Metric Identification in Mixture IRT Models

by

Youn-Jeng Choi

(Under the Direction of Allan S. Cohen)

Abstract

In item response theory (IRT), the origin and unit of the ability scale in IRT are arbitrary. This arbitrariness is referred to as scale indeterminacy or the identification problem. Standard IRT models may not fit the data when there is unexplained heterogeneity present. In such cases, a mixture IRT model, which models this heterogeneity by fitting an IRT model to latent classes in the data, may be useful. The purpose of this study was to explore the effect of three different kinds of constraints for identifying the metric in the mixture IRT (MixIRT) model: (1) equating in which an anchor item is used to anchor the metrics between latent classes, (2) person centering in which the mean of the ability parameters is set to zero after each calibration, and (3) item centering in which the mean of the item difficulty parameters is set to zero. Results based on an analysis of the empirical data indicated that the number of latent classes detected differed depending on the particular MixIRT model and constraint combination. The mean ability, proportion of group memberships, and item parameters also differed between the three constraints. Results of a simulation study are presented followed by an illustrative example using real data from the TIMSS 2011 8th grade science test. In the simulation study, the impact of the three identification methods was examined on classifications of latent class memberships and on item and ability parameter estimates for three dichotomous MixIRT models. There was no effect of identification constraint on the MixRM and Mix2PLM. Only the item anchoring constraint was found to work well with the Mix3PLM, although recovery was relatively poor for this model compared to the MixRM and Mix2PLM. When the types of constraint were compared, the person centering constraint produced the worst recovery results. Test length and sample size did not appear to have an effect on the recovery of item parameters. The longer test length improved group member ship identification. Percent of correct model selection using AIC was lower for the larger sample size. Recovery or group membership, item difficulty, and item discrimination decreased with an increase in the number of latent classes simulated. Recovery of the lower asymptote, however, was slightly better for the larger sample size and for more latent classes.

INDEX WORDS: Scale identification, Identification problem, mixture item response model, Bayesian analysis, TIMSS 2011 science test

Metric Identification in Mixture IRT Models

by

Youn-Jeng Choi

B.A., Ewha Womans University, Seoul, Korea, 2003

M.A., Ewha Womans University, Seoul, Korea, 2006

A Dissertation Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2014

Metric Identification in Mixture IRT Models

by

Youn-Jeng Choi

Approved:

Major Professor:	Allan S. Cohen
Committee:	Seock-Ho Kim
	Gary J. Lautenschlager
	Zhenqui Lu

Electronic Version Approved:

Julie Coffield
Interim Dean of the Graduate School
The University of Georgia
August 2014

DEDICATION

This is dedicated to dear Al, Dr. Seong, my family, and God.

Contents

List of Figures

List of Tables

Chapter 1

STATEMENT OF THE PROBLEM

The invariance of item and person statistics is an important assumption of item response theory (Lord, 1980; Lord & Novick, 1968). Under this assumption, item and person parameters remain invariant over different samples from the same population and over different combinations of items that fit the model (Lord, 1980). A useful benefit of this assumption is that invariance permits establishing a metric that does not change either when the same test is given to new examinees or when new items are calibrated to the same model. The assumption of invariance is one of the characteristics that differentiates IRT from classical test theory (Hambleton, Swaminathan, & Rogers, 1991).

The assumption of invariance also poses a significant problem, however, as the choice of the origin and scale for the ability metric are arbitrary (Lord, 1980). This arbitrariness is referred to as scale indeterminacy or the metric identification problem (Baker & Kim, 2004). Various methods to identify (or fix) the metric have been developed for IRT models.

When the same IRT model does not fit all members of a population, a Mixture IRT model (MixIRTM) may be appropriate. The MixIRTM is formed by an integration of an IRT model with a latent class model (Cho, Cohen, & Kim, 2013). The IRT portion of the model estimates a continuous latent variable and the latent class portion of the model estimates a categorical latent variable. Combining these two models permits examining the possibility that a population of examinees can be classified into some number of discrete latent classes, and that item and ability parameters may differ for the different classes (Bolt, Cohen, & Wollack, 2002).

Characterizing members of different latent classes is important in interpreting the meaning of the classes. Comparison of item parameter estimates between latent classes is one approach for characterizing the latent classes (Rost, 1990). In order to make such comparisons, however, the latent classes need to have a common metric. That is, estimates of model parameters need to be on the same scale in order for comparisons to be made.

There are currently three methods commonly used for developing a common metric between latent classes. The first method is concurrent calibration in which one or more items are used to anchor the metrics between classes (Bolt et al., 2002; Choi, Alexeev, & Cohen, 2014). The second method is to impose equality constraints by setting the mean of one latent class to zero and its standard deviation to one (Baker & Kim, 2004; Cho, Cohen, & Kim, in press; Cho, Cohen, & Templin, 2008; De Boeck, Cho, & Wilson, 2011; von Davier & Yamamoto, 2004). A third method is setting the sum of item difficulties to zero for each latent group (Cho & Cohen, 2010; Dai & Mislevy, 2006; Rost, 1990; Samuelsen, 2008). Although each of these methods has been reported in the literature, to date, relatively little research exists investigating the impact of these constraints on metric identification in MixIRT models.

A commonly used method is one suggested by Rost (1990) for the mixture Rasch model (MixRM) in which the mean of item difficulties is set to zero. There is somewhat less agreement, however, about constraints used for identification for the mixture 2PL model (Mix2PLM) or mixture 3PL model (Mix3PLM). Results from Choi, Alexeev, Cohen, and Kim (2010) for the Mix3PLM indicated that setting the within class average of item difficulties to zero worked well with respect to recovery of generating parameters. Likewise, fixing the mean and standard deviation of the ability estimates to zero and one, respectively, also worked well. Comparison of these three identification constraints suggested the latter constraint had no effect on the accuracy of parameter recovery. These results were based on only five replications, however, and for a relatively small number of conditions.

The purpose of this study was to explore the effects of these three methods for establishing a common metric between latent classes in MixIRT models. An example is provided using real data from the TIMSS 2011 Grade 8 Science Test to motivate the simulation study. The impact of each of these methods was then examined in the context of a simulation study on selection of the correct (i.e., the generating) model, recovery of item and latent group mean parameter estimates, and selection of the correct latent class for each examinee.

Chapter 2

THEORETICAL FRAMEWORK

2.1 Mixture Item Response Theory Models

Item response theory (IRT) is composed of a family of statistical models each designed to model the relationship between a continuous latent ability and performance on a test item. As an example, the 2-parameter logistic model (2PLM) shown below in Equation 2.1 gives the probability of a correct response for examinee j on item i:

$$P(y_{ij} = 1|\theta_j) = \frac{\exp[a_i(\theta_j - b_i)]}{1 + \exp[a_i(\theta_j - b_i)]}, \tag{2.1}$$

where θ_j is the latent ability for examinee j, a_i is the item discrimination of item i, and b_i is the difficulty of item i.

When item performance is scored dichotomously (e.g., zero for an error and one for a correct response), there are three standard models that tend to be used to model this relationship. Using the logistic form of these models, they are the 1-parameter logistic model (1PLM), the 2-parameter logistic model (2PLM), and the 3-parameter logistic model (3PLM). Each of these models assumes that latent variable (e.g., ability) is measured the same for all members of a population.

Mixture IRT (MixIRTMs) assume that there may be groups or classes of examinees that are latent in the population and for which the same IRT model does not hold. That is, there may be different values for model parameters depending on the latent class to which a given examinee belongs. The MixIRTM can be viewed as a combination of an IRT model and a latent class model.

MixIRTMs assume that an examinee population is composed of a fixed number of discrete latent classes of examinees (Cohen, Wollack, Bolt, & Mroch, 2002). All examinees who belong to a certain latent class are assumed to have unique characteristics and are assumed to be homogeneous on the categorical latent variable that differentiates one class from the others. These models may be appropriate, in other words, when a single IRT model is not the best fit to the data.

The Mixture Rasch Model (MixRM). The MixRM (Rost, 1990) is the simplest of the dichotomous MixIRTMs and is based on the assumption that an examinee population is composed of a number of discrete latent classes, each of which has unique item and ability parameters (Cohen et al., 2002). The MixRM in Equation (2.2) associates a class membership parameter, g, with each examinee. Class membership decides the relative difficulty of the items for an examinee in that class. Additionally, g also determines a latent ability parameter, θ_{jg}, which then has an effect on determination of the number of correct answers on the test. The probability of a correct response in the MixRM can be written as

$$P(y_{ij} = 1|\theta_{jg}) = \Sigma_{g=1}^{G} \pi_g \frac{\exp(\theta_{jg} - b_{ig})}{1 + \exp(\theta_{jg} - b_{ig})}, \tag{2.2}$$

where g is an index for the latent class, $g = 1, \ldots, G$; $j = 1, \ldots, N$ examinees; θ_{jg}, is the latent ability of a examinee j within class g; π_g is the proportion of examinees for each class; and b_{ig} is the Rasch difficulty parameter of item i for latent class g. In this way, the

MixRM assumes that the Rasch model fits in each latent class but may have different item and ability parameters.

The Mixture 2-Parameter Logistic Model (Mix2PLM). The Mix2PLM can be viewed as a relaxed version of the MixRM in which the item discrimination parameter is left unconstrained. Similar to the MixRM, the 2-parameter logistic model (2PLM) is assumed to hold for each class, but each class may have unique item difficulty and discrimination parameters and different ability parameters. As with the MixRM, each examinee is parameterized both by a class membership parameter (g = 1,...,G) and a within-class ability parameter (θ_{jg}). That is, the Mix2PM also associates a class membership parameter, g, with each examinee as well as a latent ability parameter, θ_g. The probability of a correct response in the Mix2PLM can be written as

$$P(y_{ij} = 1|\theta_{jg}) = \Sigma_{g=1}^{G} \pi_g \frac{\exp[a_{ig}(\theta_{jg} - b_{ig})]}{1 + \exp[a_{ig}(\theta_{jg} - b_{ig})]}, \tag{2.3}$$

where g is an index for latent class, $g = 1, \ldots, G$; j is the jth examinee among N examinees; θ_{jg} is the latent ability of examinee j within class g; π_g is the proportion of examinees for each class; a_{ig} is the discrimination parameter for item i in class g; and b_{ig} is the difficulty parameter for item i in class g.

The Mixture 3-Parameter Logistic Model (Mix3PLM). The Mix3PLM can be viewed as an extension of a Mix2PLM in which a term is added to model the lower asymptote of the item response function. A 3-parameter logistic model is assumed to hold for each class in the Mix3PLM. Item and ability parameters are allowed to differ between latent classes. Each examinee is parameterized both by a class membership parameter ($g = 1, \ldots, G$) and a within-class ability parameter (θ_{jg}). The probability of a correct response in the Mix3PLM can be written as

$$P(y_{ij} = 1|\theta_{jg}) = \Sigma_{g=1}^{G}\pi_g \left[c_{ig} + (1 - c_{ig})\frac{\exp[a_{ig}(\theta_{jg} - b_{ig})]}{1 + \exp[a_{ig}(\theta_{jg} - b_{ig})]}\right], \tag{2.4}$$

where g is an index for latent class ($g = 1, \ldots, G$), j is the jth examinee among N examinees (j = 1, ..., N examinees), θ_{jg} is the latent ability of examinee j within class g, π_g is the proportion of examinees for each class, a_{ig} is the discrimination parameter for item i in class g, b_{ig} is the difficulty parameter for item i in class g, and c_{ig} is the lower asymptote parameter for item i in class g.

2.2 Scale Identification

The property of invariance of item and ability parameters in IRT is an important difference from classical test theory (Hambleton et al., 1991). This property implies that item and ability parameters do not depend on characteristics of the examinee sample or the specific set of items used to measure ability. Thus, it allows for comparison of item and ability parameter estimates from different sets of items or from different samples of examinees by linking them to a common metric.

The choice of origin for the ability metric is arbitrary. That is, adding the same constant to every θ_j and to every b_i in this model does not change the value of $a_i(\theta_j - b_i)$ and so p_{ij} also remains unchanged. Similarly, if we multiply every θ_j and every b_i by the same constant, and divide every a_i by the same constant, there is no change in the value of the term $a_i(\theta_j - b_i)$ or of P_{ij} (Lord, 1980). Lord (1980) notes that item parameters will remain invariant for groups from the same population as long as the ability scale is not changed. This property is referred to as scale indeterminacy or metric indeterminacy.

Thus, it is necessary to fix the metric to a particular origin and unit in order to locate it.

There are three methods that are commonly used in IRT to fix the metric: item anchoring, person centering, and item centering (de Ayala, 2009). Item anchoring is usually used for multiple group analysis. In this method, some items are fixed so that the same parameters are used across groups. This is done based on either theoretical or empirical considerations or both [e.g., the items are known to function the same in each group (Bolt et al., 2002; Choi et al., 2014; Choi, Cohen, Lu, & Kim, 2014)]. Person centering sets the mean of the ability parameters to zero or the mean and standard deviation of the ability parameters to zero and unity during calibration of model parameters (Cho & Cohen, 2010; Finch & Pierson, 2011). The third method, item centering, fixes the mean of item difficulty parameters to zero during calibration (Bolt, Cohen, & Wollack, 2001; Izsák, Orrill, Cohen, & Brown, 2010; Meiser & Machunsky, 2008; Meyer, 2008). These three methods are described below in the context of IRT and are then discussed as they apply in the context of MixIRT.

Methods for Solving the Identification Problem

Item anchoring may be used when there are either theoretical or empirical reasons for fixing some set of items to given values. If item parameters are known, for example, it is possible to fix the item parameters at known values in each group. When multiple groups are analyzed, therefore, these items may be used as anchors to link the metric across groups. As an example, in the likelihood ratio test for differential item functioning (DIF), all item parameter estimates can be constrained to the same values in each group except those of the studied item (Thissen, Steinberg, & Wainer, 1993). Then the item parameters of the studied item are estimated in each group. In the person centering method, the mean of the ability parameters is set to zero after each calibration. Person centering is used in programs such as LOGIST (Wingersky, Barton, & Lord, 1982), BILOG-MG (Zimowski, Muraki, Mislevy, & Bock, 2003), PARSCALE (Muraki & Bock, 2003), and MULTILOG (Thissen, Chen, & Bock, 2003). Item centering sets the mean of the item difficulty parameters to zero following calibration. IRT programs such as WINSTEPS (Linacre, 2001a), BIGSTEPS (Linacre & Wright, 2001), and FACETS (Linacre, 2001b) use item centering (de Ayala, 2009).

Item anchoring, person centering, and item centering for handling the identification problem are also used for MixIRTMs. Item anchoring is used either by constraining some subset of items to equality between groups or by fixing some subset of items to the same values in each group. In either case, the specific subset of items will have the same item parameters in each group. The second method, person centering, is to impose equality constraints for some reference class by setting the mean of one group to zero and the unit of scale (i.e., its standard deviation) to one. The item and ability parameter estimates for the other groups are then estimated relative to the estimates for the reference group. The third method, item centering, is done by setting the sum of item difficulties to zero for each latent group. The WINMIRA program (von Davier, 2001) uses this type of item centering. Programs such as M-plus (Muthén & Muthén, 2012) and OpenBUGS (Spiegelhalter, Thomas, Best, & Lunn, 2007) allow researchers to impose any of these three methods.

In the next section, an empirical example is presented to illustrate the potential impact of the different identification constraints on establishing a common metric in MixIRTMs.

2.3 Empirical Example

Data

For this example, data were taken from the TIMSS 2011 Grade 8 Test. The TIMSS 2011 Test consists of six sets of questions: A mathematics test, a science test, a student background questionnaire, a teacher background questionnaire (focusing on mathematics and science teaching), a school background questionnaire, and a curriculum questionnaire. The 17 multiple-choice items and 8 short constructed response items (scored dichotomously) from the 2011 TIMSS 8th grade science test were analyzed for this example.

The multiple-choice items were scored correct or incorrect, and blanks were skipped and not scored. In addition, the short answer items were scored as either correct or incorrect, and blank items were skipped. The items selected for this example assessed four content domains: Biology (8 items), Chemistry (6 items), Physics (4 items), and Earth science (7 items).

Sample. Data from seven of the 45 countries participating in the TIMSS 2011 program were used for this example. The sample of 2,493 students in this data set were from the following countries: 357 students from Chinese Taipei, 410 students from Ghana, 361 students from the Republic of Korea, 464 students from Morocco, 247 students from Norway, 423 students from Singapore, and 231 students from The Ukraine. The seven nations were selected, because, as a group, their average scale scores on the test approximated high, middle and low achievement among the participating countries. Singapore, Chinese Taipei, and the Republic of Korea had the highest mean mathematics scores of 590, 564, and 560, respectively. The mean scores for Ukraine and Norway of 501 and 494, respectively, were average among participating countries. Mean scores for Morocco and Ghana were 376 and 306, respectively, and were among the lowest for participating countries.

Estimation of Model Parameters

The MixRM and Mix2PLM were estimated with each of the three identification constraints for establishing a common metric: Item anchoring (Constraint 1) was established by using a single anchor item. Person centering (Constraint 2) was done by setting the mean ability of the first latent group to zero with unit variance. Item centering (Constraint 3) was implemented by setting the mean of item difficulties to zero in each class.

Estimation of model parameters was done using Markov Chain Monte Carlo (MCMC) estimation as implemented in the OpenBUGS computer software (Spiegelhalter et al., 2007). MCMC is the sampling algorithm from probability distributions based on constructing a

Markov chain. Heidelberger and Welch's (1983) convergence diagnostics were used to determine the number of iterations as implemented in the Coda package using R (Plummer, Best, Cowles, Vines, Sarkar, & Almond, 2012).

