $\pi = 0.2$	17183.95	21263.95	13264.64	31275.77	24796.61
E=2PLMix, $J = 20$, $\pi = 0.2$	16686.66	20766.66	12767.35	30778.48	24299.33
E=M2PLMix, $J = 80$, $\pi = 0.2$	57313.8	61633.8	53592.87	72234.56	65374.27
E=2PLMix, $J = 80$, $\pi = 0.2$	56583.54	60903.54	52862.61	71504.29	64644.01
E=M2PLMix, $J = 20$, $\pi = 0.4$	17832.23	21912.23	13912.92	31924.05	25444.89
E=2PLMix, $J = 20$, $\pi = 0.4$	17866.2	21946.2	13946.89	31958.02	25478.87
E=M2PLMix, $J = 80$, $\pi = 0.4$	60665.4	64985.4	56944.47	75586.16	68725.87
E=2PLMix, $J = 80$, $\pi = 0.4$	60333.06	64653.06	56612.13	75253.81	68393.53
E=M2PLMix, $J = 20$, $\pi = 0.6$	17775.27	21855.27	13855.96	31867.09	25387.94
E=2PLMix, $J = 20$, $\pi = 0.6$	17914.7	21994.7	13995.39	32006.52	25527.36
E=M2PLMix, $J = 80$, $\pi = 0.6$	64863.29	69183.29	61142.36	79784.04	72923.76
E=2PLMix, $J = 80$, $\pi = 0.6$	64738.46	69058.46	61017.53	79659.21	72798.93

Table C.34: Model Fit Data Generation Model 2PLMix Model, $N = 1000$, Item Difficulty Ordered

Model Fit Data Generation under the 2PLMix Model, $N = 1000$, Item Difficulty Ordered					
	Estimating Model M2PLMix and M2PLMix				
	-2LL	AIC	AICA	BIC	Adj. BIC
E=M2PLMix, $J = 20$, $\pi = 0.2$	15545.17	19625.17	11625.86	29636.99	23157.84
E=2PLMix, $J = 20$, $\pi = 0.2$	14648.75	18728.75	10729.45	28740.57	22261.42
E=M2PLMix, $J = 80$, $\pi = 0.2$	58069.91	62389.91	54348.98	72990.66	66130.38
E=2PLMix, $J = 80$, $\pi = 0.2$	57217.9	61537.9	53496.97	72138.65	65278.37
E=M2PLMix, $J = 20$, $\pi = 0.4$	16288.78	20368.78	12369.47	30380.6	23901.45
E=2PLMix, $J = 20$, $\pi = 0.4$	16180.43	20260.43	12261.12	30272.25	23793.1
E=M2PLMix, $J = 80$, $\pi = 0.4$	61989.43	66309.43	58268.5	76910.18	70049.9
E=2PLMix, $J = 80$, $\pi = 0.4$	61067.89	65387.89	57346.96	75988.64	69128.36
E=M2PLMix, $J = 20$, $\pi = 0.6$	16796.04	20876.04	12876.73	30887.86	24408.71
E=2PLMix, $J = 20$, $\pi = 0.6$	16971.92	21051.92	13052.61	31063.74	24584.59
E=M2PLMix, $J = 80$, $\pi = 0.6$	65605.16	69925.16	61884.23	80525.91	73665.63
E=2PLMix, $J = 80$, $\pi = 0.6$	64582.71	68902.71	60861.77	79503.46	72643.18

Table C.35: Model Fit Data Generation Model 2PLMix Model, $N = 2000$, Item Difficulty Random

Model Fit Data Generation under the 2PLMix Model, $N = 2000$, Item Difficulty Random					
	Estimating Model M2PLMix and M2PLMix				
	-2LL	AIC	AICA	BIC	Adj. BIC
E=M2PLMix, $J = 20$, $\pi = 0.2$	35000.62	43080.62	27082.93	65708.27	52872.97
E=2PLMix, $J = 20$, $\pi = 0.2$	33776.48	41856.48	25858.79	64484.12	51648.82
E=M2PLMix, $J = 40$, $\pi = 0.2$	63877.63	72037.63	56035.24	94889.31	81926.93
E=2PLMix, $J = 40$, $\pi = 0.2$	60272.2	68432.2	52429.81	91283.88	78321.5
E=M2PLMix, $J = 20$, $\pi = 0.4$	35517.49	43597.49	27599.8	66225.13	53389.83
E=2PLMix, $J = 20$, $\pi = 0.4$	35400.12	43480.12	27482.44	66107.77	53272.47
E=M2PLMix, $J = 40$, $\pi = 0.4$	65992.85	74152.85	58150.46	97004.53	84042.15
E=2PLMix, $J = 40$, $\pi = 0.4$	64738.06	72898.06	56895.68	95749.75	82787.36
E=M2PLMix, $J = 20$, $\pi = 0.6$	35606.45	43686.45	27688.76	66314.09	53478.79
E=2PLMix, $J = 20$, $\pi = 0.6$	35849.34	43929.34	27931.66	66556.99	53721.69
E=M2PLMix, $J = 40$, $\pi = 0.6$	68903.92	77063.92	61061.54	99915.61	86953.22
E=2PLMix, $J = 40$, $\pi = 0.6$	68591.59	76751.59	60749.21	99603.27	86640.89

Table C.36: Model Fit Data Generation Model 2PLMix Model, N = 2000, Item Difficulty Ordered