The following conjugate priors were used in the estimation of the MixRM and Mix2PLM in the empirical example:

$$a_{i,g} \sim Normal(0,1) \text{ and } a_{i,g} > 0, i = 1,\ldots,n$$
$$b_{i,g} \sim Normal(0,1),\ i = 1,\ldots,n$$
$$\theta_{g,j} \sim Normal(\mu_g,1),\ j = 1,\ldots,N$$
$$\mu_g \sim Normal(0,1),\ g = 1,\ldots,G$$
$$(\pi_1,\ldots,\pi_G) \sim Dirichlet(0.5,\ldots,0.5)$$

where a is the discrimination parameter, b is the difficulty parameter, N is the total number of examinees, n is the total number of items, G is the number of latent class group, i is the ith item, j is the jth examinee, and g is the gth latent group.

The coda file for OpenBUGS contains the value of the estimate from each iteration for each parameter. This information was analyzed using the Heidelberger and Welch convergence diagnostics to determine the length of the burn-in and post-burn-in iteration chains. The burn-in arises when early iterations in Markov chain simulation are discarded to diminish the effect of the starting values (Gelman, Carlin, Stern, & Rubin, 2003; Gilks, Richardson, & Spiegelhalter, 1996). After discarding the burn-in period, the number of iterations for Bayesian estimation was determined.

For the MixRMs, a burn-in of 8,000 iterations was found to be sufficient for convergence for all parameters. For the Mix2PLM, 22,000 post-burn-in iterations were used with Constraint 1 (item anchoring with a single anchor item) and Constraint 3 (item centering with the sum of item difficulties set to zero). A burn-in of 2,000 iterations and 24,000 post-burn-in iterations were sufficient for obtaining convergence for Constraint 2 (person centering in which the first group mean and variance were set to zero and one, respectively). For the Mix2PLMs, a burn-in of 8,000 iterations and 21,000 post-burn-in iterations were used for Constraint 1. A burn-in of 9,000 iterations and 16,000 post-burn-in iterations were used for Constraint 2, and a burn-in of 3,000 iterations and 27,000 iterations for Constraint 3 were used.

RESULTS

Detection of Label Switching

When estimating MixIRTMs, it is important to monitor the estimation for possible label switching. This can be observed in real data when latent classes switch during a single MCMC chain. To determine if label switching has occurred, modes of the posterior densities for group membership were monitored. If multiple modes are present, then label switching can be assumed to have occurred. For this example, modes of posterior parameters were examined by using the group membership information in the coda output files. Cross-tabulation analyses were then done for these modes for each constraint for the MixRMs and Mix2PLMs.

Label switching between constraints was assumed when the same latent classes estimated from the two models using different constraints did not agree. The presence of label switching can be seen in Tables 2.1 and 2.2. Label switching was inferred, in other words, when different latent classes for the two models had higher percentages of agreement. When label switchings was observed, labels were switched for reporting purposes based on the highest percent of agreement for group membership.

Tables 2.1 and 2.2 show the agreement in group membership classifications between MixRMs for different identification constraints. The values on the main diagonal are shown in bold and indicate the number of exact agreements in group membership between constraints. In Table 2.1, for example, 959 simulated examinees were placed into Class 1 by both Constraints 1 (item anchoring) and 2 (person centering). This represented 38.5% of the total sample of 2,493 simulated examinees. The percent matching for group membership was 91.2% between Constraints 1 (item anchoring) and 2 (person centering) and 94.3% between Constraints 1 (item anchoring) and 3 (item centering). There was no label switching for the MixRMs with Constraints 1 and 2 but label switching existed on the MixRM with Constraint 3.

Table 2.1: Latent Class Classifications for the MixRM with Constraint 1 (Item Anchoring) and Constraint 2 (Person Centering)

MixRM	MixRM with Constraint 2					
with Constraint 1	Class 1	Class 2	Class 3	Class 4	Class 5	Total
Class 1	**959**	31	0	0	3	993
	(38.5%)	(1.2%)	(0.0%)	(0.0%)	(0.1%)	(39.8%)
Class 2	1	**239**	1	0	0	241
	(0.0%)	**(0.6%)**	(0.0%)	(0.0%)	(0.0%)	(9.7%)
Class 3	132	6	**990**	1	0	1129
	(5.3%)	(0.2%)	**(39.7%)**	(0.0%)	(0.0%)	(45.3%)
Class 4	20	14	0	**48**	3	85
	(0.8%)	(0.6%)	(0.0%)	**(1.9%)**	(0.1%)	(3.4%)
Class 5	6	0	1	0	**38**	45
	(0.2%)	(0.0%)	(0.0%)	(0.0%)	**(1.5%)**	(1.8%)
Total	1118	290	992	49	44	2493
	(44.8%)	(11.6%)	(39.8%)	(2.0%)	(1.8%)	(100.0%)

Tables 2.3 and 2.4 show the numbers and percentages of agreement of group membership classifications for the Mix2PLM with different identification constraints. The percent of agreement for classification of group membership was 80.8% between Constraints 1 and 2 and 89% between Constraints 1 and 3. There was label switching for the Mix2PLMs with Constraints 2 and 3.

Table 2.2: Latent Class Classifications for the MixRM with Constraint 1 (Item Anchoring) and Constraint 3 (Item Centering)

MixRM with Constraint 1	MixRM with Constraint 3						
	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Total
Class 1	0	0	4	**985**	1	3	993
	(0.0%)	(0.0%)	(0.2%)	**(39.5%)**	(0.0%)	(0.1%)	(39.8%)
Class 2	0	0	**231**	10	0	0	241
	(0.0%)	(0.0%)	**(9.3%)**	(0.4%)	(0.0%)	(0.0%)	(9.7%)
Class 3	7	1	3	90	16	**1012**	1129
	(0.3%)	(0.0%)	(0.1%)	(3.6%)	(0.6%)	**(40.6%)**	(45.3%)
Class 4	0	0	0	6	**79**	0	85
	(0.0%)	(0.0%)	(0.0%)	(0.2%)	**(3.2%)**	(0.0%)	(3.4%)
Class 5	0	**45**	0	0	0	0	45
	(0.0%)	**(1.8%)**	(0.0%)	(0.0%)	(0.0%)	(0.0%)	(1.8%)
Total	7	46	238	1091	96	1015	2493
	(0.3%)	(1.8%)	(9.5%)	(43.8%)	(3.9%)	(40.7%)	(100.0%)

Table 2.3: Latent Class Classifications for the Mix2PLM with Constraint 1 (Item Anchoring) and Constraint 2 (Person Centering)

Mix2PLM with Constraint 1	Mix2PLM with Constraint 2			
	Class 1	Class 2	Class 3	Total
Class 1	375	**780**	32	1187
	(15.0%)	**(31.3%)**	(1.3%)	(47.6%)
Class 2	**733**	0	2	735
	(29.4%)	(0.0%)	(0.1%)	(29.5%)
Class 3	68	2	**501**	571
	(2.7%)	(0.1%)	**(20.1%)**	(22.9%)
Total	1176	782	535	2493
	(47.2%)	(31.4%)	(21.5%)	(100.0%)

Table 2.4: Latent Class Classifications for the Mix2PLM with Constraint 1 (Item Anchoring) and Constraint 3 (Item Centering)

Mix2PLM with Constraint 1	Mix2PLM with Constraint 3 Class 1	Class 2	Class 3	Total
Class 1	**969**	31	187	1187
	(38.9%)	(1.2%)	(7.5%)	(47.6%)
Class 2	0	6	**729**	735
	(0.0%)	(0.2%)	**(29.2%)**	(29.5%)
Class 3	2	**521**	48	571
	(0.1%)	**(20.9%)**	(1.9%)	(22.9%)
Total	971	558	964	2493
	(38.9%)	(22.4%)	(38.7%)	(100.0%)

Model Selection

Two information indices are reported for each of the MixIRTM analyses, the Bayesian information criterion (BIC: Schwartz, 1978) and Akaike's Information Criterion (AIC: Akaike, 1973). Values for AIC and BIC are reported in Table 2.5 and smaller AIC and BIC values indicate the better model. Both AIC and BIC suggest different numbers of latent classes depending on the constraint used. Based on AIC, a five latent classes were detected using item anchoring, six latent classes were detected using person centering, and seven latent classes were detected using item centering. Using BIC, five latent classes were detected using item anchoring and person centering and six classes were detected using item centering.

AIC and BIC values in Table 2.6 show the number of latent classes detected for the Mix2PLM. Based on AIC, four latent classes were detected using item anchoring and person centering. Using item centering, a 3-group solution was suggested. Using BIC, a 3-group solution was the best fit to the data for the Mix2PLM for all three constraints.

Table 2.5: Model Comparison Information Criteria for MixRMs

	AIC			BIC		
Latent Classes	Constraint 1 (Item Anchoring)	Constraint 2 (Person Centering)	Constraint 3 (Item Centering)	Constraint 1 (Item Anchoring)	Constraint 2 (Person Centering)	Constraint 3 (Item Centering)
1	67670	67670	67670	67820	67820	67820
2	65890	65870	65870	66190	66170	66180
3	65210	65180	65200	65660	65640	65660
4	64910	64810	64850	65510	65430	65470
5	**64520**	64460	64490	**65280**	**65240**	65270
6	64520	**64420**	64240	65430	65350	**65170**
7			**64190**			65280

Table 2.6: Model Comparison Information Criteria for Mix2PLMs

	AIC			BIC		
Latent Classes	Constraint 1 (Item Anchoring)	Constraint 2 (Person Centering)	Constraint 3 (Item Centering)	Constraint 1 (Item Anchoring)	Constraint 2 (Person Centering)	Constraint 3 (Item Centering)
1	66430	66460	66460	66580	66750	66760
2	65300	65550	65210	65890	66140	65810
3	64660	64550	**64570**	**65540**	**65450**	**65470**
4	**64370**	**64330**	64570	65540	65530	65770

Comparison of Class Means and Latent Group Proportions between Constraints

Comparison of model parameters between latent classes requires that the parameter estimates are expressed on a common scale. Additional equating or scale transformation was not required for comparisons of scale parameters within each constraint as this is what these constraints are designed to do. Comparisons of scale parameters between constraints for the same model, however, did require an additional scale transformation. Mean and sigma equating was used for this set of transformations. The means for each of the five classes using each of the constraints are reported in Table 2.7. The means appear to be different between the three constraints for the MixRMs. In particular, the difference between means for Constraint 2 (person centering) and Constraint 3 (item centering) appears to be large (see Table 2.7).

Table 2.7: Ability Means of Latent Classes for MixRM and Mix2PLM

	Mixture Rasch Model			Mixture 2PL Model		
1	0.38	0.00	0.78	-1.24	-1.43	-1.58
2	0.77	0.59	1.12	1.34	-0.07	0.86
3	-1.18	-1.47	-0.72	0.90	0.92	0.79
4	0.15	-0.50	0.46			
5	1.45	0.90	1.46			
6			-0.24			

The latent class means for Constraints 1 (item anchoring) and 3 (item centering) for the Mix2PL model do not appear to be similar. In addition, the means for the second latent class using Constraint 2 (person centering) also look different from those for the other two constraints. The first and third latent class means for Constraints 2 (person centering) were similar to those for Constraints 1 and 3 (see Table 2.7).

The proportions of examinees classified into Classes 1 to 6 by the MixRM using each of the constraints are reported in Table 2.8. The proportions for the different constraints in MixRM look somewhat similar although the proportions for the first and third classes differ for each of the three constraints.

Table 2.8: Proportions of Latent Classes for MixRM and Mix2PLM

	Mixture Rasch Model			Mixture 2PL Model		
Latent Classes	Constraint 1 (Item Anchoring)	Constraint 2 (Person Centering)	Constraint 3 (Item Centering)	Constraint 1 (Item Anchoring)	Constraint 2 (Person Centering)	Constraint 3 (Item Centering)
1	39.8	44.8	43.8	47.6	31.4	38.9
2	9.7	11.6	9.5	29.5	47.2	38.7
3	45.3	39.8	40.7	22.9	21.5	22.4
4	3.4	2.0	3.9			
5	1.8	1.8	1.8			
6			0.3			

The proportions of examinees classified into each of the three latent classes detected with the Mix2PLM also are reported in Table 2.8. The Mix2PLMs had different proportions of class membership for the different constraints. The proportions in Class 3 look similar for the three constraints, however, the proportions of Classes 1 and 2 look different for the different constraints. These results suggest that students were assigned to different latent classes by the different constraints.

Comparison of Joint Classifications

Group memberships were compared to examine the effects of the constraints on classification of examinees into latent classes after solving the label switching problem. Joint classifications for Constraint 1 (item anchoring) and Constraint 2 (person centering) for the MixRM are given in Table 2.9. Agreement between constraints is calculated by taking the sum of the numbers on the main diagonal. Results in this table indicate 91.2% (n = 2,274) agreement. Agreement between Constraints 1 (item anchoring) and 3 (item centering) was 94.3% (n = 2,352)(see Table 2.10). Agreement between Constraint 2 and Constraint 3 was 92.7% (n = 2,312)(see Table 2.11).

For the Mix2PLM, there was 80.8% agreement in classification (n = 2014) between Con-

Table 2.9: Group Membership Classification for MixRMs Using Constraints 1 and 2

MixRM with Constraint 1	MixRM with Constraint 2					
	Class 1	Class 2	Class 3	Class 4	Class 5	Total
Class 1	**959**	31	0	0	3	993
	(38.5%)	(1.2%)	(0.0%)	(0.0%)	(0.1%)	(39.8%)
Class 2	1	**239**	1	0	0	241
	(0.0%)	**(9.6%)**	(0.0%)	(0.0%)	(0.0%)	(9.7%)
Class 3	132	6	**990**	1	0	1129
	(5.3%)	(0.2%)	**(39.7%)**	(0.0%)	(0.0%)	(45.3%)
Class 4	20	14	0	**48**	3	85
	(0.8%)	(0.6%)	(0.0%)	**(1.9%)**	(0.1%)	(3.4%)
Class 5	6	0	1	0	**38**	45
	(0.2%)	(0.0%)	(0.0%)	(0.0%)	**(1.5%)**	(1.8%)
Total	1118	290	992	49	44	2493
	(44.8%)	(11.6%)	(39.8%)	(2.0%)	(1.8%)	(100.0%)

Table 2.10: Group Membership Classification for MixRMs Using Constraints 1 and 3

MixRM with Constraint 1	MixRM with Constraint 3						
	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Total
Class 1	**985**	4	3	1	0	0	993
	(39.5%)	(0.2%)	(0.1%)	(0.0%)	(0.0%)	(0.0%)	(39.8%)
Class 2	10	**231**	0	0	0	0	241
	(0.4%)	**(9.3%)**	(0.0%)	(0.0%)	(0.0%)	(0.0%)	(9.7%)
Class 3	90	3	**1012**	16	1	7	1129
	(3.6%)	(0.1%)	**(40.6%)**	(0.6%)	(0.0%)	(0.3%)	(45.3%)
Class 4	6	0	0	**79**	0	0	85
	(0.2%)	(0.0%)	(0.0%)	**(3.2%)**	(0.0%)	(0.0%)	(3.4%)
Class 5	0	0	0	0	**45**	0	45
	(0.0%)	(0.0%)	(0.0%)	(0.0%)	**(1.8%)**	(0.0%)	(1.8%)
Total	1091	238	1015	96	46	7	2493
	(43.8%)	(9.5%)	(40.7%)	(3.9%)	(1.8%)	(0.3%)	(100.0%)

Table 2.11: Group Membership Classification for MixRMs Using Constraints 2 and 3

MixRM with Constraint 2	MixRM with Constraint 3 Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Total
Class 1	**1033**	0	60	19	6	0	1118
	(41.4%)	(0.0%)	(2.4%)	(0.8%)	(0.2%)	(0.0%)	(44.8%)
Class 2	39	**237**	0	14	0	0	290
	(1.6%)	**(9.5%)**	(0.0%)	(0.6%)	(0.0%)	(0.0%)	(11.6%)
Class 3	16	1	**955**	11	2	7	992
	(0.6%)	(0.0%)	**(38.3%)**	(0.4%)	(0.1%)	(0.3%)	(39.8%)
Class 4	0	0	0	**49**	0	0	49
	(0.0%)	(0.0%)	(0.0%)	**(2.0%)**	(0.0%)	(0.0%)	(2.0%)
Class 5	3	0	0	3	**38**	0	44
	(0.1%)	(0.0%)	(0.0%)	(0.1%)	**(1.5%)**	(0.0%)	(1.8%)
Total	1091	238	1015	96	46	**7**	2493
	(43.8%)	(9.5%)	(40.7%)	(3.9%)	(1.8%)	**(0.3%)**	(100.0%)

straint 1 and Constraint 2 (see Table 2.12). Agreement between Constraint 1 and Constraint 3 agreement for the Mix2PLM was 89% (n = 2,219), and 90.9% agreement (n = 2,267) between Constraint 2 and Constraint 3 in the Mix2PLM (see Tables 2.13 and 2.14).

Table 2.12: Latent Group Classification for the Mix2PL Model Using Constraints 1 and 2

Mix2PLM with Constraint 1	Mix2PLM with Constraint 2 Class 1	Class 2	Class 3	Total
Class 1	**780**	375	32	1187
	(31.3%)	(15.0%)	(1.3%)	(47.6%)
Class 2	0	**733**	2	735
	(0.0%)	**(29.4%)**	(0.1%)	(29.5%)
Class 3	2	68	**501**	571
	(0.1%)	(2.7%)	**(20.1%)**	(22.9%)
Total	782	1176	535	2493
	(31.4%)	(47.2%)	(21.5%)	(100.0%)

All comparisons, except for those between the Constraints 1 and 3 for the MixRM, indicated that the different constraints resulted in some lack of agreement in latent group classifications. The group membership agreement between constraints used with the MixRM

Table 2.13: Latent Group Classification for the Mix2PL Model Using Constraints 1 and 3

Mix2PLM with Constraint 1	Mix2PLM with Constraint 3 Class 1	Class 2	Class 3	Total
Class 1	**969**	187	31	1187
	(38.9%)	(7.5%)	(1.2%)	(47.6%)
Class 2	0	**729**	6	735
	(0.0%)	**(29.2%)**	(0.2%)	(29.5%)
Class 3	2	48	**521**	571
	(0.1%)	(1.9%)	**(20.9%)**	(22.9%)
Total	971	964	558	2493
	(38.9%)	(38.7%)	(22.4%)	(100.0%)

Table 2.14: Latent Group Classification for the Mix2PL Model Using Constraints 2 and 3

Mix2PLM with Constraint 2	Mix2PLM with Constraint 3 Class 1	Class 2	Class 3	Total
Class 1	**782**	0	0	782
	(31.4%)	(0.0%)	(0.0%)	(31.4%)
Class 2	177	**963**	36	1176
	(7.1%)	**(38.6%)**	(1.4%)	(47.2%)
Class 3	12	1	**522**	535
	(0.5%)	(0.0%)	**(20.9%)**	(21.5%)
Total	971	964	558	2493
	(38.9%)	(38.7%)	(22.4%)	(100.0%)

was higher than that when used with the Mix2PLM.

Comparisons of Item Parameter Estimates

As noted above, mean and sigma equating was used to obtain a common metric between latent classes and across the different constraints for each MixIRTM. This permitted direct comparisons such as plotting difficulty estimates on the same graph in order to observe differences between constraints. Item difficulty parameter estimates for the different constraints for the MixRM are shown for classes 1, 2, and 4 in Tables 2.15, 2.16, and 2.17.

Table 2.15: Difficulty Parameters For the MixRM Under Constraint 1

Item No.	TIMSS Item ID	Class 1	Class 2	Class 3	Class 4	Class 5
1	S032611	0.89	0.89	0.89	0.89	0.89
2	S032614	-0.14	-0.31	1.04	-0.09	0.69
3	S032156	-0.13	-0.50	1.18	0.83	-0.16
4	S032056	0.44	-0.18	2.67	0.39	2.23
5	S032087	0.50	0.80	1.30	1.25	-0.11
6	S032279	1.98	1.36	1.10	1.37	0.73
7	S032238	0.37	-1.56	1.53	-0.78	-0.54
8	S032160	-0.58	-0.10	0.01	0.06	-0.72
9	S032654	0.43	0.04	1.10	0.19	1.22
10	S032126	0.21	0.18	1.19	-0.72	1.64
11	S032510	-1.18	-1.47	0.36	-0.02	-0.07
12	S032158	-0.83	-0.04	0.14	-0.08	-0.93
13	S052093	-1.77	-1.35	-1.14	-0.75	-1.04
14	S052088	-1.13	-1.26	0.31	0.17	2.64
15	S052030	0.58	1.05	0.61	0.33	1.03
16	S052080	-0.39	-0.65	0.13	2.60	0.05
17	S052091	0.19	-0.85	1.39	-0.26	0.70
18	S052152	0.50	-0.06	1.57	1.28	-0.72
19	S052136	-0.07	-0.18	1.49	-0.05	0.61
20	S052046	-2.89	-2.60	0.05	-0.94	-1.24
21	S052254	0.30	2.21	1.07	1.45	-0.24
22	S052207	0.79	0.25	2.34	0.26	-0.48
23	S052297	-1.04	1.65	0.62	0.07	0.03
24	S052032	1.67	1.69	3.02	2.61	2.74
25	S052106	1.19	0.26	2.18	0.24	1.04

As can be seen in Figure 2.1, plots of difficulty parameter estimates in Class 1 were

Table 2.16: Difficulty Parameters For the MixRM Under Constraint 2

Item No.	TIMSS Item ID	Class 1	Class 2	Class 3	Class 4	Class 5
1	S032611	0.92	1.78	0.10	1.00	0.33
2	S032614	-0.15	0.29	0.66	0.25	0.68
3	S032156	-0.10	0.14	0.77	0.87	0.04
4	S032056	0.50	0.42	2.36	1.12	1.96
5	S032087	0.47	1.41	0.94	1.15	0.33
6	S032279	1.82	2.03	0.59	1.32	1.31
7	S032238	0.41	-0.99	1.00	0.13	-0.32
8	S032160	-0.61	0.46	-0.43	0.02	-0.28
9	S032654	0.38	0.65	0.67	0.41	1.21
10	S032126	0.14	0.76	0.84	-0.13	1.36
11	S032510	-1.10	-0.88	-0.02	0.01	0.02
12	S032158	-0.86	0.51	-0.35	0.13	-0.66
13	S052093	-1.80	-0.86	-1.52	-0.96	-0.69
14	S052088	-1.03	-0.66	-0.05	0.09	1.96
15	S052030	0.53	1.61	0.08	0.36	0.97
16	S052080	-0.42	0.13	-0.28	1.71	0.76
17	S052091	0.21	-0.25	0.94	0.22	0.60
18	S052152	0.53	0.58	1.04	1.58	-0.21
19	S052136	-0.06	0.41	1.15	0.35	0.66
20	S052046	-2.64	-2.10	-0.34	-0.58	-1.02
21	S052254	0.25	2.65	0.60	1.46	0.23
22	S052207	0.81	0.84	1.93	0.91	0.06
23	S052297	-0.97	2.02	0.21	0.08	0.16
24	S052032	1.67	2.29	2.68	2.61	2.85
25	S052106	1.19	0.84	1.69	0.98	1.18

Table 2.17: Difficulty Parameters For the MixRM Under Constraint 3

Item No.	TIMSS Item ID	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
1	S032611	0.93	1.27	0.49	0.23	-0.23	-0.56
2	S032614	-0.13	-0.30	-0.09	-0.49	0.28	0.23
3	S032156	-0.11	-0.49	0.04	0.42	-0.52	0.06
4	S032056	0.43	-0.12	1.55	0.08	1.64	1.08
5	S032087	0.51	0.84	0.03	0.83	-0.35	0.30
6	S032279	1.96	1.40	0.06	0.96	0.42	-0.05
7	S032238	0.34	-1.53	0.81	-1.12	-0.84	-0.18
8	S032160	-0.57	-0.10	-1.12	-0.40	-0.87	-0.41
9	S032654	0.43	0.04	0.04	-0.26	0.74	0.16
10	S032126	0.21	0.20	0.04	-1.11	1.08	0.36
11	S032510	-1.15	-1.50	-0.81	-0.45	-0.47	-0.47
12	S032158	-0.85	-0.01	-0.74	-0.48	-1.21	-0.65
13	S052093	-1.77	-1.40	-2.54	-1.14	-1.27	-1.08
14	S052088	-1.10	-1.30	-0.96	-0.25	1.78	-0.05
15	S052030	0.57	1.10	-0.32	-0.12	0.48	-0.21
16	S052080	-0.36	-0.67	-1.13	2.12	-0.18	-0.41
17	S052091	0.19	-0.87	0.44	-0.67	0.17	0.10
18	S052152	0.48	-0.01	0.81	0.92	-0.91	0.04
19	S052136	-0.04	-0.15	0.27	-0.45	0.20	0.32
20	S052046	-2.95	-2.64	-0.96	-1.34	-1.48	-0.96
21	S052254	0.30	2.28	0.09	1.13	-0.46	0.04
22	S052207	0.80	0.29	1.30	-0.13	-0.56	0.63
23	S052297	-0.99	1.71	-0.46	-0.39	-0.31	-0.36
24	S052032	1.68	1.70	1.85	2.24	2.22	1.44
25	S052106	1.19	0.26	1.31	-0.13	0.63	0.62

essentially on top of one another. The plots of difficulty parameter estimates for Class 2 were a little different between the three constraints. In class 2, Constraint 2 (person centering) had difficulty parameter estimates which were higher than Constraints 1 (item anchoring) and 3 (item centering) (see Figure 2.2). The results from Classes 3 to 5 indicated that difficulty parameter estimates clearly differed for all three constraints (see Figures 2.3 to 2.5). Correlations are reported in Table 2.18) showing the degree of the relationships between estimates for the different constraints.

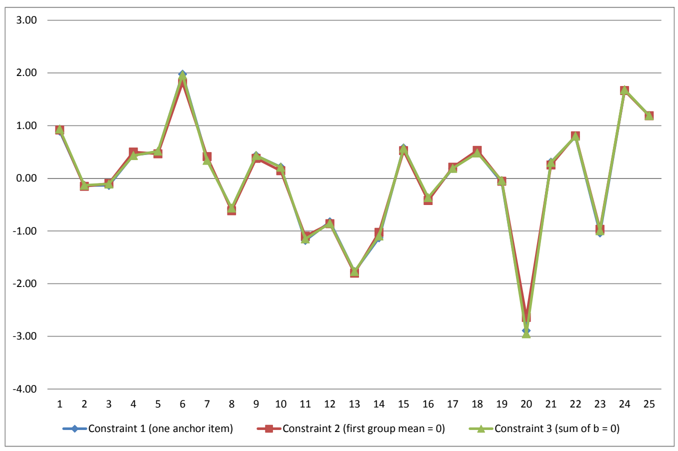


Figure 2.1: Difficulty Estimates Comparison in the MixRM for Class 1

Correlations Between Parameter Estimates. Correlations between parameter estimates for the same latent class should be high if constraints had no impact. To examine this conjecture, correlations were calculated between estimates for each MixIRTM with each of the constraints. The correlations for the MixRM are presented in Table 2.18. Correlations for the Mix2PLM are presented in Table 2.25. The following notation is used in these tables: C1&A indicates a parameter estimate in Class 1 under Constraint 1 (item anchoring); C2&P indicates Class 2 for Constraint 2 (person centering); C2&I indicates class 2 for Constraint 3 (item centering), etc. In this notation, constraints are indicated as Constraint 1 = A,

Figure 2.2: Difficulty Estimates Comparison in the MixRM for Class 2

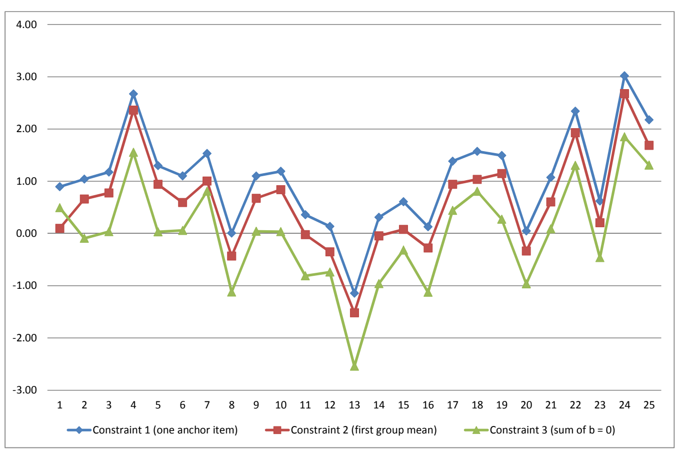


Figure 2.3: Difficulty Estimates Comparison in the MixRM for Class 3

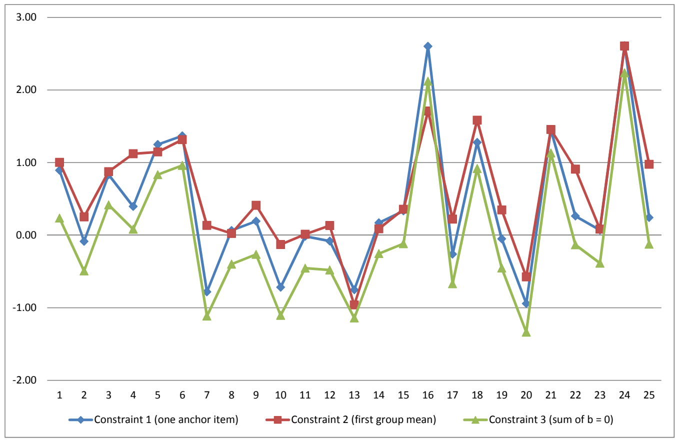


Figure 2.4: Difficulty Estimates Comparison in the MixRM for Class 4

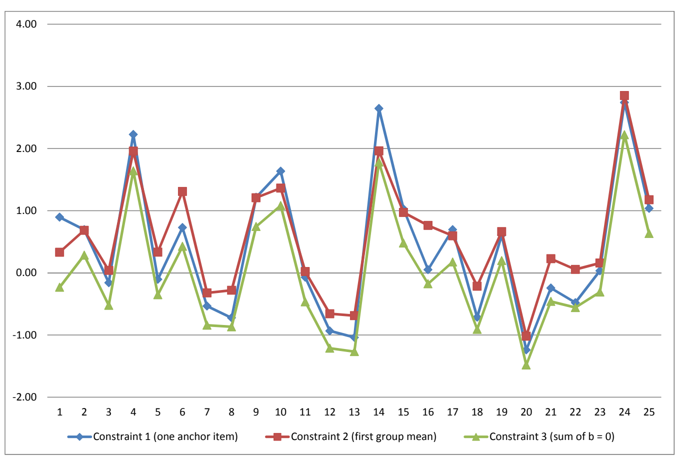


Figure 2.5: Difficulty Estimates Comparison in the MixRM for Class 5

Constraint 2 = P, and Constraint 3 = I.

Correlations in Table 2.18 indicate that difficulty parameter estimates were similar between the three constraints within the same latent class. Most of these correlations, in fact, were .99. The correlations between Constraints 1 and 2 ($r = .918$) and between Constraint 2 and 3 ($r = .928$) in Class 4, however, were slightly smaller, suggesting that there might be some effect of constraints in Class 4. The sample size for Class 4 was very small ($n = 85$ (3.4%) for Constraint 1, $n = 49$ (2.0%) for Constraint 2, and $n = 96$ (3.9%) for Constraint 3). The slightly smaller correlations may be due to the instability resulting from these small sample sizes (see Table 2.8).

Discrimination parameter estimates under each constraint for the Mix2PLM are given in Tables 2.19, 2.20, and 2.21 and difficulty parameter estimates under each constraint for this model are given in Tables 2.22, 2.23, and 2.24. Correlations were computed between parameter estimates to help detect constraint effects on item parameter estimation. Correlations between discrimination parameters are given in Table 2.25.

One would expect correlations between the estimates for the same parameters within the same latent class but between different constraints to be close to 1 if there were no differences due to constraints. Although all within class correlations were high and significant ($p < .01$), they were not all close to 1. Correlations in Table 2.25 indicated there were some differences in discrimination parameter estimates in class 2 under Constraints 1 and 2. The correlations between discrimination parameters for all three constraints for class 3, however, were close to unity, suggesting there was very little effect between constraints in this latent class.

Correlations for these analysis was reported to get more statistically reasonable comparison (see Table 2.26). Correlations in Table 2.26 indicated that most difficulty parameter estimates differed between constraints within latent class. Exceptions were correlations between Constraints 1 and 3 in class 1 ($r = .980$) and Constraints 1 and 2 in class 3 ($r = .997$).

Table 2.18: Correlations for Difficulty Parameters in the MixRM

	C1&P	C1&I	C2&A	C2&P	C2&I	C3&A	C3&P	C3&I	C4&A	C4&P	C4&I	C5&A	C5&P	C5&I
C1&A	.998**	1.000**	.665**	.700**	.669**	.731**	.688**	.752**	.531**	.728**	.537**	.462*	.574**	.490*
C1&P		.998**	.651**	.686**	.656**	.759**	.715**	.782**	.529**	.737**	.534**	.472*	.579**	.498*
C1&I			.670**	.705**	.674**	.729**	.686**	.748**	.536**	.729**	.541**	.468*	.579**	.495*
C2&A				.996**	.998**	.385	.354	.393	.565**	.597**	.563**	.290	.393	.317
C2&P					.997**	.398*	.363	.411*	.602**	.633**	.596**	.311	.413*	.331
C2&I						.385	.349	.402*	.561**	.597**	.556**	.287	.380	.305
C3&A							.994**	.978**	.329	.656**	.356	.502*	.571**	.550**
C3&P								.952**	.314	.633**	.345	.514**	.590**	.573**
C3&I									.300	.645**	.321	.444*	.489*	.469*
C4&A										.918**	.997**	.257	.434*	.297
C4&P											.928**	.348	.515**	.394
C4&I												.256	.442*	.305
C5&A													.960**	.986**
C5&P														.986**

Note. * $p < .05$ (2-tailed), ** $p < .01$ (2-tailed)

Table 2.19: Discrimination Parameters For the Mix2PLM Under Constraint 1

Item No.	TIMSS Item ID	Class 1	Class 2	Class 3
1	S032611	0.46	0.47	0.42
2	S032614	1.10	0.66	1.29
3	S032156	0.65	0.91	1.11
4	S032056	1.22	0.50	1.19
5	S032087	0.56	0.94	0.88
6	S032279	0.52	0.78	0.44
7	S032238	0.47	1.29	1.76
8	S032160	0.45	0.81	0.72
9	S032654	0.54	0.61	1.18
10	S032126	1.22	0.53	0.84
11	S032510	0.68	0.95	1.11
12	S032158	0.45	1.15	0.48
13	S052093	0.76	1.09	0.82
14	S052088	1.03	0.48	0.83
15	S052030	0.38	1.04	0.46
16	S052080	0.26	0.52	0.64
17	S052091	1.09	1.11	0.76
18	S052152	0.46	1.12	1.39
19	S052136	1.21	0.56	1.38
20	S052046	0.86	2.02	1.47
21	S052254	0.40	1.02	0.51
22	S052207	1.73	1.26	1.30
23	S052297	0.92	1.31	0.21
24	S052032	1.24	0.61	1.12
25	S052106	1.20	1.15	0.49