Model Fit Data Generation under the 2PLMix Model, $N = 2000$, Item Difficulty Ordered					
	Estimating Model M2PLMix and M2PLMix				
	-2LL	AIC	AICA	BIC	Adj. BIC
E=M2PLMix, $J = 20$, $\pi = 0.2$	31293.03	39373.03	23375.34	62000.68	49165.38
E=2PLMix, $J = 20$, $\pi = 0.2$	29485.96	37565.96	21568.27	60193.6	47358.3
E=M2PLMix, $J = 40$, $\pi = 0.2$	60885.33	69045.33	53042.95	91897.02	78934.63
E=2PLMix, $J = 40$, $\pi = 0.2$	57070.36	65230.36	49227.98	88082.05	75119.66
E=M2PLMix, $J = 20$, $\pi = 0.4$	32712.03	40792.03	24794.34	63419.68	50584.38
E=2PLMix, $J = 20$, $\pi = 0.4$	32566.25	40646.25	24648.56	63273.9	50438.6
E=M2PLMix, $J = 40$, $\pi = 0.4$	64201.79	72361.79	56359.41	95213.48	82251.09
E=2PLMix, $J = 40$, $\pi = 0.4$	62486.55	70646.55	54644.16	93498.23	80535.85
E=M2PLMix, $J = 20$, $\pi = 0.6$	33728.12	41808.12	25810.44	64435.77	51600.47
E=2PLMix, $J = 20$, $\pi = 0.6$	34043.9	42123.9	26126.22	64751.55	51916.25
E=M2PLMix, $J = 40$, $\pi = 0.6$	68032.16	76192.16	60189.77	99043.84	86081.46
E=2PLMix, $J = 40$, $\pi = 0.6$	67711.67	75871.67	59869.29	98723.35	85760.97