Table 2.20: Discrimination Parameters For the Mix2PLM Under Constraint 2

Item No.	TIMSS Item ID	Class 1	Class 2	Class 3
1	S032611	0.25	1.00	0.57
2	S032614	1.02	0.83	1.30
3	S032156	0.40	1.22	1.07
4	S032056	1.17	0.96	1.26
5	S032087	0.51	1.10	0.93
6	S032279	0.45	0.51	0.47
7	S032238	0.41	1.39	1.77
8	S032160	0.15	0.89	0.69
9	S032654	0.51	0.64	1.36
10	S032126	1.11	0.66	0.84
11	S032510	0.48	1.07	1.14
12	S032158	0.20	1.27	0.50
13	S052093	0.58	0.83	0.82
14	S052088	0.81	0.50	0.82
15	S052030	0.27	0.99	0.47
16	S052080	0.15	0.59	0.67
17	S052091	0.80	1.14	0.77
18	S052152	0.37	1.39	1.43
19	S052136	1.02	0.81	1.35
20	S052046	0.34	2.12	1.42
21	S052254	0.31	1.19	0.59
22	S052207	1.54	1.26	1.32
23	S052297	0.53	1.44	0.20
24	S052032	1.44	0.72	1.16
25	S052106	1.05	1.16	0.47

Table 2.21: Discrimination Parameters For the Mix2PLM Under Constraint 3

Item No.	TIMSS Item ID	Class 1	Class 2	Class 3
1	S032611	0.29	1.06	0.61
2	S032614	1.07	0.78	1.39
3	S032156	0.49	1.09	1.18
4	S032056	1.23	0.78	1.36
5	S032087	0.53	1.12	0.97
6	S032279	0.50	0.64	0.48
7	S032238	0.45	1.46	1.96
8	S032160	0.26	0.83	0.74
9	S032654	0.55	0.63	1.45
10	S032126	1.21	0.60	0.93
11	S032510	0.61	1.04	1.21
12	S032158	0.29	1.28	0.51
13	S052093	0.65	0.90	0.86
14	S052088	0.87	0.48	0.88
15	S052030	0.33	1.07	0.50
16	S052080	0.18	0.55	0.70
17	S052091	0.98	1.16	0.83
18	S052152	0.42	1.35	1.53
19	S052136	1.15	0.69	1.51
20	S052046	0.59	2.10	1.52
21	S052254	0.36	1.14	0.61
22	S052207	1.70	1.30	1.44
23	S052297	0.74	1.35	0.22
24	S052032	1.35	0.69	1.23
25	S052106	1.16	1.22	0.51

Table 2.22: Difficulty Parameters For the Mix2PLM Under Constraint 1

Item No.	TIMSS Item ID	Class 1	Class 2	Class 3
1	S032611	0.52	0.60	0.83
2	S032614	-0.48	-0.34	0.03
3	S032156	0.32	-0.09	-0.09
4	S032056	1.13	0.31	0.13
5	S032087	0.86	0.20	1.00
6	S032279	0.89	2.04	2.03
7	S032238	1.45	0.84	-0.51
8	S032160	-1.40	-0.97	-0.58
9	S032654	0.42	0.04	0.48
10	S032126	-0.38	0.01	-0.02
11	S032510	-1.00	-1.15	-0.91
12	S032158	-0.93	-0.92	-1.46
13	S052093	-2.97	-1.35	-2.04
14	S052088	-1.13	-1.87	-1.33
15	S052030	0.33	0.55	0.58
16	S052080	-0.62	-1.47	-0.53
17	S052091	-0.07	0.73	-0.99
18	S052152	1.71	0.59	0.51
19	S052136	-0.20	-0.29	0.12
20	S052046	-1.39	-1.71	-1.69
21	S052254	1.20	-0.05	2.28
22	S052207	0.26	1.10	0.60
23	S052297	-0.80	-0.66	-0.14
24	S052032	1.44	2.35	1.96
25	S052106	0.76	1.60	-0.29

Table 2.23: Difficulty Parameters For the Mix2PLM Under Constraint 2

Item No.	TIMSS Item ID	Class 1	Class 2	Class 3
1	S032611	0.72	0.51	1.08
2	S032614	-0.21	-0.77	0.01
3	S032156	0.06	0.84	-0.09
4	S032056	0.86	1.10	0.19
5	S032087	0.30	0.59	1.06
6	S032279	1.95	0.62	1.92
7	S032238	0.68	0.97	-0.51
8	S032160	-0.93	-0.83	-0.69
9	S032654	0.08	0.27	0.52
10	S032126	0.03	-0.62	-0.03
11	S032510	-0.94	-0.99	-0.85
12	S032158	-0.75	-0.48	-1.27
13	S052093	-2.45	-3.68	-2.05
14	S052088	-2.13	-1.11	-1.42
15	S052030	0.34	0.21	0.53
16	S052080	-1.27	0.16	-0.40
17	S052091	0.44	-0.11	-1.02
18	S052152	0.58	1.54	0.55
19	S052136	0.02	-0.23	0.08
20	S052046	-1.19	-0.87	-1.73
21	S052254	0.15	1.19	2.24
22	S052207	0.84	0.21	0.56
23	S052297	-0.56	-0.63	-0.15
24	S052032	2.13	1.62	1.90
25	S052106	1.31	0.55	-0.37

Table 2.24: Difficulty Parameters For the Mix2PLM Under Constraint 3

Item No.	TIMSS Item ID	Class 1	Class 2	Class 3
1	S032611	0.79	1.03	0.78
2	S032614	-0.62	0.01	-0.22
3	S032156	0.74	-0.07	0.00
4	S032056	1.11	0.17	0.72
5	S032087	0.78	0.98	0.29
6	S032279	0.72	1.82	1.86
7	S032238	1.16	-0.43	0.75
8	S032160	-1.04	-0.65	-1.04
9	S032654	0.26	0.48	0.07
10	S032126	-0.53	0.01	-0.03
11	S032510	-1.06	-0.81	-0.96
12	S032158	-0.63	-1.29	-0.76
13	S052093	-3.34	-1.97	-2.07
14	S052088	-1.11	-1.34	-2.07
15	S052030	0.23	0.52	0.45
16	S052080	-0.06	-0.44	-1.41
17	S052091	-0.15	-0.92	0.55
18	S052152	1.64	0.49	0.61
19	S052136	-0.26	0.12	-0.14
20	S052046	-1.26	-1.62	-1.32
21	S052254	1.19	2.08	0.12
22	S052207	0.15	0.54	0.92
23	S052297	-0.78	-0.15	-0.59
24	S052032	1.47	1.76	2.15
25	S052106	0.60	-0.29	1.36

Table 2.25: Correlations between Discrimination Parameters in the Mix2PLM

	C1&P	C1&I	C2&A	C2&P	C2&I	C3&A	C3&P	C3&I
C1&A	.939**	.980**	-.024	-.013	-.058	.278	.243	.257
C1&P		.985**	-.216	-.209	-.241	.311	.299	.307
C1&I			-.107	-.105	-.143	.308	.286	.297
C2&A				.880**	.921**	.139	.086	.084
C2&P					.980**	.240	.204	.206
C2&I						.190	.158	.158
C3&A							.991**	.993**
C3&P								.999**

Note. ** $p < .01$ (2-tailed)

Table 2.26: Correlations between Difficulty Parameters in Mix2PLM

	C1&M	C1&S	C2&A	C2&M	C2&S	C3&A	C3&M	C3&S
C1&A	.858**	.980**	.778**	.952**	.773**	.760**	.767**	.846**
C1&M		.822**	.957**	.800**	.764**	.757**	.753**	.995**
C1&S			.714**	.989**	.763**	.746**	.759**	.800**
C2&A				.670**	.703**	.702**	.691**	.978**
C2&M					.732**	.719**	.729**	.768**
C2&S						.998**	.999**	.750**
C3&A							.997**	.743**
C3&M								.739**

Note. ** $p < .01$ (2-tailed)

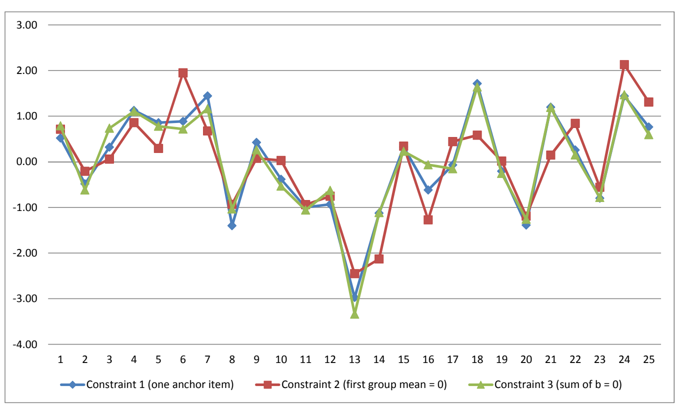


Figure 2.6: Difficulty Estimates Comparison in the Mix2PLM for Class 1

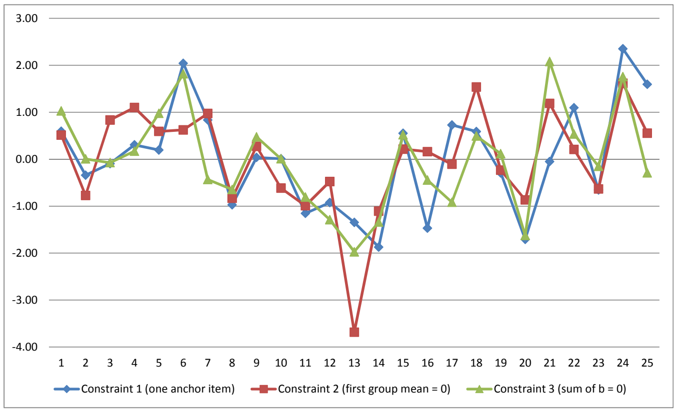


Figure 2.7: Difficulty Estimates Comparison in the Mix2PLM for Class 2

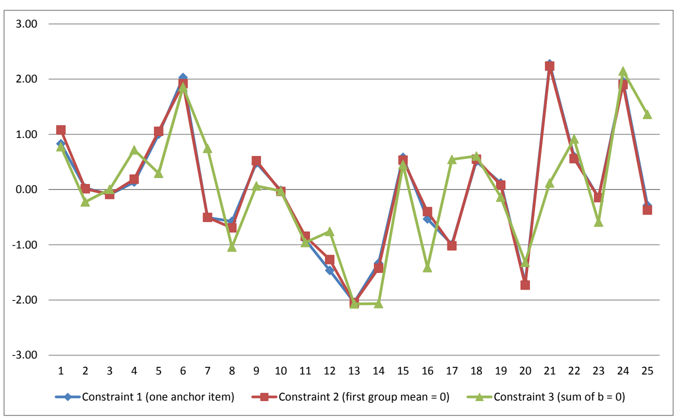


Figure 2.8: Difficulty Estimates Comparison in the Mix2PLM for Class 3

Comparison plots were done to examine further the possible effects of the constraints on item of difficulty parameter estimates. Plots between Constraints 1 and 3 suggests the parameter estimates were similar in Class 1. However, for Constraint 2, the difficulty parameter estimates appear to be different from those obtained with Constraints 1 and 3 in class 1 (see Figure 2.6). Nineteen items (Items 1, 2, 3, 4, 5, 6, 7, 10, 12, 13, 14, 16, 17, 18, 20, 21, 22, 24, and 25) appeared to have different difficulty parameter estimates across the three constraints in Class 2 (see Figure 2.7). The difficulty parameters between Constraints 1 and 2 were similar in Class 3, but those between Constraints 1 and 3 and between Constraints 2 and 3 in class 3 were somewhat different (see Figure 2.8). These plots suggest that for the Mix2PLM, the three constraints provided somewhat different estimates of item difficulty.

Conclusions

The three constraints were applied to establish a common metric between latent classes for a MixRM and a Mix2PLM. The data set was taken from the TIMSS 2011 8th Grade Science Test. Results from the MixRM suggested that each of the constraints had a somewhat different effect on item difficulty estimates, ability estimates, numbers of latent classes, classifications of examinees into latent classes, and proportions of membership in each latent class. Similar results were observed for the Mix2PLM with the exception that the same number of latent classes was extracted using all three constraints.

Results based on this data set clearly differed depending on the MixIRT model and on the constraint used. A major purpose of this dissertation, therefore, was to examine this issue in greater depth with an eye to better understanding the impact of these kinds of constraints. In addition, it is also possible that results may differ depending on conditions in the data. To examine this latter point, we also considered the effects of different test lengths and sample sizes (described in the next section) on developing a common metric for MixIRT models.

Chapter 3

METHODS AND RESEARCH DESIGN

3.1 Estimation

Bayesian Estimation for Mixture IRT Models Using a Markov Chain Monte Carlo Algorithm

MCMC estimation algorithms have been used widely in IRT and have been found to make it possible to estimate more complex types of item response models (Kim, 2001; Patz & Junker, 1999a, 1999b). MCMC methods estimate the full conditional posterior distribution of each parameter being estimated. Markov chain estimates the posterior using a sample from the parameter's posterior distribution at the stages. The sample mean for a parameter over the post-burn-in iterations of the MCMC chain can be taken as the parameter estimate. When mixture distributions are estimated, a class membership parameter for each observation (i.e., examinee) at each stage of the chain is sampled. A class membership is sampled for each examinee along with a continuous ability, θ_{jg}, at each stage of the Markov chain. The class membership parameter is proportional to the probability of the examinee's membership for a given examinee (θ_{jg}). The item parameters in each class are decided based on the frequency with which each examinee is sampled into each class. The posterior probability of the examinee's membership in each class is also determined by the frequency of sampling each examinee into each class (Bolt et al., 2002).

Prior Distributions

To estimate the MixIRTMs using MCMC, priors need to be specified for all item and ability parameters to be estimated. The following conjugate priors were used in this example as well as for the simulation study that follows:

$$
\begin{aligned}
&a_{ig} \sim Normal(0,1) \text{ and } a_{ig} > 0, i = 1, \ldots, n \\
&b_{ig} \sim Normal(0,1),\ i = 1, \ldots, n \\
&c_{ig} \sim Beta(5,17),\ i = 1, \ldots, n \\
&\theta_{jg} \sim Normal(\mu_g, 1),\ j = 1, \ldots, N \\
&\mu_g \sim Normal(0,1),\ g = 1, \ldots, G \\
&(\pi_1, \ldots, \pi_G) \sim Dirichlet(0.5, \ldots, 0.5)
\end{aligned}
$$

where a is the discrimination parameter, b is the difficulty parameter, c is the lower asymptote parameter, N is the total number of examinees, n is the total number of items, G is the number of latent class group, i is the ith item, j is the jth examinee, and g is the gth latent group.

There is some agreement that item parameters follows log-normal distribution for difficulty, normal distribution for discrimination, beta distribution for lower asymptote in IRT model (du Toit, 2003). Based on this agreement, IRT computer programs such as BILOG and MULTILOG use default or user assigned priors for Bayesian estimates (Zimowski et al., 2003; Thissen et al., 2003). Although the calculation of posterior distribution is difficult, the calculation of posterior can be easier when prior distribution and posterior distribution are under the same distribution family. It refers to as conjugate prior (Gelman et al., 2003).

The Bayesian defaults for prior from BILOG-MG computer program were used for the difficulty and lower asymptote parameters (du Toit, 2003). The prior default for the discrimination parameter follows the log-normal distribution with mean $= 0$ and $SD = 0.5$ in the BILOG-MG. The prior for the discrimination parameter follows the normal distribution with mean $= 0$ and $SD = 1$ and should be larger than zero in this study. This prior distribution was also used by Li, Cohen, Kim, and Cho (2009).

Label Switching

Label switching is a concern because it can lead to difficulties in interpretation. This kind of problem arises in MCMC estimation when the model components are ordered arbitrarily in mixture models (Sperrin, Jaki, & Wit, 2010). The term label switching was proposed by Render and Walker (1984) to describe the invariance of the likelihood estimates when the mixture components were relabeled (Stephens, 2000). Label switching must be addressed before convergence diagnostics since it is a prerequisite of convergence of an MCMC sampler (Jasra, Holmes, & Stephens, 2005).

There are two types of label switching. The first one arises when the latent classes are reordered multiple times over the course of an MCMC chain during a run of an MCMC sampler (Sperrin, Jaki, & Wit, 2010). There are some methods to handle this type of label switching. Three commonly used methods are imposing artificial identifiability constraints on the model parameters, implementing a relabelling algorithms (Stephens, 1997), and imposing invariant loss functions (Celeux, Hurn, & Robert, 2000).

The second type of label switching arises when the latent class switches among the replications in a simulation study (Choi et al., 2010; Li et al., 2009). This type can cause the difficulty in interpretation of results since the latent class will have taken different orders in each replication. In a simulation study, since the generating values for item parameters for each latent class and the group membership parameters are already known, estimated parameters can be compared with the generating values. If label switching is observed, then relabelling can be done to re-align the latent classes based on matching estimates with the generating parameters.

Two ways can be used to handle label switching. The first method uses group membership parameters and the second used item parameters and the second method uses group membership parameters to check for label switching. With respect to the first method, a cross-tabulation can be done to compare the generating group membership information to group membership estimates. For the second method, comparing item parameters, the generating parameter for items in each latent class can be compared to the parameter estimates. The pattern of matches can then be compared and the latent classes relabeled to the closest pattern.

Monitoring Convergence

The initial iterations in a Markov chain are referred to as burn-in iterations and are assumed to reflect some effect of the starting values. These are discarded in order to diminish the effect of the starting values (Gelman et al., 2003; Gilks et al., 1996). Iterations after the burn-in are used to obtain estimates of parameters. Convergence diagnostics are used to determine how many iterations to retain following burn-in. This requires examining each iteration in a chain for each parameter.