Appendix D

Code

Estimation of 2PL Model

```
#include <boost/math/distributions/normal.hpp>
using boost::math::normal;
#include <iostream>
#include <iomanip>
#include <limits>
#include <armadillo>
#include <stdio.h>
#include <stdlib.h>
#include <cstdio>
#include <random>
#include <vector>
#include <numeric>
#include <algorithm>
#include <iterator>
#include <cmath>
#include <stdio.h>
#include <math.h>
#include <time.h>
#include <iomanip>
#include <time.h>


// Part 1: Namespace define
using namespace std;
using namespace arma;

// Part 2: To generate random numbers each new time
std::default_random_engine g_generator(time(0));

/////////////////////////////////////////////////////
// 1. Structure Template for Each chain
/////////////////////////////////////////////////////
// c) 2PL
struct cur2PL_2chain
{
  int N;
  int J;
  int BI;
  int IT;
  int thin;
  double s2t; double MHt;
  double s2b; double MHb;
  double s2a; double MHa;
  arma::vec theta;
  arma::vec alpha;
  arma::vec beta;
  arma::mat X;
};

/////////////////////////////////////////////////////
```

```
// 2. Structure Template for Estimated Parameters
/////////////////////////////////////////////////////
// c) 2PL
struct contain2_2PL
{
  arma::mat thetaestcon_chainA_2PL;
  arma::mat thetaestcon_chainB_2PL;
  arma::mat alphaestcon_chainA_2PL;
  arma::mat alphaestcon_chainB_2PL;
  arma::mat betaestcon_chainA_2PL;
  arma::mat betaestcon_chainB_2PL;
};

// ----------------------------------------
//
// Functions
//
// ----------------------------------------
/////////////////////////////////////////////////////
// C1. General Functions
/////////////////////////////////////////////////////
arma::mat getdata(FILE * pFile, int N, int J){
  arma::mat X(N,J);
  double data[N][J];

  for(int i = 0; i < N; i++){
    for(int j = 0; j < J; j++){
      if (fscanf(pFile,"%lf", &data[i][j])==1){
        X(i,j) = data[i][j];

      } else{
        cout << "Failure" << endl;
      }
    }
  }
  return X;

}
void writecsv(arma::mat M, const char *dir, const char *fname){
  // Make sure the directory exists
  char cmd[4096];
  sprintf(cmd, "mkdir -p %s", dir);
  system(cmd);      // execute the mkdir command

  // Concatenate the directory and filename
  char path[4096];
  char ttpath[4096];

  sprintf(ttpath, "%s%s%s%s", dir,"_", fname,".txt");
  sprintf(path, "%s/%s", dir, ttpath);

  // Open the file path
  FILE * pFile = fopen (path,"w");
  if (pFile == NULL) {
    printf("ERROR: writecsv could not open %s for writing\n", fname);
    exit(1);
  }

  // Write out the data
  for(int i = 0; i < M.n_rows; i++){
    for(int j = 0; j < M.n_cols; j++){
      fprintf(pFile," %0.6f",M(i,j));
    }
    fprintf(pFile,"\n");
  }

  // Close the file
  fclose (pFile);
}
arma::vec colSumsRcpp(arma::mat x) {
  int J = x.n_cols;
  arma::vec out(J);
  for (int j = 0; j < J; j++) {
    double total = 0;
    total = sum(x.col(j));
    out[j] = total;
  }
  return out;
}
arma::vec rowSumsRcpp(arma::mat x) {
  int N = x.n_rows;
  arma::vec out(N);
  for (int i = 0; i < N; i++) {
    double total = 0;
```

```
    total = sum(x.row(i));
    out[i] = total;
  }
  return out;
}
arma::vec rangecpp(const int begin, const int end){
  arma::vec ptr(end-1);
  for (int i=begin; i < end; i++){
    ptr[i-1]=i;
  }
  return ptr;
}
/////////////////////////////////////////////////////
// C2. Random Distributions Functions
/////////////////////////////////////////////////////
arma::vec  dnormcpp(arma::vec nrandomVar, double mu,double sig){
  arma::vec  A(nrandomVar.size());
  for (int i = 0; i < nrandomVar.size(); i++) {
    double y = (1/sqrt(2*M_PI* pow(sig,2.0) ))*
    exp((-0.5*pow((nrandomVar[i] - mu),2))/pow(sig,2));
    A[i] = y;
  }
  return A;

}
arma::vec rnormcpp(int J){

  //  default_random_engine generator;
  normal_distribution<double> distribution(0,1);
  arma::vec B(J);

  for (int j=0; j<J;j++) {
    double number = distribution(g_generator);
    B[j]=number;
  }
  return B;
}
arma::vec  dlnormcpp(arma::vec nrandomVar, double mu, double sig){
  arma::vec  A(nrandomVar.size());
  for (int i = 0; i < nrandomVar.size(); i++) {
    double y = (1/(nrandomVar[i]*sig*sqrt(2*M_PI)))*
    exp((-0.5*pow(( log(nrandomVar[i]) - mu),2))/pow(sig,2));
    A[i] = y;
  }
  return A;

}
arma::vec runifcpp(int J, double a, double b){

  //  default_random_engine generator;
  uniform_real_distribution<double> distribution(a,b);
  arma::vec B(J);

  for (int j=0; j<J;j++) {
    double number = distribution(g_generator);
    B[j]=number;
  }
  return B;
}
/////////////////////////////////////////////////////
// C3. Initialize Parameters Functions
/////////////////////////////////////////////////////
arma::vec intT(arma::mat X){
  normal s;
  arma::vec caltheta = 0.025 + (rowSumsRcpp( X)/X.n_cols)*0.95;
  arma::vec inttheta(X.n_rows);
  for (int j = 0; j < X.n_rows; j++) {
    double f =  quantile(s,caltheta[j]);
    inttheta[j]=f;
  }
  return inttheta;
}
arma::vec intB(arma::mat X){
  normal s;
  arma::vec calbeta = 0.025 + (colSumsRcpp( X)/X.n_rows)*0.95;
  arma::mat intbeta(X.n_cols,1);
  for (int i = 0; i < X.n_cols; i++) {
    arma::vec f =  quantile(s,calbeta[i]) + runifcpp(1,-0.01,0.01);
    intbeta.row(i)=-f;
  }
  return intbeta;
}
/////////////////////////////////////////////////////
// C4. Speedness Functions
```

```
////////////////////////////////////////////////////
arma::vec probsp_IW(arma::vec tau, arma::vec theta){
  int N=theta.size();
  arma::vec nsp(N);
  for (int i= 0; i < N; i++){
    arma::vec tmp1 = tau + (1-tau) / (1+exp(-theta[i]));
    nsp[i]=tmp1[0];
  }

  return nsp;
}
arma::mat genpmatm2pl(int J, arma::vec taunsp, int omega){
  int N = taunsp.size();
  arma::vec items = rangecpp(1,J);
  arma::mat spvec1(J-1,N);
  for (int i= 0; i < N; i++){
    arma::vec tmp1 = ((pow(items,omega) -  pow(items-1,omega))
    * (1-taunsp(i))) / pow(J-1,omega);
    spvec1.col(i) = tmp1;
  }
  arma::mat spvec2 = join_cols(spvec1, arma::trans(taunsp));
  return spvec2;