Commonly used convergence diagnostics methods include the Brook, Gelman, & Rubin convergence diagnostic (Gelman & Rubin, 1992), the Geweke convergence diagnostic (Geweke, 1992), the Heidelberger and Welch convergence diagnostic (Heidelberger & Welch, 1983), and the Raftery and Lewis convergence diagnostic (Raftery & Lewis, 1992). The first method can be used when two or more parallel chains are run simultaneously. It was not used in this study as only single chains were estimated. The Heidelberger and Welch (1983) convergence diagnostic can be used when a single chain is run and was used for this study as implemented in the Coda software package using R (Plummer et al., 2012).

The Heidelberger and Welch convergence diagnostic consists of two tests. A stationarity test using the Cramer-von-Mises statistic and a halfwidth test. If the posterior means of selected iterations are non-stationary, the test is repeated after discarding the first 10% of the iterations. This process is stopped when the resulting chain passes the stationarity test or when more than 50% of the iterations have been discarded (Smith, 2007). Passing the stationarity test is taken to indicate that convergence has been realized and the number of burn-in and post burn-in is determined at that point. Table 3.1 provides Heidelberger and Welch convergence diagnostic output used in this study. $b1$ in Table 3.1 is the item difficulty parameter for Class 1. Convergence was obtained since the item difficulties for all 20 items in Class 1 passed the stationarity test.

A halfwidth test is based on the chain that has passed the stationarity test for each parameter being estimated. If the halfwidth of the credibility interval for the posterior mean is less than a specified accuracy of this mean, the halfwidth test can be interpreted to mean the parameter has been estimated with acceptable accuracy. If the halfwidth test fails, a longer run of the MCMC chain is required to improve the accuracy of the estimate (Smith, 2007). For monitoring convergence, the success of stationarity test was counted for all item difficulty parameters. We determined the burn-in and post burn-in iterations when all item difficulty parameter passed the stationarity test.

The ratio of the standard deviation of the parameter to the MC standard error for the parameter was also monitored. The usual criterion for this ratio is that the MC standard error should be not more than 5% of the standard deviation of the parameter. In addition, the credibility interval was considered to monitor the convergence especially for Mix3PLM. A credibility interval in Bayesian statistics is analogous to confidence intervals. The computer

Table 3.1: Heidelberger & Welch Convergence Diagnostic for the Example in Chapter 2

	Stationarity Test	Start Iteration	P-value	Halfwidth Test	Mean	Halfwidth
b1[1]	passed	1	0.206	passed	-0.648	0.004
b1[2]	passed	1	0.570	passed	-0.456	0.004
b1[3]	passed	1	0.732	passed	-0.075	0.004
b1[4]	passed	1	0.093	passed	0.512	0.004
b1[5]	passed	1	0.740	passed	0.506	0.004
b1[6]	passed	1	0.351	passed	-1.994	0.005
b1[7]	passed	1	0.437	passed	-1.848	0.005
b1[8]	passed	1	0.222	passed	-1.443	0.004
b1[9]	passed	1	0.298	passed	-1.326	0.004
b1[10]	passed	1	0.471	passed	-0.935	0.004
b1[11]	passed	1	0.126	passed	-0.464	0.004
b1[12]	passed	1	0.468	passed	-0.242	0.004
b1[13]	passed	1	0.175	passed	0.045	0.004
b1[14]	passed	1	0.412	passed	0.382	0.004
b1[15]	passed	1	0.825	passed	0.494	0.004
b1[16]	passed	1	0.905	passed	0.936	0.004
b1[17]	passed	1	0.831	passed	1.384	0.004
b1[18]	passed	1	0.682	passed	1.321	0.004
b1[19]	passed	1	0.487	passed	1.619	0.005
b1[20]	passed	1	0.360	passed	1.973	0.005

software OpenBUGS (Spiegelhalter, Thomas, Best, & Lunn, 2007) provides a 95% credibility interval for all parameter estimates as a default. When the posterior means are placed within the 95% credibility interval, there is an evidence for convergence.

Model Selection

If an IRT model does not fit the data, the model does not support an accurate interpretation of the latent variable (e.g., of ability). When the competing models are nested, the likelihood ratio test can be used for model selection. When models are not nested, then information indices can be used to help inform model selection. In this study, model selection indices were used as MixIRT models are not nested.

Congdon (2003) and Gill (2002) suggest the Bayesian information criterion (BIC; Schwartz, 1978) and the Akaike's Information Criterion (AIC; Akaike, 1973) for model selection in a Bayesian context. Kass (1993) and Kass and Raftery (1995) also suggested using the BIC as a substitute for full calculation of the Bayes factor because calculation without specifying priors is possible in BIC. Bayesian researchers have proposed somewhat different methods as well as the AIC and BIC for model selection: the deviance information criterion (DIC; Spiegelhalter Best, Carlin, & van der Linde, 2002), Hannan-Quinn information criterion (HQC; Hannan & Quinn, 1979; Claeskens & Hjort, 2008), posterior predictive model checks (PPMC; Gelman et al., 2003), and the pseudo-Bayes factor (PsBF; Geisser & Eddy 1979; Gelfand & Dey, 1994). The equations were described at Li et al. (2009) and Hannan and Quinn (1979).

Li et al. (2009) found BIC to be more accurate than AIC, DIC, PPMC, and PsBF for model selection with MixIRT models. Claeskens and Hjort (2008) introduced HQC as a BIC-like criterion but hesitated to use it because of the unclearness of the HQC equation. They concluded that both AIC and the BIC have good properties for model selection because AIC is efficient and the BIC is consistent. Based on the conclusion of Li et al. (2009) and

Claeskens and Hjort (2008), AIC and BIC were determined as a model selection index.

The AIC is very useful in comparing and selecting non-nested models but AIC has a strong bias toward models that overestimated with extra parameters and AIC tended to select the more complex model with an increase in sample size (Gill, 2002; Carlin & Louis, 2001; Sawa, 1978; Li et al., 2009).

AIC and BIC indices can be calculated as

$$AIC = \overline{D(\xi)} + 2p \tag{3.1}$$

$$BIC = \overline{D(\xi)} + p \log N, \tag{3.2}$$

where $\overline{D(\xi)}$ is the posterior mean of the deviance in MCMC estimation, ξ denotes all parameters in the model, p is the number of parameters, and N is the number of examinees. The number of latent classes was determined using the Bayesian information criterion (BIC) as suggested by Li et al. (2009) and AIC was also monitored.

3.2 Simulation Study Design

Conditions Simulated

A simulation study was performed to examine the impact of different identification constraints in the context of three MixIRT models: MixRM, Mix2PLM, and Mix3PLM. The design of the simulation study included the three constraints used in the example described above. These were Constraint 1 (item anchoring), Constraint 2 (person centering), and Constraint 3 (item centering).

In this study, there were four competing candidate models. These included models with from 1- to 4-latent groups. Each iteration calculated posterior mean of the deviance and then calculate AIC and BIC. AIC and BIC were calculated after mean of posterior mean. Lower AIC and BIC values indicate the preferred model.

Li et al. (2009) report differences in recovery for different sample sizes for each of the three dichotomous mixture IRT models, the MixRM, Mix2PLM, and Mix3PLM. The following simulation conditions were used in this study: two sample sizes (600 examinees and 2,400 examinees), two test lengths (20 and 40 items), three different cases of latent groups (1-, 2-, and 3-groups) with different proportions of simulated groups for each of these three MixIRT models (see Table 3.2). For different proportions of simulated groups, 60% and 40% were used for two latent group solution and three latent group solution used three different proportions, 60%, 30%, and 10%, respectively. The simulation conditions are not fully crossed because proportions of group membership and sample sizes differ depending on the MixIRT model being simulated.

Table 3.2: Simulation Conditions

	Number of		Sample Size		
Group	Latent Groups	Proportions	$N = 600$	$N = 1200$	$N = 2400$
G1	1		600	1,200	2,400
G2	2	.60 : .40	360 / 240	720 / 480	1,440 / 960
G3	3	.60 : .30 : .10	360 / 180 / 60	720 / 360 / 120	1,440 / 720 / 240

Note. G1 = one group, G2 = two groups, G3 = three groups

Twenty replications for each condition were generated. Random numbers seeds were used to generate the 60 data sets (= 20 replications $\times$ 3 latent groups). The random numbers were generated without replacement using the generator at the following website: http://stattrek.com/statistics/random-number-generator.aspx. The random numbers were generated in August 2012. Each random number was used as a seed for generating data for each condition using N(0,1) in a program written in R. Code for this program is given in Appendix A.

The first 20 random numbers were used for the 20 replications for the theta parameter generated for the 1-group case and the first group of the 2-group and of 3-group cases. The next 20 random numbers were used for the 20 replications for the theta parameter generations for the second group of the 2-group and 3-group cases, and the last 20 random numbers were used for generating the 20 replications for the theta parameter generations of the third group of the 3-group case.

Table 3.3: Simulated Performance Patterns

Type of Knowledge	Group 1	Group 2	Group 3
1	Good	Average	Poor
2	Average	Poor	Good
3	Poor	Good	Average

There were 20 replications of the 108 conditions: Three identification constraints × two test lengths × two sample sizes × one to three latent groups × three mixture IRT models yields 108 conditions.

Three types of knowledge were simulated in each test as suggested by Li et al. (2009). The generating parameters for the knowledge type are given in Table 3.3. The three types of knowledge are related to sets of responses. Items 1 to 5 have same item parameters for the three groups. Items 6 to 10 measure Type 1 knowlege, Items 11 to 15 measure Type 2 of knowledge, and Items 16 to 20 measure Type 3 knowledge (see Table 3.4). Group 1 is simulated to have good performance in first type of knowledge, average performance in second type of knowledge, and poor performance in last type of knowledge. Three different groups were simulated as performing differently based on the type of knowledge.

Item parameters were modified from Li et al. (2009). Two discrimination parameters were used: A good performance was simulated with a discrimination of 2; a value of 1 was used for average or poor performing groups. Three lower asymptote parameters, .25, .2, and .1, were assigned to high difficulty, medium difficulty, and low difficulty items, respectively. The item parameters for Group 1 were used for the 1-group model. Data for the 2-group model were simulated using the item parameters for Groups 1 and 2. Data for the 3-group model were simulated using the item parameters for Groups 1, 2, and 3. For the 40-item condition, the pattern for the 20-item test was used twice.

Table 3.4: Generating parameters for MixIRT Model Simulations: 25% Anchor Items

Type of			Group 1			Group 2			Group 3		
Knowledge	Anchor	Item	b	a	c	b	a	c	b	a	c
	anchor	1	-0.50	1	.20	-0.50	1	.20	-0.50	1	.20
	anchor	2	-0.50	1	.20	-0.50	1	.20	-0.50	1	.20
	anchor	3	0.00	1	.20	0.00	1	.20	0.00	1	.20
	anchor	4	0.50	1	.20	0.50	1	.20	0.50	1	.20
	anchor	5	0.50	1	.20	0.50	1	.20	0.50	1	.20
1		6	-2.00	2	.10	-0.50	1	.20	1.00	1	.25
		7	-1.75	2	.10	-0.25	1	.20	1.25	1	.25
		8	-1.50	2	.10	0.00	1	.20	1.50	1	.25
		9	-1.25	2	.10	0.25	1	.20	1.75	1	.25
		10	-1.00	2	.10	0.50	1	.20	2.00	1	.25
2		11	-0.50	1	.20	1.00	1	.25	-2.00	2	.10
		12	-0.25	1	.20	1.25	1	.25	-1.75	2	.10
		13	0.00	1	.20	1.50	1	.25	-1.50	2	.10
		14	0.25	1	.20	1.75	1	.25	-1.25	2	.10
		15	0.50	1	.20	2.00	1	.25	-1.00	2	.10
3		16	1.00	1	.25	-2.00	2	.10	-0.50	1	.20
		17	1.25	1	.25	-1.75	2	.10	-0.25	1	.20
		18	1.50	1	.25	-1.50	2	.10	0.00	1	.20
		19	1.75	1	.25	-1.25	2	.10	0.25	1	.20
		20	2.00	1	.25	-1.00	2	.10	0.50	1	.20

Note. a = discrimination parameter; b = difficulty parameter; c = lower asymptote parameter

Table 3.5 presents the numbers of hours needed to complete these runs. The MixRM took between 3.6 and 26 hours to run the data for one condition using OpenBUGS. The Mix2PLM took between 3 and 50 hours and the Mix3PLM took between 9 and 197 hours to complete.

Recovery Evaluation

A recovery analysis was performed to evaluate the accuracy of the estimates of item and group mean parameters. It is important to determine whether the data were simulated

Table 3.5: OpenBUGS Running Times for One Replication

Model	Test Length	Sample Size	Time (hour)
MixRM	20	600	3
MixRM	20	2400	17
MixRM	40	600	6
MixRM	40	2400	26
Mix2PLM	20	600	3
Mix2PLM	20	2400	29
Mix2PLM	40	600	7
Mix2PLM	40	2400	50
Mix3PLM	20	600	9
Mix3PLM	20	2400	71
Mix3PLM	40	600	18
Mix3PLM	40	2400	197

as intended in order to be able to make the intended inferences as to how the different constraints can be expected to affect parameter estimation. The following three indices were calculated to determine accuracy of recovery: BIAS, root mean square error (RMSE), and Pearson correlations.

$$Bias(\widehat{b}) = E(\widehat{b}) - b \tag{3.3}$$

$$RMSE(\widehat{b}) = \sqrt{E[(\widehat{b} - b)^2]} \tag{3.4}$$

$$Cor(\widehat{b}, b) = \frac{Cov(\widehat{b}, b)}{\sigma_{\widehat{b}}\sigma_b} \tag{3.5}$$

Let $\widehat{b}$ be a point estimator for a parameter b. Then $\widehat{b}$ is defined as an unbiased estimator if $E(\widehat{b}) = b$. If not, $\widehat{b}$ is said to be biased. Based on this definition, the bias of a point estimator $\widehat{b}$ is given by Equation 3.3.

RMSE was computed by the square root of mean square error (MSE: the average of the square of the distance between the estimator and its target or generating parameter) in Equation 3.4. If the bias and RMSE increased, an estimator is said to be bad and estimated parameter is considered not to recover generating parameter well.

Harwell, Stone, Hsu, and Kirisci (1996) suggested using the correlation between estimated and true parameters as a criterion in Monte Carlo studies. Using correlations makes it possible to compare variables with different metrics. However, the correlation only explains the rank ordering between the estimated and true parameters and there is no clear criterion to compare the magnitude of correlation. For this study, Pearson correlations (Cor) were computed to evaluate the accuracy of recovery analysis as well as bias and RMSE. The correlation between the estimator and its generating parameter was calculated by Equation 3.5.

The bias, RMSE, and Pearson correlation were computed across items, latent class groups, and replications by Equations 3.6 and 3.8.

$$Bias(\hat{b}) = \frac{\sum_{r=1}^{R}\sum_{g=1}^{G}\sum_{i=1}^{I}(\widehat{b}_{igr} - b_{ig})}{RGI} \tag{3.6}$$

$$RMSE(\hat{b}) = \sqrt{\frac{\sum_{r=1}^{R}\sum_{g=1}^{G}\sum_{i=1}^{I}(\widehat{b}_{igr} - b_{ig})^2}{RGI}} \tag{3.7}$$

$$Cor(\widehat{b}, b) = \frac{1}{R}\sum_{r=1}^{R}\frac{Cov(\widehat{b}_{igr}, b_{ig})}{\sigma_{\widehat{b}_{igr}}\sigma_{b_{ig}}}, \tag{3.8}$$

where $\widehat{b}_{igr}$ is the estimated item difficulty parameter for item i in latent group g for rth replication, b_{ig} is the generating true value of item difficulty for item i in latent group g, R is the number of replications ($r = 1, \ldots, R$), I ($i = 1, \ldots, I$) is the number of items, and G ($g = 1, \ldots, G$) is the number of latent classes in the model being estimated.

Linking of Metrics for Recovery Analysis

In order to estimate BIAS and RMSE, item parameter estimates need to be on the same scale as the generating parameters. The metrics of estimates from each replication were transformed to the metric of generating parameters using the mean and sigma equating method (de Ayala, 2009; Hambleton & Swaminathan, 1985; Kolen & Brennan, 2004). The mean and sigma is a simple and widely used method for IRT equating. The equating constants A and B for the linear transformation were computed as follows:

$$A = \frac{S_{b_B}}{S_{b_T}} \tag{3.9}$$

$$B = \bar{b}_B - A\bar{b}_T, \tag{3.10}$$

where b_B is a generating item difficulty to be placed onto the base metric of the generating parameter, b_T is the estimated item difficulty on the target scale, S is the standard deviation and $\bar{b}$ is the mean of item difficulty parameters (Hambleton & Swaminathan, 1985). The estimated parameters from each replication were equated and placed onto the metric of the generating parameter using the mean and sigma method as shown in Equations 3.11 to 3.13. BIAS, RMSE, and correlations were computed after this transformation.

$${b_i}^* = Ab_i + B = \frac{S_{b_B}}{S_{b_T}} b_i + (\bar{b}_B - \frac{S_{b_B}}{S_{b_T}} \bar{b}_T) \tag{3.11}$$

$${a_i}^* = \frac{a_i}{A} = a_i \frac{S_{b_T}}{S_{b_B}} \tag{3.12}$$

$$c_i{}^* = c_i, \tag{3.13}$$

where b_i is item difficulty parameter of item i, a_i is item discrimination parameter of item i, and c_i is lower asymptote parameter of item i. $b_i{}^*$, $a_i{}^*$, and $c_i{}^*$ represent estimated parameters transformed by mean and sigma equating method. They were calculated from the Table 3.4.

Chapter 4

RESULTS

The purpose of this study was to compare the effects of three different constraints for identifying the metric in MixIRT models. In this chapter, we present results from a simulation study examining the effects of these on three different dichotomous MixIRT models, the MixRM, the Mix2PLM and the Mix3PLM. The following conditions were generated: two different test lengths (20 and 40 items), two different sample sizes (600 and 2,400 examinees), three different numbers of latent classes (1-, 2-, and 3- latent classes), and three different constraints (item anchoring, person centering, and item centering).