}
arma::vec twoplmixsp(int J, arma::vec tau, int omega){

  arma::vec items = rangecpp(1,J);
  arma::vec tmp1 = ((pow(items,omega) -  pow(items-1,omega))
  * (1-tau)) / pow(J-1,omega);
  arma::vec spvec= join_cols(tmp1,tau);
  return spvec;
}
arma::mat genetacpp(arma::vec delta, arma::vec items, double kappa){

  int J = items.max();
  int N = delta.size();
  arma::vec div_item = 1/items;

  arma::mat tmp1(N,J);
  for(int i=0;i<N;i++){
    tmp1.row(i) = arma::trans(div_item);
  }


  arma::mat tmp2 = delta * arma::trans(ones<vec>(J));
  arma::mat tmp3 = tmp1 % tmp2;
  arma::mat tmp4 = tmp3 - ones<mat>(N,J);
  arma::mat tmp5(N,J);
  for(int i=0;i<N;i++){
    for(int j=0;j<J;j++){
      if(tmp4(i,j)< 0) {
        tmp5(i,j) = 1;
      } else
      {
        tmp5(i,j) = 0;
      }
    }
  }

  arma::vec sped= kappa * (J - items);

  arma::vec tmp6(N);
  for(int i=0;i<N;i++){
    tmp6.row(i) = sped [delta(i)-1];
  }
  arma::mat sp_it = tmp6 * arma::trans(ones<vec>(J));
  arma::mat outp = sp_it % tmp5;
  return  outp;

}
arma::vec gencandd(arma::vec delta){

  int N = delta.size(); int J = delta.max();
  arma::vec newdelta(N);

  std::uniform_int_distribution<int> distribution(-4,4);
  for (int i=0; i < N; i++){

    // randomly add/sub number

    int tmp =delta[i] + distribution(g_generator);
    if(tmp <=0){
      newdelta[i] = 1;
    } else if(tmp>J){
```

```
      newdelta[i] = J;
    }else{
      newdelta[i] = tmp;
    }
  }
  return newdelta;
}
arma::vec priord1(arma::vec de, arma::mat pmat){
  int N = pmat.n_cols;
  arma::vec  priorprob(N);
  for (int i=0; i < N; i++) {
    priorprob[i] = pmat.col(i)[de[i]-1];
  }
  return priorprob;
}
arma::vec priord2(arma::vec d, arma::vec p){
  int N = d.size();
  arma::vec pd(N);
  for(int ii=0; ii<N; ii++) {
    int dd = d[ii];
    pd[ii] = p[dd-1];
  }
  return pd;
}
/////////////////////////////////////////////////////
// C5. MCMC General Functions
/////////////////////////////////////////////////////
arma::vec dec_ifelse(arma::vec uni, arma::vec lik){
  arma::vec C(uni.size());
  for(int j; j<uni.size();j++){
    if(uni[j]<=lik[j]){
      C[j]=1;
    } else{
      C[j]=0;
    }
  }
  return C;
}
arma::vec starvold(arma::vec uni, arma::vec star, arma::vec old){
  arma::vec D(uni.size());


  for(int j=0; j<uni.size();j++){
    if(uni[j]==1){
      D[j]=star[j];
    } else{
      D[j]=old[j];
    }

  }
  return D;
}
arma::mat logprobRcpp_speeded(arma::mat X,arma::vec t,
arma::vec a, arma::vec b, arma::mat e){
  int N = X.n_rows ;int J = X.n_cols ;
  arma::vec onetheta(J);
  arma::vec onealpha(N);
  arma::vec onebeta(N);

  onetheta.fill(0);
  onealpha.fill(0);
  onebeta.fill(0);

  arma::mat THETA = t * arma::trans(onetheta.ones());
  arma::mat ALPHA  = onealpha.ones() *  arma::trans(a);
  arma::mat BETA  = onebeta.ones() *  arma::trans(b);

  arma::mat ETA = e;

  arma::mat LOGIT  = exp(ALPHA % (THETA - BETA - ETA));
  arma::mat PROB = LOGIT/(1+LOGIT);
  arma::mat LL(N,J);

  LL = arma::log(PROB) % X + arma::log((1-PROB)) % (1-X);

  return LL;
}
arma::mat logprobRcpp_regular(arma::mat X,arma::vec t,
arma::vec a,arma::vec b){

  int N = X.n_rows ;int J = X.n_cols ;
  arma::vec onetheta(J);
  arma::vec onealpha(N);
  arma::vec onebeta(N);
```

```
  onetheta.fill(0);
  onealpha.fill(0);
  onebeta.fill(0);

  arma::mat THETA = t * arma::trans(onetheta.ones());
  arma::mat ALPHA  = onealpha.ones() *  arma::trans(a);
  arma::mat BETA  = onebeta.ones() *  arma::trans(b);


  arma::mat LOGIT  = exp(ALPHA % (THETA - BETA));
  arma::mat PROB = LOGIT/(1+LOGIT);
  arma::mat LL(N,J);
  LL.fill(0);

  LL = arma::log(PROB) % X + arma::log((1-PROB)) % (1-X);

  return LL;
}
/////////////////////////////////////////////////////
// C6. Sampling Functions
/////////////////////////////////////////////////////
/////////////////////////////////////////////////////
// 3. THETA
/////////////////////////////////////////////////////
////////////////////////////////////////////////////
// c) 2PL
void sampletheta_2pl_2chains(cur2PL_2chain& infoA,
cur2PL_2chain& infoB) {
  int N = infoA.N; int J = infoA.J;
  // A
  arma::vec thetaoldA = infoA.theta; arma::vec thetastarA =
  infoA.MHt*rnormcpp(N) + thetaoldA;
  // OLD-A
  arma::vec  tmpoldA   = rowSumsRcpp
  (logprobRcpp_regular(infoA.X,thetaoldA,infoA.alpha,infoA.beta));
  arma::vec  prioroldA = log(dnormcpp(thetaoldA,0.0,infoA.s2t));
  arma::vec  tmpoldprA = tmpoldA + prioroldA;
  // STAR-A
  arma::vec  tmpstarA   = rowSumsRcpp(logprobRcpp_regular(infoA.X,thetastarA,
  infoA.alpha,infoA.beta));
  arma::vec  priorstarA = log(dnormcpp(thetastarA,0.0,infoA.s2t));
  arma::vec  tmpstarprA = tmpstarA + priorstarA;
  arma::vec  likA = exp(tmpstarprA - tmpoldprA);
  arma::vec  uniA = runifcpp(N,0.0,1.0);
  arma::vec decA = dec_ifelse(uniA, likA);
  infoA.theta = starvold(decA,thetastarA,thetaoldA);

  //B
  arma::vec thetaoldB = infoB.theta; arma::vec thetastarB =
  infoB.MHt*rnormcpp(N) + thetaoldB;
  // OLD-B
  arma::vec  tmpoldB   = rowSumsRcpp(logprobRcpp_regular(infoB.X,thetaoldB,infoB.
  alpha,infoB.beta));
  arma::vec  prioroldB = log(dnormcpp(thetaoldB,0.0,infoB.s2t));
  arma::vec  tmpoldprB = tmpoldB + prioroldB;
  // STAR-B
  arma::vec  tmpstarB   = rowSumsRcpp(logprobRcpp_regular(infoB.X,thetastarB,
  infoB.alpha,infoB.beta)) ;
  arma::vec  priorstarB = log(dnormcpp(thetastarB,0.0,infoB.s2t));
  arma::vec  tmpstarprB = tmpstarB + priorstarB;
  arma::vec  likB = exp(tmpstarprB - tmpoldprB);
  arma::vec  uniB = runifcpp(N,0.0,1.0);
  arma::vec decB = dec_ifelse(uniB, likB);
  infoB.theta = starvold(decB,thetastarB,thetaoldB);
}
/////////////////////////////////////////////////////


/////////////////////////////////////////////////////
// 4. ALPHA
/////////////////////////////////////////////////////
/////////////////////////////////////////////////////////////////////////////////////////////
// c) 2PL
void samplealpha_2pl_2chains(cur2PL_2chain& infoA, cur2PL_2chain& infoB) {
  int N = infoA.N; int J = infoA.J;