4.1 Results of the Simulation Study

Monitoring Convergence

Three convergence diagnostics were used to determine convergence, the Heidelberger and Welch (1983) convergence diagnostics, the ratio of the standard deviation of the parameter estimate to the MC standard error for the parameter estimate, and the 95% credibility interval. For the MixRM and Mix2PLM, estimation chains were monitored using Heidelberger and Welch convergence diagnostics and the ratio of the standard deviation of the parameter to the MC standard error for the parameter. The 95% credibility interval was used to monitor the convergence of the Mix3PLM.

Tables E.1 and E.2 in Appendix E contain the convergence results from the Heidelberger & Welch test for the MixRM and Mix2PLM. These tables show the convergence results for the MixRM and Mix2PLM, person centering, 20- and 40-items, 600 and 2,400 examinees, and three-group solution. The total number of parameters passing the stationarity test was counted to evaluate convergence. For 20-item condition, the total number passing the test should be 200 (20 for the one-group model, 40 for the two-group model, 60 for the three-group, 80 for the four-group model) and 400 for the 40-item condition (40 for the one-group model, 80 for the two-group model, 120 for the three-group model, 160 for the four-group model). Convergence of all parameters was achieved when the value in the last column in Tables E.1 and E.2 equaled 1,200. As can be seen in these tables, convergence was good but not perfect.

The chain for the MixRM was found to have converged perfectly for all parameters after a burn-in of 5,000 iterations and 5,000 post-burn-in iterations (see Table E.1). The chain for the Mix2PLM did not converge perfectly, but was very close with 1,185 estimates passing among the 1,200 total with a burn-in of 6,000 iterations and 11,000 post burn-in iterations (see Table E.2). Autocorrelations, density plots, and history plots also provided evidence of convergence for the MixRM and Mix2PLM (see Figures F.1 to F.3).

The Mix3PLM failed to converge after 35,000 iterations based on the Heidelberger and Welch diagnostics and the ratio of the standard deviation of the parameter estimate to the MC standard error for the parameter estimate. Autocorrelation plots, density plots, and history plots also failed to show convergence (see Figures G.1 to G.3). As an alternative approach, the 95% credibility interval was used. A credibility interval in Bayesian statistics is analogous to confidence intervals in frequentist statistics. The computer software OpenBUGS (Spiegelhalter et al., 2007) used in this study provides a 95% credibility interval as a default for all parameter estimates. Based on the 95% credibility interval, the Mix3PLMs were considered to have converged after a burn-in of 6,000 iterations and 11,000 post burn-in iterations (see Tables E.3 to E.10).

To test convergence using the ratio of the standard deviation of the parameter to the MC standard error for the parameter, MC standard error and standard deviation were derived from OpenBUGS stat output files and the ratio was calculated. This method also supported the results of Heidelberger & Welch diagnostics for MixRM and Mix2PLM. However, the parameters for Mix3PLM are not converged using this method. Only the method using 95% credibility interval provided the evidence of convergence for Mix3PLM.

Model Selection

Exploratory MixRM, Mix2PLM, and Mix3PLM analyses were done to determine the best fitting model to the simulated data. The criterion used for model selection was BIC as suggested by Li et al. (2009). AIC was provided as a comparison index. The percentages in Table 4.1 indicate the number of correct model selection decisions for each condition.

Model selection for the MixRM using BIC was close to 100 percent correct for all conditions except one condition for Constraint 1 (item anchoring) for 20 items, 2,400 simulees, and the 2-group model had 95 percent agreement. All model selections were perfect for the Mix2PLM. For the Mix3PLM, however, there were some low percentages of correct model selection. These occurred under Constraint 2 (person centering) and Constraint 3 (item centering). These results are discussed below.

Model Selection using BIC

The percentages of correct detections shown in Table 4.1 are also plotted in Figures 4.1 to 4.3 to help provide another indication of the effects of the three constraints. In the figures, the three lines indicated the type of MixIRTM (i.e., the MixRM, Mix2PLM, and

Table 4.1: Percent of Correct Model Selections for the MixIRTM

Constraint	Item	Sample	Latent Classes	BIC MixRM	BIC Mix2PLM	BIC Mix3PLM	AIC MixRM	AIC Mix2PLM	AIC Mix3PLM
1. Item Anchoring	20	600	1	100	100	100	95	100	100
			2	100	100	100	90	100	100
			3	100	100	100	95	100	100
		2400	1	100	100	100	85	100	100
			2	95	100	95	**70**	90	95
			3	100	100	95	**65**	95	95
	40	600	1	100	100	100	100	100	100
			2	100	100	100	100	100	100
			3	100	100	100	100	100	100
		2400	1	100	100	90	100	100	90
			2	100	100	100	100	100	100
			3	100	100	100	100	100	100
2. Person Centering	20	600	1	100	100	100	90	100	100
			2	100	100	100	90	100	100
			3	100	100	**40**	100	100	100
		2400	1	100	100	**45**	**65**	100	**45**
			2	100	100	90	**75**	95	**65**
			3	100	100	90	**55**	100	**60**
	40	600	1	100	100	90	100	100	**65**
			2	100	100	95	100	100	95
			3	100	100	**25**	100	100	100
		2400	1	100	100	**45**	100	100	**45**
			2	100	100	95	100	100	95
			3	100	100	95	100	100	85
3. Item Centering	20	600	1	100	100	100	95	100	100
			2	100	100	100	95	100	100
			3	100	100	**40**	100	100	100
		2400	1	100	100	100	**75**	100	100
			2	100	100	100	**75**	95	95
			3	100	100	100	**70**	95	90
	40	600	1	100	100	100	100	100	100
			2	100	100	95	100	100	95
			3	100	100	**25**	100	100	100
		2400	1	100	100	**25**	100	100	**25**
			2	100	100	**60**	100	100	**60**
			3	100	100	**30**	100	100	**70**
MixIRTM				99.86	100.00	82.36	91.25	99.17	88.06
Constraint 1				99.58	100.00	98.33	91.67	98.75	98.33
Constraint 2				100.00	100.00	75.83	89.58	99.58	79.58
Constraint 3				100.00	100.00	72.92	92.50	99.17	86.25
	20-items			99.72	100.00	88.61	82.50	98.33	91.39
	40 items			100.00	100.00	76.11	100.00	100.00	84.72
		n=600		100.00	100.00	82.78	98.33	100.00	97.22
		n=2400		99.72	100.00	80.83	85.28	98.33	78.61
			1-class	100.00	100.00	82.92	92.08	100.00	80.83
			2-class	99.58	100.00	94.17	91.25	98.33	91.67
			3-class	100.00	100.00	70.00	90.42	99.17	91.67

Mix3PLM) and the y-axis describes the percentages of correct detection using BIC index. The simulation conditions are described along the x-axis (I2 indicates 20-items, I4 indicates 40-items, S6 indicates 600 examinees, S24 indicates 2,400 examinees, G1 to G3 indicate one-group to three-group models). For example, I2S6G1 indicates the condition with 20-items, 600 examinees, one group solution under 20 replications.

Model Selection using BIC under Constraint 1 (Item Anchoring). The plots in Figure 4.1 for Constraint 1 clearly indicate that the three MixIRTMs provided nearly the same results for all conditions for both the MixRM and Mix2PLM. Those places where the plots separate mainly occur where less than 100 percent correct detections were observed. The percent of correct model selection was 99.58% for MixRM, 100.00% for Mix2PLM, and 98.33% for Mix3PLM in Table 4.1. Constraint 1 (item anchoring) did not affect the model selection for any of the MixIRTMs.

Model Selection using BIC Under Constraint 2 (Person Centering). Unlike results for the MixRM and Mix2PLM, there were clearly some problems for the Mix3PLM, with Constraint 2 (person centering) (See Figure 4.2. For the smaller sample size ($n = 600$), the Mix3PLM detected fewer correct models simulated with 3 latent classes. Likewise, for the larger sample size (n = 2,400), the Mix3PLM detected fewer models simulated with 1 latent class. For 3 latent groups and the smaller sample size conditions, correct model selection was 40 percent for 20 items and 25 percent for 40 items (see Figure 4.2). When these data sets were generated, the numbers of students for each group in the small sample ($n = 600$) with three groups condition were 360, 180, and 60, respectively. It is possible that the small sample size for class 3 of 60 examinees might not be sufficient to estimate all the parameters in the Mix3PLM. In addition, increasing test length to 40 items but with the same smaller sample ($n = 600$) may not have provided sufficient information for accurate estimation of all model parameters.

In addition, for the larger sample size (n = 2,400) and one latent group condition in the

Mix3PLM, only 45 percent correct detections were observed for both the 20- and 40-item tests. The one group solution is the usual IRT with no latent classes. In this case, it appears that under Constraint 2 (person centering), model selection did not work well for the usual 3PL IRT model. These results suggest that person centering affected model selection using the BIC index, when the larger sample size was simulated for the one latent class model (i.e., a 3PL model without any latent classes).

Model Selection using BIC Under Constraint 3 (Item Centering). Model selection was perfect for the MixRM and Mix2PLM under Constraint 3 (item centering). Problems were observed for the Mix3PLM under this constraint. For the smaller sample size (n = 600) and 3-latent group model conditions, correct model selection was 40 % for the 20-item test and 25 % for the 40-item test condition (see Figure 4.3). This result was similar to what was observed under Constraint 2 (person centering). For the longer test (40 items) × larger sample (n = 2,400) condition, the percent of correct model selection was low regardless of the number of latent groups simulated (e.g., one group = 25 %, two groups = 60 %, three groups = 30 %). One conclusion appears to be that Constraint 3 (item centering) did not work well in two conditions, one was the smaller sample size and the 3-latent group model and the other was the longer test length and larger sample size regardless of the number of latent groups in the model.

When the three types of constraints were compared, there was no great difference in model selection accuracy between the MixRM and the Mix2PLM. When Constraint 2 (person centering) and Constraint 3 (item centering) were compared for the Mix3PLM, there was no difference in the percentage of correct model selections for the smaller sample size (600 examinees). Constraint 2 (person centering) had higher a percentage than Constraint 3 (item centering), however, in the larger sample size condition regardless of test length or the number of latent classes in the model. Constraint 3 (item centering) had higher percentages of correct model selection than Constraint 2 (person centering) in the smaller sample size regardless of test length or number of latent classes. The percentage were low for two simulation conditions (the 40-items, 600 examinees, and 3-group condition and the 40-items, 2,400 examinees, and 1-group condition) for both Constraints 2 and 3 (see Table 4.1).

Overall, model selection for the MixRM and Mix2PLM was close to 100 percent for all three constraints. Model selection under Constraint 1 (item anchoring) was better for the Mix3PLM than under Constraints 2 or 3. There was poor model selection in the mix3PLM with Constraint 2 (person centering) when the smaller sample size and 3-group model were simulated and when the larger sample size and 1-group model were simulated. Under Constraint 3 (item centering), low percentages of correct detections were observed for the Mix3PLM for the longer test (i.e., 40 items) $\times$ larger sample size (i.e., n = 2,400).

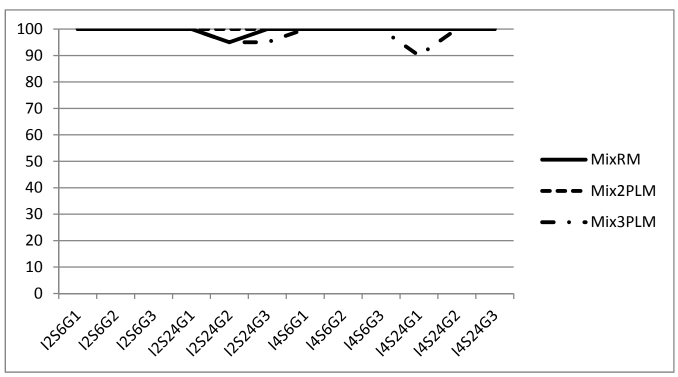


Figure 4.1: Percent Correct Detections with BIC for Constraint 1 (Item Anchoring)

Model Selection using AIC

Although BIC was used as the criterion for model selection in this study, results for AIC are provided here for comparison purposes, as AIC is also often used for model selection. The main problem with AIC is that it has been found to be sensitive to model complexity (Li

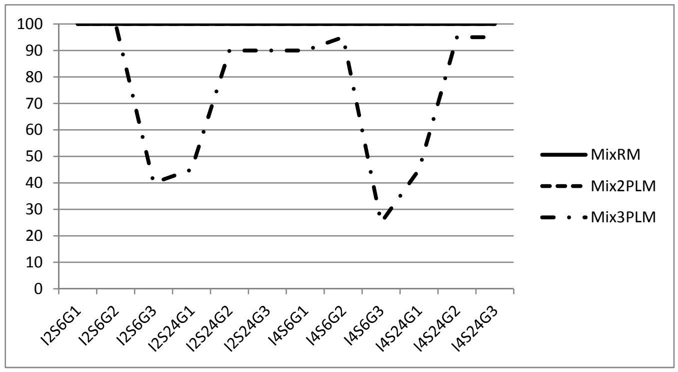


Figure 4.2: Percent Correct Detections with BIC for Constraint 2 (Person Anchoring)

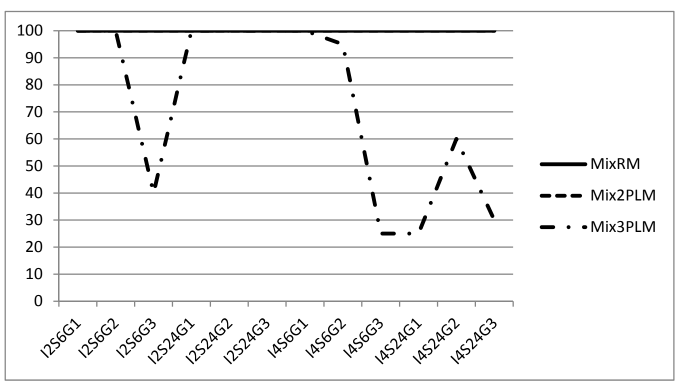


Figure 4.3: Percent Correct Detections with BIC for Constraint 3 (Item Centering)

et al., 2009). That is, AIC has been shown to have a tendency to select the more complex model regardless of the generating (i.e., the correct) model.

Table 4.1 presents model selection results for AIC under each constraint. These correct detection rates are also plotted Figures 4.4 to 4.6.

Model Selection using AIC under Constraint 1 (Item Anchoring). The plots in Figure 4.4 indicate that correct model selection was poor only for the shorter test (i.e., 20 items) $\times$ larger sample size (n = 2,400) condition for the MixRM. The percent of correct model selection with the shorter test $\times$ larger sample size was low for this model with correct detection percentages of 85, 70, and 65 for the one- to three-latent group models, respectively. AIC detected the correct model relatively well under Constraint 1, however, for the Mix2PLM and Mix3PLM.

Model Selection using AIC under Constraint 2 (Person Centering). When Constraint 2 (person centering) was used, low percentages of correct detection using AIC were observed for the 20-item $\times$ 2,400 sample size condition for both the MixRM and the Mix3PLM (see also Figure 4.5). For the MixRM, low these percentages were observed: 65 % for the 1-group model, 75 % for the 2-group model, and 55 % for the 3-group model. Low percentages of correct model selection also were observed for the Mix3PLM of 45 percent for the one-group condition, 65 percent for the two-group condition, and 60 percent for the three-group condition. In addition, 65 percent of correct model selections were observed for the shorter 20-item test $\times$ smaller sample size (n = 600) $\times$ one-group condition. For the longer test (i.e., 40 items) $\times$ larger sample size (i.e., n = 2,400) condition, 45 percent of correct detection were observed for the one-group and 85 percent for the three-group conditions.

Model Selection using AIC under Constraint 3 (Item Centering). Correct models were selected between 95 and 100 percent of the time for the Mix2PLM (see also Figure 4.6). For the MixRM and Mix3PLM, however, there were some conditions with low percentages of correct selection. For the MixRM, the shorter test length (20 items) × larger sample size ($n = 2{,}400$) condition had lower percentages of correct model selection (75 percent for the one- and two-group conditions and 70 percent for the three-group condition). For the longer test length (40 items) × larger sample size ($n = 2{,}400$) condition, the Mix3PLM had low percentages of correct model selection regardless of the number of latent classes in the model (25 percent for the one-group model, 60 percent for the two-group model, 70 percent for the three-group model. As can be seen in Table 4.1, this pattern was similar to that observed for Constraint 3 for the Mix3PLM when BIC was used.

Overall, AIC results for the Mix2PLM indicated good model selection results for all three constraints. For the MixRM, however, the percent of correct model selection decreased for shorter test lengths (20 items) × larger sample size ($n = 2{,}400$). Selection of the correct model was good for the Mix3PLM under Constraint 1 (item anchoring). Under Constraints 2 (person centering) and 3 (item centering), however, the Mix3PLM had lower percentages of correct selection.