  // A
  arma::vec  alphaoldA = infoA.alpha; arma::vec  alphastarA = infoA.MHa*rnormcpp(J) + alphaoldA;
  // OLD-A
  arma::vec  tmpoldA   = colSumsRcpp(logprobRcpp_regular(infoA.X,infoA.theta,alphaoldA,infoA.beta));
  arma::vec  prioroldA = log(dlnormcpp(alphaoldA,0.0,infoA.s2a));
  arma::vec  tmpoldprA = tmpoldA + prioroldA;
  // STAR-A
```

```
  arma::vec  tmpstarA   = colSumsRcpp(logprobRcpp_regular(infoA.X,infoA.theta,alphastarA,infoA.beta));
  arma::vec  priorstarA = log(dlnormcpp(alphastarA,0.0,infoA.s2a));
  arma::vec  tmpstarprA = tmpstarA + priorstarA;
  arma::vec  likA = exp(tmpstarprA - tmpoldprA);
  arma::vec  uniA = runifcpp(J,0.0,1.0);
  arma::vec decA = dec_ifelse(uniA, likA);
  infoA.alpha = starvold(decA,alphastarA,alphaoldA);

  // B
  arma::vec  alphaoldB = infoB.alpha; arma::vec  alphastarB = infoB.MHa*rnormcpp(J) + alphaoldB;
  // OLD-B
  arma::vec  tmpoldB   = colSumsRcpp(logprobRcpp_regular(infoB.X,infoB.theta,alphaoldB,infoB.beta));
  arma::vec  prioroldB = log(dlnormcpp(alphaoldB,0.0,infoB.s2a));
  arma::vec  tmpoldprB = tmpoldB + prioroldB;
  // STAR-B
  arma::vec  tmpstarB   = colSumsRcpp(logprobRcpp_regular(infoB.X,infoB.theta,alphastarB,infoB.beta));
  arma::vec  priorstarB = log(dlnormcpp(alphastarB,0.0,infoB.s2a));
  arma::vec  tmpstarprB = tmpstarB + priorstarB;
  arma::vec  likB = exp(tmpstarprB - tmpoldprB);
  arma::vec  uniB = runifcpp(J,0.0,1.0);
  arma::vec decB = dec_ifelse(uniB, likB);
  infoB.alpha = starvold(decB,alphastarB,alphaoldB);
}
/////////////////////////////////////////////////////////////////////////////////////////////////

/////////////////////////////////////////////////////
// 5. BETA
/////////////////////////////////////////////////////
/////////////////////////////////////////////////////////////////////////////////////////////////
// c) 2PL
void samplebeta_2pl_2chains(cur2PL_2chain& infoA, cur2PL_2chain& infoB) {
  int N = infoA.N; int J = infoA.J;
  // A
  arma::vec  betaoldA = infoA.beta;arma::vec  betastarA = infoA.MHb*rnormcpp(J) + betaoldA;
  // OLD-A
  arma::vec  tmpoldA   = colSumsRcpp(logprobRcpp_regular(infoA.X,infoA.theta,infoA.alpha,betaoldA));
  arma::vec  prioroldA = log(dnormcpp(betaoldA,0.0,infoA.s2b));
  arma::vec  tmpoldprA = tmpoldA + prioroldA;
  // STAR-A
  arma::vec  tmpstarA   = colSumsRcpp(logprobRcpp_regular(infoA.X,infoA.theta,infoA.alpha,betastarA)) ;
  arma::vec  priorstarA = log(dnormcpp(betastarA,0.0,infoA.s2b));
  arma::vec  tmpstarprA = tmpstarA + priorstarA;
  arma::vec  likA = exp(tmpstarprA - tmpoldprA);
  arma::vec  uniA = runifcpp(J,0.0,1.0);
  arma::vec  decA = dec_ifelse(uniA, likA);
  infoA.beta = starvold(decA,betastarA,betaoldA);
  // B
  arma::vec  betaoldB = infoB.beta;arma::vec  betastarB = infoB.MHb*rnormcpp(J) + betaoldB;
  // OLD-B
  arma::vec  tmpoldB   = colSumsRcpp(logprobRcpp_regular(infoB.X,infoB.theta,infoB.alpha,betaoldB));
  arma::vec  prioroldB = log(dnormcpp(betaoldB,0.0,infoB.s2b));
  arma::vec  tmpoldprB = tmpoldB + prioroldB;
  // STAR-B
  arma::vec  tmpstarB   = colSumsRcpp(logprobRcpp_regular(infoB.X,infoB.theta,infoB.alpha,betastarB)) ;
  arma::vec  priorstarB = log(dnormcpp(betastarB,0.0,infoB.s2b));
  arma::vec  tmpstarprB = tmpstarB + priorstarB;
  arma::vec  likB = exp(tmpstarprB - tmpoldprB);
  arma::vec  uniB = runifcpp(J,0.0,1.0);
  arma::vec  decB = dec_ifelse(uniB, likB);
  infoB.beta = starvold(decB,betastarB,betaoldB);
}
/////////////////////////////////////////////////////////////////////////////////////////////////

/////////////////////////////////////////////////////
// 7. Controlling Sampling Functions
/////////////////////////////////////////////////////
/////////////////////////////////////////////////////////////////////////////////////////////////
/////////////////////////////////////////////////////////////////////////////////////////////////
// c) 2PL
void bmcmcup_2pl_2chains(cur2PL_2chain& infoA, cur2PL_2chain& infoB){
 sampletheta_2pl_2chains(infoA,infoB);
 samplealpha_2pl_2chains(infoA,infoB);
 samplebeta_2pl_2chains(infoA,infoB);
}
/////////////////////////////////////////////////////////////////////////////////////////////////
/////////////////////////////////////////////////////
// 8. Main MCMC Functions
/////////////////////////////////////////////////////
/////////////////////////////////////////////////////////////////////////////////////////////////
// c) 2PL
contain2_2PL mainfun_2pl_2chains(cur2PL_2chain & infoA, cur2PL_2chain & infoB){
  // Label important variables for simulations
  int bi   = infoA.BI;    // Burn-In
  int keep = infoA.IT;    // Iterations Kept
```

```
  int thin = infoA.thin;  // Thinning
  int N    = infoA.N;     // Sample Size
  int J    = infoA.J;     // Number of Items


  // Burning
  for(int ii=0; ii<bi; ii++) {
    bmcmcup_2pl_2chains(infoA,infoB);
    cout << "Burn-In: "<< ii+1 << endl;
  }


  arma::mat matthetaA(N,(keep/thin));
  arma::mat matthetaB(N,(keep/thin));
  arma::mat matalphaA(J,(keep/thin));
  arma::mat matalphaB(J,(keep/thin));
  arma::mat matbetaA( J,(keep/thin));
  arma::mat matbetaB( J,(keep/thin));

  matthetaA.fill(0);
  matthetaB.fill(0);
  matalphaA.fill(0);
  matalphaB.fill(0);
  matbetaA.fill(0);
  matbetaB.fill(0);

  int count=0;
  for(int i=0; i<keep; i++) {
    bmcmcup_2pl_2chains(infoA,infoB);

    cout << "Real Iterations: "<< i+1 << endl;

    // Thinning Process HOLLLLA:)
    if((i % infoA.thin)==0){
      int pl = (count);
      matthetaA.col(pl)  = infoA.theta;
      matthetaB.col(pl)  = infoB.theta;
      matalphaA.col(pl)  = infoA.alpha;
      matalphaB.col(pl)  = infoB.alpha;
      matbetaA.col(pl)  = infoA.beta;
      matbetaB.col(pl)  = infoB.beta;
      count ++;
    }
  }
  contain2_2PL info_contain_2PL  =
    {
    // Container for simulations estimates
    matthetaA,matthetaB,
    matalphaA,matalphaB,
    matbetaA,matbetaB
    };
  return info_contain_2PL;

}
////////////////////////////////////////////////////////////////////////////////////////////////////
```