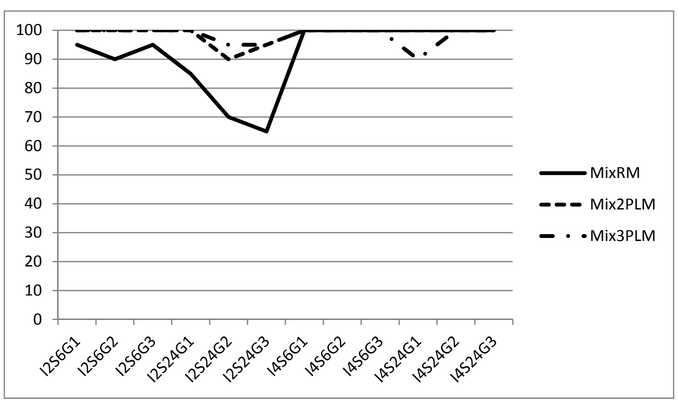


Figure 4.4: The Percent of Correct Number of Latent Classes with Item Anchoring Constraint Based on AIC

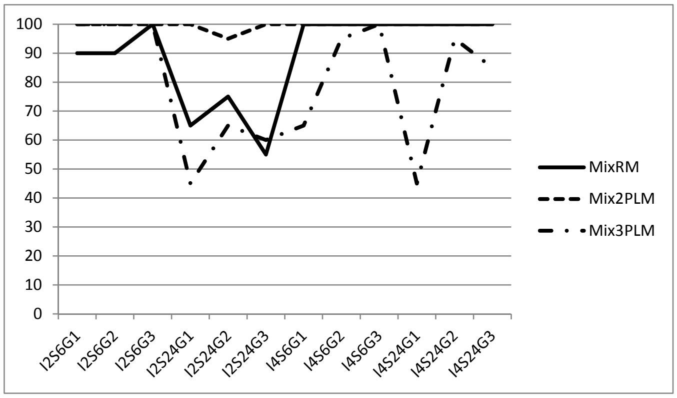


Figure 4.5: The Percent of Correct Number of Latent Classes with Person Centering Constraint Based on AIC

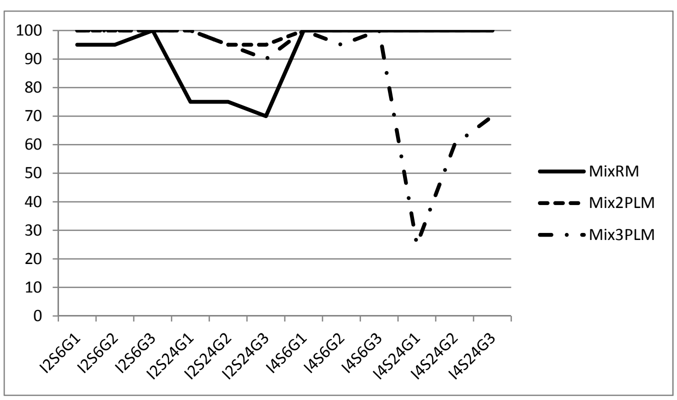


Figure 4.6: The Percent of Correct Number of Latent Classes with Item Centering Constraint Based on AIC

Comparison of Model Selection for AIC and BIC

As can be seen in Table 4.1, for the MixRM and Mix2PLM, BIC had more correct model selections than AIC. For the Mix3PLM, BIC had more often correct model selections than AIC except for three conditions: longer test length × smaller sample size × 3-class using Constraints 2 (person centering) and 3 (item centering), longer test length × larger sample size × 3-class using Constraint 3 (item centering). This is agreement with results from Li et al. (2009) which found that BIC made more correct model selections for all three MixIRTMs.

Comparisons Between BIC and AIC for the MixRM and Mix2PLM. For the MixRM and Mix2PLM, BIC selected the correct model more often than did AIC for all conditions (i.e., type of constraint, test length, sample size, and number of latent group) at more than 99.5%. AIC and BIC, however, had about the same close to 100% of correct model selections for the longer test length in the MixRM and for the longer test length, smaller sample size, and 1-group model in the Mix2PLM.

Comparisons Between BIC and AIC for the Mix3PLM. AIC and BIC seemed to work least well in Mix3PLM than in either the MixRM or Mix2PLM. AIC and BIC were equally accurate at selection of the correct model with 98.33 percent under Constraint 1 (item anchoring). However, AIC more often selected the correct model than BIC under Constraints 2 (person centering) and 3 (item centering). For both test lengths, AIC was more accurate than BIC. In addition, for the small sample size, AIC was more accurate, whereas BIC was more accurate for the large sample size. For the 1-class and 2-class models, BIC was more accurate and slightly better. For the 3-class model, however, AIC was much better with 91.67 percent correct compared to 70 percent correct for BIC.

Among the 36 condition, AIC and BIC selected correct models for 25 conditions. BIC selected more correct models for 6 conditions and AIC did so for 5 conditions. For the small sample size, 3-class model, and for Constraints 2 (person centering) and 3 (item centering), AIC was more accurate. For the longer test length, larger sample size, 3-class models, AIC also was more accurate. BIC was more accurate when the shorter test length, larger sample size, and Constraint 2 (person centering) was used. For the longer test length, larger sample size, and 3-class model, and Constraint 2 (person centering) and for the shorter test length, larger sample size, 2- and 3-class models, and Constraint 3 (item centering), BIC was slightly more accurate.

Label Switching

The possibility of label switching needs to be monitored in a simulation study with MixIRT models in order to determine recovery of group membership. Label switching occurred in this study when the latent classes switched between replications. As was noted earlier, when latent classes switch on a given replication, it can cause confusion in interpretation of results (Li et al., 2009). This type of label switching is easily observed in simulation studies because the generating parameters are known and can be compared with the estimated parameters for each of the latent classes (Cho, Cohen, & Kim, 2006; Li et al., 2009).

In this study, replications were monitored to determine whether or not label switching had occurred. When it was observed, the problem was solved by comparing frequencies between generated group membership and the posterior mode estimates of group membership. The latent classes of each replication were switched based on the frequency comparisons prior to the recovery analysis.

Table 4.2 indicates how many time label switching was observed in the simulation study. Percent was calculated by the number of times label switching was observed over the number of correct model selection based on BIC index. For example, the condition having 20 replications with the Mix3PLM, Constraint 3 (item centering), 40-items, 2,400 examinees, and 3-class model had 100 % (= 6/6) model selection. This condition, however, had only 6 data sets and the correct model was selected in all six. In addition, thought, label switching was observed in all six data sets. There was not much difference in percent correct any of the

MixIRTMs, test length, or sample size. However, the percent of label switching was very low (14%), when Constraint 1 (item anchoring) was used in the Mix2PLM. When the number of latent classes in the model increased, the percent of label switching increased as well.

Recovery Analysis

A recovery analysis was done to evaluate whether the estimation algorithms based on each of the constraints had an effect on recovery of the generating parameters from the simulated data. The item and latent group mean parameters used to generate the data were compared to the estimated item and latent group mean parameters. Recovery of item and latent group mean parameters were assessed using bias, root mean square error (RMSE), and Pearson correlations between the generating parameters and the estimated parameters. Before recovery analysis, all estimated parameters were been placed onto the metric of the generating parameters using the mean and sigma equating method.

RECOVERY OF ITEM AND LATENT GROUP MEAN PARAMETERS

Recovery Analysis for the MixRM. Bias, RMSE, and correlations for the MixRM between generating parameters and parameter estimates are reported in Table 4.3 for each of the simulation conditions. All bias and correlation statistics appeared to indicate that generating parameters were recovered successfully. Bias values were all zero for item difficulty and very small for latent group means, ranging between -.002 and .002. Correlations between generating values and item difficulty estimates were all high, ranging from .979 to .999.

RMSEs for item difficulty ranged from .049 to .229. Most RMSEs were less than .144, suggesting relatively good recovery. Values above .2 can be taken as indicating conditions that may be causing problems with recovery. These values are presented in bold in Table 4.3. RMSEs of .144 to .160 were observed in the 2-class $\times$ n = 600 conditions for all three constraints suggesting the algorithm had difficulty recovering the generating values for these

Table 4.2: Percent of Label Switching under 20 replications for MixIRTMs

Constraint	Item	Sample	Latent Classes	MixRM	Mix2PLM	Mix3PLM
1. Item anchoring	20	600	1	0	0	0
			2	65	10	60
			3	75	60	100
		2400	1	0	0	0
			2	74	0	90
			3	50	75	100
	40	600	1	0	0	0
			2	0	0	5
			3	100	25	100
		2400	1	0	0	0
			2	0	0	10
			3	100	0	100
2. Person centering	20	600	1	0	0	0
			2	0	100	50
			3	95	100	38
		2400	1	0	0	0
			2	35	15	94
			3	100	100	56
	40	600	1	0	0	0
			2	5	15	11
			3	100	90	100
		2400	1	0	0	0
			2	0	100	100
			3	100	100	95
3. Item centering	20	600	1	0	0	0
			2	0	100	100
			3	100	0	50
		2400	1	0	0	0
			2	0	100	90
			3	100	0	10
	40	600	1	0	0	0
			2	100	100	100
			3	90	85	60
		2400	1	0	0	0
			2	100	100	75
			3	100	10	100
MixIRTM				41	36	47
Constraint 1				39	14	47
Constraint 2				36	52	45
Constraint 3				49	41	49
	20-items			39	37	47
	40-items			44	35	48
		n = 600		41	38	43
		n = 2400		42	33	51
			1-class	0	0	0
			2-class	32	53	65
			3-class	93	54	76

conditions. Values above .2 were observed for the 3-class $\times$ $n = 600$ conditions for all three constraints. These results suggest that recovery was negatively affected by the small sample size (i.e., $n = 600$) as the number of latent classes increased (see also Figure 4.7). The type of identification constraint did not appear to affect recovery of item difficulties as indexed by RMSEs. Recovery of the generating item and group mean parameters was also generally satisfactory.

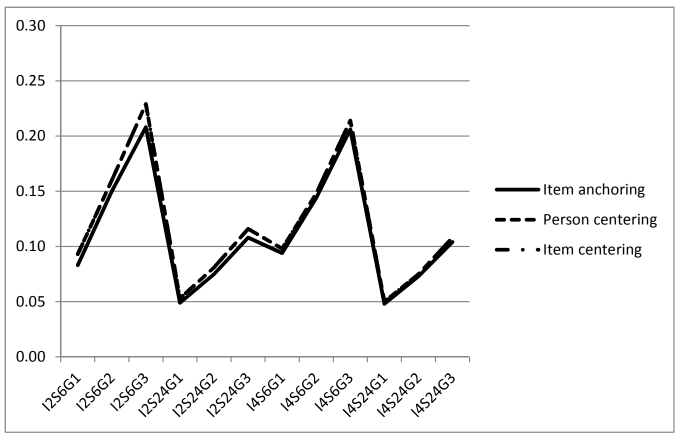


Figure 4.7: RMSEs for Item Difficulties in the MixRM

Recovery Analysis for the Mix2PLM. Bias statistics for item difficulty and discrimination estimates are given in Table 4.4. Bias statistics were all zero for item difficulty. RMSEs for item difficulty were also generally small, most ranged from .059 to .133. There were some, however, that ranged from .164 to .222. These latter values all occurred for the 3-class model in the small sample ($n = 600$) conditions for all three constraints. This was similar to the RMSE results for item difficulty for the MixRM. Correlations for item difficulty were high and ranged from .980 to .999. Recovery of item difficulty for the Mix2PLM was generally good with the possible exceptions of the 3-class model for the small sample conditions. The type of constraint did not appear to affect recovery of item difficulty.

Table 4.3: BIAS, RMSE and Correlations (Cor.) of Difficulty (β) Parameters and Latent Group Mean ($\bar{\theta}$) in MixRM over 20 Replications

Constraint	Item	Sample	Latent Classes	Item Difficulty BIAS	Item Difficulty RMSE	Item Difficulty Cor.	Latent Group Mean BIAS	Latent Group Mean RMSE
1. Item Anchoring	20	600	1	0.000	0.083	0.997	0.001	0.004
			2	0.000	0.150	0.991	0.001	0.008
			3	0.000	**0.208**	0.983	0.002	0.016
		2400	1	0.000	0.049	0.999	-0.001	0.004
			2	0.000	0.075	0.998	-0.001	0.008
			3	0.000	0.108	0.996	-0.002	0.014
	40	600	1	0.000	0.094	0.996	0.001	0.003
			2	0.000	0.144	0.992	0.000	0.015
			3	0.000	**0.206**	0.983	0.000	0.017
		2400	1	0.000	0.048	0.999	0.000	0.001
			2	0.000	0.073	0.998	-0.001	0.009
			3	0.000	0.104	0.996	0.000	0.005
2. Person Centering	20	600	1	0.000	0.093	0.997	0.001	0.003
			2	0.000	0.160	0.990	0.001	0.007
			3	0.000	**0.229**	0.979	0.001	0.015
		2400	1	0.000	0.053	0.999	-0.001	0.004
			2	0.000	0.081	0.998	-0.001	0.008
			3	0.000	0.116	0.995	-0.002	0.016
	40	600	1	0.000	0.098	0.996	0.001	0.003
			2	0.000	0.148	0.991	0.000	0.014
			3	0.000	**0.214**	0.982	-0.001	0.014
		2400	1	0.000	0.050	0.999	0.000	0.001
			2	0.000	0.075	0.998	-0.001	0.009
			3	0.000	0.108	0.995	0.000	0.005
3. Item Centering	20	600	1	0.000	0.093	0.997	0.002	0.008
			2	0.000	0.160	0.990	0.002	0.012
			3	0.000	**0.229**	0.979	0.001	0.018
		2400	1	0.000	0.053	0.999	0.003	0.014
			2	0.000	0.081	0.998	-0.005	0.021
			3	0.000	0.116	0.995	-0.002	0.011
	40	600	1	0.000	0.098	0.996	-0.001	0.004
			2	0.000	0.148	0.991	0.000	0.014
			3	0.000	**0.214**	0.982	-0.001	0.017
		2400	1	0.000	0.050	0.999	-0.001	0.003
			2	0.000	0.075	0.998	0.001	0.027
			3	0.000	0.108	0.995	0.000	0.002
Constraint 1				0.000	0.112	0.994	0.000	0.009
Constraint 2				0.000	0.119	0.993	0.000	0.008
Constraint 3				0.000	0.119	0.993	0.000	0.013
	20-items			0.000	0.119	0.993	0.000	0.011
	40-items			0.000	0.114	0.994	0.000	0.009
		$n = 600$		0.000	0.154	0.990	0.001	0.011
		$n = 2400$		0.000	0.079	0.997	-0.001	0.009
			1-class	0.000	0.072	0.998	0.000	0.004
			2-class	0.000	0.114	0.994	0.000	0.013
			3-class	0.000	0.163	0.988	0.000	0.013

Note. When RMSE is larger than .2 or correlation is less than .8, the values are bold.

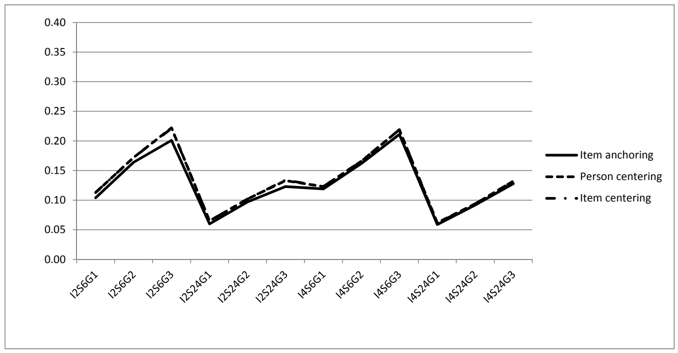


Figure 4.8: RMSEs of Item Difficulties by the Type of Identification Constraint in the Mixture 2PL Model

Bias statistics for item discrimination parameters were also relatively small, ranging from .002 to .043. RMSEs for item discrimination parameters appeared to be higher for the small sample conditions, ranging from .150 to .316. RMSEs in the large sample conditions were slightly smaller, ranging from .080 to .199. This result suggests a sample size effect on recovery of item discrimination. In addition, RMSEs for the 2-class and 3-class conditions were larger than those in the 1-class conditions. Correlations for item discrimination generally moderately high to high, ranging from .772 to .985. Correlations for the 1-class conditions ranged from .938 to .985. Correlations for the 2-class conditions ranged from .861 to .965. Correlations for the 3-class conditions ranged from .772 to .928. These correlations suggest that the number of latent classes in the model had an impact on the recovery of the discrimination parameters.

The bias and RMSE for latent group mean parameter estimates appeared to be recovered successfully. All bias values were less than .005 and all RMSE statistics were less than .027.