Parallel Programming

```
#include <stdio.h>
#include <stdlib.h>
#include <pthread.h>
#include <iostream>
#include <string>
#include <vector>

using namespace std;

// command arguments
char *g_progname;        // The name of the command to run
char *g_filelist;        // Path to the file containing arguments to split
int   g_nProcesses;      // How many cores does your machine have?

// the list of tasks from a file
vector<string> g_inputs;

void *runprogram(void *pRank);

int main(int argc, char **argv)
{

  //
  // 1. Start Timer for Code *Only needed for One Replication
  //
  clock_t t1,t2;
  t1=clock();


  char line[4096];

  //
  // Read command arguments
  //
  if (argc != 4) {
    fprintf(stderr, "usage:\n");
    fprintf(stderr, "  ./runProgram progname filelist nProcesses\n");
    return 1;
  }
  g_progname   =      argv[1];
  g_filelist   =      argv[2];
  g_nProcesses = atoi(argv[3]);

  //
  // Read the file
  //
  FILE *fin = fopen(g_filelist, "r");
  if (fin == NULL) {
    fprintf(stderr, "ERROR: could not open %s for reading\n", g_filelist);
    return 1;
  }
  while (!feof(fin)) {
    fscanf(fin, "%s", line);
    fprintf(stderr, "input[%d]: %s", (int)g_inputs.size(), line);
    g_inputs.push_back( line );
    printf("\n");
  }
  fclose(fin);

  //
  // Spawn each task manager individually
  //
  pthread_t *threads = (pthread_t*)malloc(g_nProcesses*sizeof(pthread_t));
  for (int i=0; i<g_nProcesses; i++) {
    pthread_create(&(threads[i]), NULL, runprogram, (void*)i);
  }

  //
  // Wait for each manager to complete
  //
  for (int i=0; i<g_nProcesses; i++) {
    pthread_join(threads[i], NULL);
  }
  free(threads);

  printf("runProgram Success!\n");

  // //
  // // 13. Start Timer for Code *Only needed for One Replication
```

```
  // //
  // t2=clock();
  // float diff ((float)t2-(float)t1);
  // float seconds = diff / CLOCKS_PER_SEC;
  // cout<<seconds<<endl;
  //
  // return 0;
}

void *runprogram(void *pRank)
{
  char cmd[4096];

  // The rank is passed "as" a pointer
  //  (because a pointer is an integer)
  size_t rank = (size_t)pRank;

  // The start and end task are based on the rank
  int nTasks = g_inputs.size();
  int startTask = nTasks *  rank    / g_nProcesses;
  int   endTask = nTasks * (rank+1) / g_nProcesses;

  for (int i=startTask; i<endTask; i++) {
    fprintf(stderr, "--------------------\n");
    fprintf(stderr, "- task manager %d running task %s\n", (int)rank, g_inputs[i].c_str());
    fprintf(stderr, "--------------------\n");
    system("mkdir -p progress");

    sprintf(cmd, "./%s %s > progress/%s\n", g_progname, g_inputs[i].c_str(), g_inputs[i].c_str());
    printf("%s\n", cmd);
    system(cmd);
  }
  return NULL;
}
```

Appendix E

Item Parameters

Table E.1: The item parameters when $J = 20$ and Item Difficulty ordering is Random

Item Parameters			
		α	β
Item	1	1	-1.008
Item	2	1.55	1.313
Item	3	0.991	1.458
Item	4	1.729	0.753
Item	5	0.943	-1.639
Item	6	1.773	0.52
Item	7	1.625	-0.287
Item	8	1.459	0.257
Item	9	1.12	-0.403
Item	10	1.224	1.859
Item	11	1.665	-0.33
Item	12	1.604	-1.59
Item	13	0.903	-0.401
Item	14	1.462	1.651
Item	15	1.736	-0.982
Item	16	0.846	0.515
Item	17	1.042	-0.718
Item	18	1.74	0.19
Item	19	1.289	-1.049
Item	20	1.026	1.016

Table E.2: The item parameters when $J = 40$ and Item Difficulty ordering is Random

		Item Parameters	
		α	β
Item	1	1	-1.008
Item	2	1.55	1.313
Item	3	0.991	1.458
Item	4	1.729	0.753
Item	5	0.943	-1.639
Item	6	1.773	0.52
Item	7	1.625	-0.287
Item	8	1.459	0.257
Item	9	1.12	-0.403
Item	10	1.224	1.859
Item	11	1.665	-0.33
Item	12	1.604	-1.59
Item	13	0.903	-0.401
Item	14	1.462	1.651
Item	15	1.736	-0.982
Item	16	0.846	0.515
Item	17	1.042	-0.718
Item	18	1.74	0.19
Item	19	1.289	-1.049
Item	20	1.026	1.016
Item	21	1	-1.008
Item	22	1.55	1.313
Item	23	0.991	1.458
Item	24	1.729	0.753
Item	25	0.943	-1.639
Item	26	1.773	0.52
Item	27	1.625	-0.287
Item	28	1.459	0.257
Item	29	1.12	-0.403
Item	30	1.224	1.859
Item	31	1.665	-0.33
Item	32	1.604	-1.59
Item	33	0.903	-0.401
Item	34	1.462	1.651
Item	35	1.736	-0.982
Item	36	0.846	0.515
Item	37	1.042	-0.718
Item	38	1.74	0.19
Item	39	1.289	-1.049
Item	40	1.026	1.016

Appendix F

Data Generation Algorithm

Within the 2PLMix and M2PLMix models, the location at which speededness begins was generated under the Multinomial$[\pi_1, \pi_2, ..., \pi_J]$ and Multinomial$[\pi_1(\theta_i), \pi_2(\theta_i), ..., \pi_J(\theta_i)]$, respectively. The ability parameter, θ_i, was previously generated which was then used to determine the examinee i's probability distribution of not being speeded based on Eq. 3.4. The development of the probability distribution for speededness with respect to the 2PLMix and M2PLMix models can be seen in the code below. The following code also shows how η_i was calculated based on a δ_i for the 2PlMix model and δ_i and θ_i for the M2PLMix model.

```
kappa <- 0.2; omega <- 2; lambda <- 0.4; pi <- 0.4
################## 2PLMix #####################
## 1. Vector of items from 1 to J - 1
items=1:(J[1]-1)

## 2. Probability of being speeded on items 1 to J-1
prob1_Jm1_2 = as.matrix(((items^omega-(items-1)^omega)/
((J-1)^omega))
* (1-pi))

## 3. Probability of being speeded on items 1 to J
prob_2PLMix <- as.matrix(as.numeric(rbind(prob1_Jm1_2,pi))

## 4. Generate Delta from Distribution
delta <- sample(1,J, prob_2PLMix)

## 5.
## 5. Generate Eta via Two function
gen.eta = function(kap,del,J){
##  Equation 2.9 to determine
## the speededness effect
gam = rbind(as.matrix(kap*((J-1):1)),0)
## Determine the speededness effect for
## examinee i based on delta
amount.sp <- gam[del,]
## Information used to make vector of
## speededness effect for examinee i
aff.spot <- cbind(del,amount.sp,J)
## Use of function to make
## vector of speededness effect
eta = t(apply(aff.spot,1,replace.val))
return(eta)
}
replace.val = function(spot.aff){
## Determine the number of items
Jc = spot.aff[3]
## Create a vector of 0s
tmp = rep(0,Jc)
## Replace location where speededness
## begins by speededness effects
tmp[spot.aff[1]:Jc]=spot.aff[2]
return(tmp)
}

eta <-  gen.eta(kappa,delta,J)
##########################
```

```
################## M2PLMix #####################
## 1. Baseline probability of not
## being speeded for examinee I EQ: 3.4
pi.lambda = lambda + (1-lambda)*(1 / (1+exp(-theta[i])))

## 2. Probability of being speeded on items 1 to J-1
prob1_Jm1_m2 = as.matrix(((items^omega-(items-1)^omega)/
((J-1)^omega)) *
(1- pi.lambda))

## 3. Probability of being speeded on items 1 to J
prob_M2PLMix <- as.matrix(as.numeric(rbind(prob1_Jm1_m2,
pi.lambda))

## 4. Generate Delta from Distribution
delta <- sample(1,J, prob_M2PLMix)

## 5. Generate Eta based on previously defined functions
eta <-  gen.eta(kappa,delta,J)
##########################
```

Under the GPC model, the ability parameter, θ_i, the proportion of the test complete, η_i, and speededness intensity, λ_i, for examinee i, in which an association was created using a copula function. This was done by the following steps

1. Define the mean parameters used within marginal distribution
2. Define the correlation structure used to associate speededness parameters and ability
3. Generate 3 parameters for each examinee from a multinomial distribution with means based on the desired distributions and correlation structure from the 2^{nd} step
4. Find the cumulative distribution (probability) value for each value that was generated from step 3
5. Use each probability value within each marginal distribution to find the corresponding quantile, marginal distributions can be seen as follows:
   - $\theta \sim N(0,1)$
   - $\eta \sim \beta(2,2)$
   - $\lambda \sim Log - N(0,1)$

The mechanism for associating ability with speededness in the GPC model can been in the following code:

```
require(mvtnorm)
N <- 1000;J<- 40
## Correlations between
## t: theta; e: eta; l: lambda;
clt<-0.5;cle<-0.25;cet<- 0.25
## Means
m_t <- 0; m_e <- 2 ; m_l <- 1
#Correlation matrix
S <- matrix(c(1,cet,cle,cet,1,
clt,cle,clt,1),3,3)
#Our gaussian variables
AB <- rmvnorm(mean=c(m_t,m_e,m_l),sig=S,n=N)
U <- pnorm(AB)
## Distribution between parameters
theta  <- qnorm(U[,1],)
eta    <- qbeta(U[,2],2,2)
lambda <- qlnorm(U[,3],1)
```

Once these parameters were generated with respect to each model, they were used to find the i^{th} examinee's probability of getting the j^{th} item correct noted as P_{ij}. This implies that an NxJ matrix of probabilities was created. This is done by using the parameters for each model to determine the probability of obtaining the correct solution. A response matrix of $0's$ and $1's$ was created based on these probabilities based on the following equation,

$$X_{ij} = \begin{cases} 1 & P_{ij} > U \\ 0 & P_{ij} \leq U \end{cases}$$

in which $U \sim U(0,1)$, is a continuous distribution.

Appendix G

Estimation Algorithm

The following are the steps used to estimate θ, α, β, and δ. The priors and full conditional distributions for each parameter were discussed in section 3.3

Algorithm for θ_i

1. Initialize θ_i^{k-1} is found from the Z-score of $0.025 + 0.95 * \sum_i X_i$ (previous iterate)
2. Determine θ_i based on the following:

$$\theta_i = \theta_i^{k-1} + 0.2 * \epsilon$$

   where $\epsilon \sim N(0, 1)$

3. Calculate the prior probabilities for θ_i^{k-1} and θ_i using Normal$(0, 1)$
4. Calculate the likelihood of $f(\theta_i^{k-1}|X, rest)$ and $f(\theta_i|X, rest)$, $rest$ are the other parameters being estimated
5. Multiply the prior probability of θ_i to the likelihood of θ_i for each item for the previous iterate and candidate parameter to determine the acceptance probability $\alpha(\theta_i, \theta_i^{k-1})$
6. The previous iterate is determined by the following:

$$\theta_i^{k-1} = \begin{cases} \theta_i^{k-1} & \alpha(\theta_i, \theta_i^{k-1}) < X \\ \theta_i & \alpha(\theta_i, \theta_i^{k-1}) > X \end{cases}$$

   where $X \sim U(0, 1)$, a continuous uniform distribution.

Algorithm for α_j

1. Initialize $\alpha_j^{k-1} = 1$ (previous iterate)
2. Determine α_j based on the following:

$$\alpha_j = \alpha_j^{k-1} + 0.15 * \epsilon$$

   where $\epsilon \sim N(0, 1.185)$

3. Calculate the prior probabilities for α_j^{k-1} and α_j using Log-Normal$(0.3, 1)$

4. Calculate the likelihood of $f(\alpha_j^{k-1}|X, rest)$ and $f(\alpha_j|X, rest)$, $rest$ are the other parameters
5. Multiply the prior probability of α_j to the likelihood of α_j for each examinee for the previous iterate and candidate parameter to determine the acceptance probability $\alpha(\alpha_j, \alpha_j^{k-1})$
6. The previous iterate is determined by the following:

$$\alpha_j^{k-1} = \begin{cases} \alpha_j^{k-1} & \alpha(\alpha_j, \alpha_j^{k-1}) < X \\ \alpha_j & \alpha(\alpha_j, \alpha_j^{k-1}) > X \end{cases}$$

where $X \sim U(0,1)$, a continuous uniform distribution.

Algorithm for β_j

1. Initialize β_j^{k-1} is found from the -Z-score of $0.025 + 0.95 * \sum_j X_j$ (previous iterate)
2. Determine β_j based on the following:

$$\beta_j = \beta_j^{k-1} + 0.25 * \epsilon$$

where $\epsilon \sim N(0, 35)$

3. Calculate the prior probabilities for β_j^{k-1} and β_j using Normal$(0,1)$
4. Calculate the likelihood of $f(\beta_j^{k-1}|X, rest)$ and $f(\beta_j|X, rest)$, $rest$ are the other parameters being estimated
5. Multiply the prior probability of β_j to the likelihood of β_j for each examinee for the previous iterate and candidate parameter to determine the acceptance probability $\alpha(\beta_j, \beta_j^{k-1})$
6. The previous iterate is determined by the following:

$$\beta_j^{k-1} = \begin{cases} \beta_j^{k-1} & \alpha(\beta_j, \beta_j^{k-1}) < X \\ \beta_j & \alpha(\beta_j, \beta_j^{k-1}) > X \end{cases}$$

where $X \sim U(0,1)$, a continuous uniform distribution.

Algorithm for δ_i

1. Initialize $\delta_i^{k-1} = J$ (previous iterate)
2. Determine δ_i based on the following:

$$\delta_i = \begin{cases} 1 & \delta_i^{k-1} + \epsilon < 1 \\ J & \delta_i^{k-1} + \epsilon > J \\ \delta_i^{k-1} + \epsilon & \text{otherwise} \end{cases}$$

where $\epsilon \sim U(-3,3)$, a discrete uniform distribution.

3. Determine η_i^{k-1} and η_i based on δ_i^{k-1} and δ_i

$$\eta = (0, 0, 0, ..., \gamma_\delta, \gamma_\delta, \gamma_\delta, ..., \gamma_\delta)$$

   where $\gamma_\delta = \kappa(J - \delta)$

4. Calculate the prior probabilities for δ_i^{k-1} and δ_i based on a Multinomial $[\pi_1(\theta_i), \pi_2(\theta_i), ..., \pi_J(\theta_i)]$

5. Calculate the likelihood of $f(X|\delta_i^{k-1})$ and $f(X|\delta_i)$, based on η_i^{k-1} and η_i

6. Multiply the prior probability of δ_i to the likelihood of δ_i for each item for the previous iterate and candidate parameter to determine the acceptance probability $\alpha(\delta_i, \delta_i^{k-1})$

7. The previous iterate is determined by the following:

$$\delta_i^{k-1} = \begin{cases} \delta_i^{k-1} & \alpha(\delta_i^{k-1}, \delta_i) < X \\ \delta_i & \alpha(\delta_i^{k-1}, \delta_i) > X \end{cases}$$

   where $X \sim U(0, 1)$, a continuous uniform distribution.

The algorithm found the iterates for the parameters in the following ordered:

1. θ
2. α
3. β
4. δ