Table 4.4: BIAS, RMSE and Correlations (Cor.) of Difficulty (β) and Discrimination (a) Parameters and Latent Group Mean ($\bar{\theta}$) in Mix2PLM over 20 Replications

Constraint	Item	Sample	Latent Classes	Item Difficulty BIAS	Item Difficulty RMSE	Item Difficulty Cor.	Item Discrimination BIAS	Item Discrimination RMSE	Item Discrimination Cor.	Latent Group Mean BIAS	Latent Group Mean RMSE
1. Item Anchoring	20	600	1	0.000	0.104	0.996	0.039	0.157	0.947	0.003	0.015
			2	0.000	0.164	0.989	0.037	**0.228**	0.873	0.001	0.013
			3	0.000	**0.201**	0.984	0.014	**0.286**	0.800	0.005	0.024
		2400	1	0.000	0.060	0.999	0.034	0.092	0.985	-0.001	0.004
			2	0.000	0.097	0.996	0.042	0.143	0.955	-0.002	0.011
			3	0.000	0.123	0.994	0.043	0.181	0.922	0.000	0.009
	40	600	1	0.000	0.119	0.994	0.028	0.150	0.948	0.001	0.005
			2	0.000	0.162	0.990	0.023	**0.219**	0.882	0.001	0.007
			3	0.000	**0.211**	0.982	0.003	**0.265**	0.819	0.002	0.027
		2400	1	0.000	0.059	0.999	0.014	0.080	0.984	0.001	0.003
			2	0.000	0.092	0.997	0.015	0.118	0.965	-0.001	0.005
			3	0.000	0.128	0.993	0.025	0.171	0.928	0.000	0.009
2. Person Centering	20	600	1	0.000	0.113	0.995	0.043	0.169	0.938	0.003	0.013
			2	0.000	0.172	0.988	0.037	**0.239**	0.862	0.001	0.005
			3	0.000	**0.222**	0.980	0.009	**0.313**	**0.776**	0.004	0.024
		2400	1	0.000	0.065	0.999	0.034	0.097	0.983	-0.001	0.004
			2	0.000	0.102	0.996	0.043	0.150	0.950	-0.002	0.018
			3	0.000	0.133	0.993	0.043	0.199	0.909	0.001	0.005
	40	600	1	0.000	0.123	0.994	0.028	0.156	0.943	0.001	0.004
			2	0.000	0.166	0.989	0.022	**0.224**	0.878	0.001	0.005
			3	0.000	**0.219**	0.981	0.003	**0.279**	0.807	0.001	0.008
		2400	1	0.000	0.062	0.999	0.013	0.083	0.983	0.001	0.002
			2	0.000	0.094	0.997	0.016	0.122	0.963	-0.001	0.011
			3	0.000	0.132	0.993	0.026	0.177	0.924	0.001	0.014
3. Item Centering	20	600	1	0.000	0.113	0.995	0.043	0.169	0.938	0.004	0.020
			2	0.000	0.172	0.988	0.038	**0.240**	0.861	0.000	0.018
			3	0.000	**0.221**	0.980	0.007	**0.316**	**0.772**	0.005	0.029
		2400	1	0.000	0.065	0.999	0.034	0.097	0.983	-0.004	0.019
			2	0.000	0.102	0.996	0.043	0.150	0.950	0.001	0.005
			3	0.000	0.134	0.993	0.043	0.199	0.908	0.001	0.011
	40	600	1	0.000	0.123	0.994	0.028	0.156	0.943	0.003	0.013
			2	0.000	0.166	0.989	0.023	**0.224**	0.878	0.001	0.008
			3	0.000	**0.219**	0.981	0.002	**0.280**	0.805	0.003	0.031
		2400	1	0.000	0.061	0.999	0.013	0.083	0.983	-0.001	0.003
			2	0.000	0.094	0.997	0.016	0.122	0.963	0.001	0.011
			3	0.000	0.132	0.993	0.026	0.177	0.924	0.001	0.012
Constraint 1				0.000	0.127	0.993	0.026	0.174	0.917	0.001	0.011
Constraint 2				0.000	0.134	0.992	0.026	0.184	0.910	0.001	0.009
Constraint 3				0.000	0.134	0.992	0.026	0.184	0.909	0.001	0.015
	20-items			0.000	0.131	0.992	0.035	0.190	0.906	0.001	0.014
	40-items			0.000	0.131	0.992	0.018	0.171	0.918	0.001	0.010
		$n = 600$		0.000	0.166	0.988	0.024	0.226	0.871	0.002	0.015
		$n = 2400$		0.000	0.096	0.996	0.029	0.136	0.953	0.000	0.009
			1-class	0.000	0.089	0.997	0.029	0.124	0.963	0.001	0.009
			2-class	0.000	0.132	0.993	0.030	0.182	0.915	0.000	0.010
			3-class	0.000	0.173	0.987	0.020	0.237	0.858	0.002	0.017

Note. When RMSE is larger than .2 or correlation is less than .8, the values are bold.

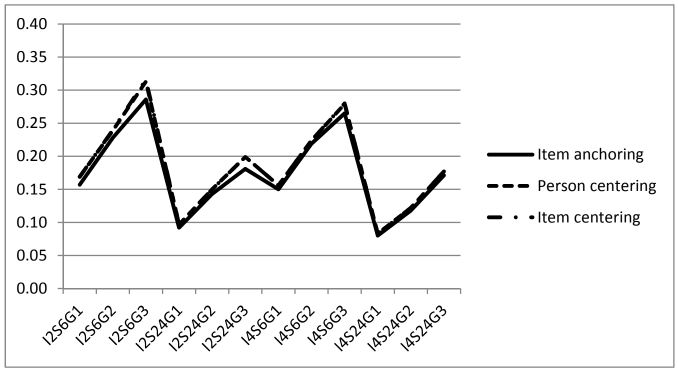


Figure 4.9: RMSEs of Item Discriminations by the Type of Identification Constraint in the Mixure 2PL Model

Recovery Analysis for the Mix3PLM. Table 4.5 presents the Bias, RMSE, and correlation results for recovery analysis of item and group mean parameters in the Mix3PLM. Bias and RMSE values were all close to zero for group mean parameters indicating these parameters were recovered well. Bias for the item difficulty and lower asymptote parameters were close to zero as well, although there were some simulation conditions which had absolute Bias values higher than .3 for the item discrimination parameters. These conditions included the 2,400 students $\times$ one group condition for Constraint 2 (person centering) and for Constraint 3 (item centering) in the 40-items $\times$ 2,400 students $\times$ one group condition.

RMSEs for the Mix3PLM are plotted in Figures 4.10 to 4.12 for each of the three identification constraints. The pattern of RMSEs for item difficulty and discrimination appear similar with each other. The RMSEs for the lower asymptote parameters in Figures 4.12 clearly indicate that all conditions had lower RMSEs except for the 20-items $\times$ 2,400 students one group condition for Constraint 2 (person centering).

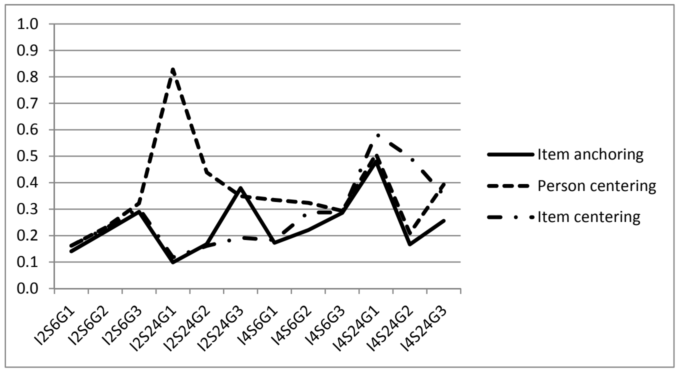


Figure 4.10: RMSEs for Item Difficulties for the Mix3PLM

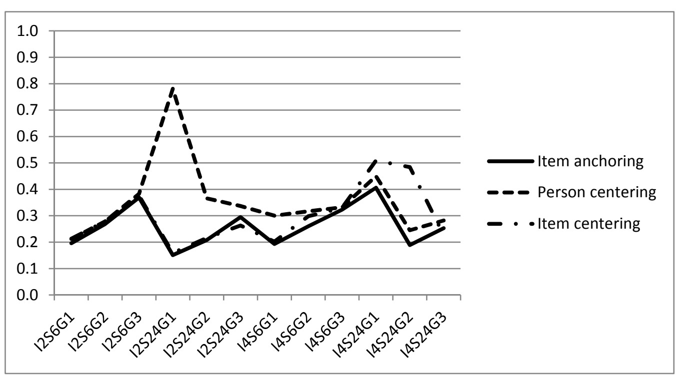


Figure 4.11: RMSEs for Item Discriminations for the Mix3PLM

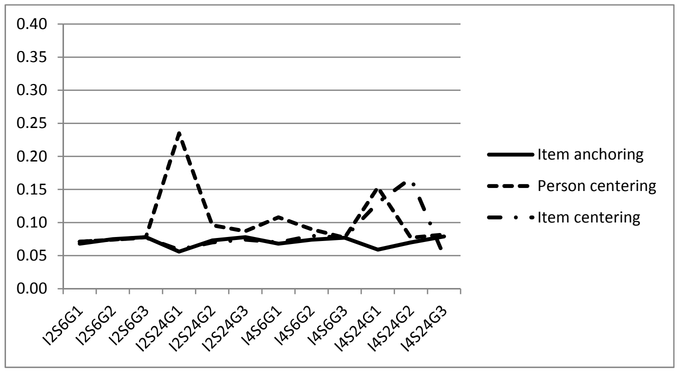


Figure 4.12: RMSEs for Item Lower Asymptotes for the Mix3PLM

In addition, Constraint 1 (item anchoring) had the highest RMSE values for the longer test length. There were no differences across the different constraints for the shorter test length and smaller sample size. Constraint 2 (person centering) had higher RMSEs for item difficulty and discrimination parameters on the 20-items × 2,400 students conditions for one group. All three constraints had high RMSE values for item difficulties and discriminations in the 40-items × 2,400 students conditions for one group. The 20-items × 2,400 students conditions for the three group model had higher RMSE values for item difficulties under Constraint 1 (item anchoring) and Constraint 2 (person centering). Constraint 2 (person centering) and Constraint 3 (item centering) had higher RMSE values for item difficulty in the 40-items × 2,400 students condition for the three-group model.

Correlations appeared to differ depending on constraint. For item difficulty, correlations for Constraint 1 (item centering) were higher, averaging .973. Those for Constraint 2 (person centering) were lower averaging .935, and those for Constraint 3 (item centering) were the

Table 4.5: BIAS, RMSE and Correlations of Item Parameters (β, a, c) and Latent Group Mean ($\bar{\theta}$) in Mix3PLM

Constraint	Item	Sample	Latent Classes	Item Difficulty (β) BIAS	RMSE	Cor.	Item Discrimination (a) BIAS	RMSE	Cor.	Lower Asymptote (c) BIAS	RMSE	Cor.	Latent Group Mean (θ) BIAS	RMSE
1. Item Anchoring	20	600	1	0.000	0.141	0.992	-0.032	0.196	0.917	0.025	0.068	**-0.177**	0.003	0.014
			2	0.000	**0.215**	0.982	-0.089	**0.269**	0.829	0.027	0.075	**-0.421**	0.002	0.020
			3	0.000	**0.290**	0.966	-0.174	**0.369**	**0.722**	0.028	0.078	**-0.489**	0.001	0.025
		2400	1	0.000	0.099	0.996	-0.012	0.151	0.957	0.017	0.056	**0.342**	-0.001	0.006
			2	0.000	0.168	0.989	-0.027	**0.208**	0.905	0.025	0.073	**-0.129**	-0.003	0.015
			3	0.000	**0.380**	0.942	-0.105	**0.295**	0.828	0.021	0.078	**-0.317**	0.003	0.035
	40	600	1	0.000	0.173	0.988	-0.026	0.193	0.915	0.025	0.068	**-0.057**	0.000	0.000
			2	0.000	**0.221**	0.981	-0.076	**0.260**	0.842	0.029	0.074	**-0.308**	-0.001	0.013
			3	0.000	**0.286**	0.967	-0.134	**0.323**	**0.775**	0.029	0.077	**-0.446**	0.005	0.038
		2400	1	0.000	**0.480**	0.908	-0.148	**0.406**	0.842	0.020	0.059	**0.244**	0.000	0.001
			2	0.000	0.167	0.989	-0.050	0.189	0.920	0.023	0.070	**-0.029**	-0.002	0.021
			3	0.000	**0.256**	0.974	-0.064	**0.253**	0.859	0.027	0.079	**-0.207**	-0.002	0.023
2. Person Centering	20	600	1	0.000	0.162	0.990	-0.016	**0.212**	0.899	0.028	0.071	**-0.128**	0.003	0.011
			2	0.000	**0.230**	0.979	-0.093	**0.279**	0.821	0.029	0.074	**-0.417**	0.001	0.008
			3	0.000	**0.322**	0.958	-0.187	**0.383**	**0.720**	0.028	0.078	**-0.501**	0.003	0.019
		2400	1	0.000	**0.828**	**0.724**	**-0.511**	**0.781**	**0.604**	0.075	**0.235**	**-0.347**	-0.001	0.004
			2	0.000	**0.439**	0.923	-0.096	**0.366**	0.829	0.029	0.096	**-0.144**	-0.002	0.010
			3	0.000	**0.349**	0.951	-0.127	**0.337**	**0.796**	0.027	0.087	**-0.280**	-0.005	0.038
	40	600	1	0.000	**0.335**	0.955	-0.134	**0.300**	0.883	0.040	0.108	**-0.279**	-0.001	0.003
			2	0.000	**0.324**	0.958	-0.108	**0.317**	0.814	0.034	0.090	**-0.319**	-0.002	0.013
			3	0.000	**0.294**	0.966	-0.138	**0.332**	**0.768**	0.030	0.077	**-0.445**	0.005	0.051
		2400	1	0.000	**0.508**	0.897	**-0.309**	**0.450**	0.888	0.046	0.153	**-0.266**	0.000	0.000
			2	0.000	**0.209**	0.983	-0.067	**0.245**	0.886	0.027	0.077	**-0.042**	-0.001	0.008
			3	0.000	**0.393**	0.939	-0.087	**0.282**	0.847	0.028	0.082	**-0.220**	0.000	0.021
3. Item Centering	20	600	1	0.000	0.162	0.990	-0.013	**0.213**	0.897	0.029	0.071	**-0.131**	0.002	0.011
			2	0.000	**0.227**	0.979	-0.092	**0.279**	0.821	0.028	0.074	**-0.429**	0.001	0.022
			3	0.000	**0.304**	0.963	-0.174	**0.376**	**0.721**	0.030	0.077	**-0.496**	0.001	0.027
		2400	1	0.000	0.118	0.995	-0.006	0.159	0.950	0.020	0.059	**0.308**	-0.001	0.004
			2	0.000	0.161	0.990	-0.014	**0.216**	0.898	0.025	0.070	**-0.106**	-0.002	0.010
			3	0.000	0.192	0.985	-0.070	**0.263**	0.848	0.026	0.074	**-0.302**	0.000	0.027
	40	600	1	0.000	0.183	0.987	-0.020	**0.202**	0.906	0.028	0.070	**-0.052**	-0.001	0.003
			2	0.000	**0.288**	0.967	-0.099	**0.298**	0.821	0.031	0.080	**-0.327**	-0.002	0.017
			3	0.000	**0.287**	0.967	-0.133	**0.330**	**0.768**	0.030	0.077	**-0.450**	0.006	0.049
		2400	1	0.000	**0.584**	0.865	**-0.405**	**0.509**	0.876	-0.002	0.130	**-0.388**	0.000	0.001
			2	0.000	**0.497**	0.902	-0.228	**0.484**	0.824	0.062	0.167	**-0.223**	-0.004	0.019
			3	0.000	**0.354**	0.100	-0.053	**0.227**	0.097	0.005	0.049	**-0.025**	0.002	0.039
Constraint 1				0.000	0.240	0.973	-0.078	0.259	0.859	0.025	0.071	-0.166	0.000	0.018
Constraint 2				0.000	0.366	0.935	-0.156	0.357	0.813	0.035	0.102	-0.282	0.000	0.016
Constraint 3				0.000	0.280	0.891	-0.109	0.296	0.786	0.026	0.083	-0.218	0.000	0.019
	20-items			0.000	0.266	0.961	-0.102	0.297	0.831	0.029	0.083	-0.231	0.000	0.017
	40-items			0.000	0.324	0.905	-0.127	0.311	0.807	0.028	0.088	-0.213	0.000	0.018
		n = 600		0.000	0.247	0.974	-0.097	0.285	0.824	0.029	0.077	-0.326	0.001	0.019
		n = 2400		0.000	0.343	0.892	-0.132	0.323	0.814	0.028	0.094	-0.118	-0.001	0.016
			1-class	0.000	0.314	0.941	-0.136	0.314	0.878	0.029	0.096	-0.078	0.000	0.005
			2-class	0.000	0.262	0.969	-0.087	0.284	0.851	0.031	0.085	-0.241	-0.001	0.015
			3-class	0.000	0.309	0.890	-0.121	0.314	0.729	0.026	0.076	-0.348	0.002	0.033

Note. When the absolute value of bias is larger than .3, RMSE is larger than .2, or correlation is less than .8, the values are bold.

lowest, averaging .891. For item discrimination, correlations for Constraint 1 were highest, averaging .859. Correlations for Constraint 2 were .813, and those for Constraint 3 were the lowest, averaging .786. Correlations for the lower asymptote were negative and low. The average correlation for Constraint 1 was -.166, -.282 for Constraint 2, and -.218 for Constraint 3.

Relationship between Model Selection and Recovery Analysis

The lack of correct model detections may be related to poor recovery results. Bias, RMSE, and Pearson correlation calculated for item recovery analysis are displayed in Tables 4.3 to 4.5. There was a high percent of correct model selection for the MixRM and Mix2PLM and no apparent affect on model selection for these models. However, the Mix3PLM had a low percent of correct model selection under Constraint 2 (person centering) and Constraint 3 (item centering) (see Table 4.1 and Figures 4.1 to 4.3). These results appear to be related to poor recovery of item parameters.

When Constraint 1 (item anchoring) was used with the Mix3PLM, BIAS and RMSE values were lower than when Constraints 2 (person centering) and 3 (item centering) were used. In addition, all correlations were higher with Constraints 2 and 3 (see Table 4.5). When the conditions having the higher RMSE values (i.e., larger than .5) for item difficulty are examined, all conditions had lower percent of correct model detection (e.g., RMSE = .828 and 45% of correct model selection) using BIC for condition with Constraint 2 (person centering), 20-items, 2,400 examinees, and 1-group model; RMSE = .584 and 25% of correct model selection for a Constraint 3 (item centering), 40-items, 2,400 examinees, and 1-group model.) It seems likely that the lack of correct model selection is related to poor recovery. The Mix3PLM having had the lowest percent of correct model selection and also had poorer recovery results.

Recovery of Group Membership

Recovery of group membership was examined by calculating the percentage of correct group identifications for each condition and comparing that with the percentage of examinees who were simulated as being in that latent group (see Table 4.6). Only correct model selections were considered for this comparison. If the correct model was indicated by the BIC index, this was considered an correct model selection for purposes of the recovery analysis.

The latent class group membership was recovered well in the MixRM and Mix2PLM. The percentage of correct group identifications ranged from 94% to 100% for the MixRM and 97% to 100% for the Mix2PLM. For the Mix3PLM, one condition under Constraint 2 (person centering) for the 20-items × 2,400 students for the two-group model had 84% correct identifications. All other conditions had correct identifications of 90% or higher (See Table 4.6).

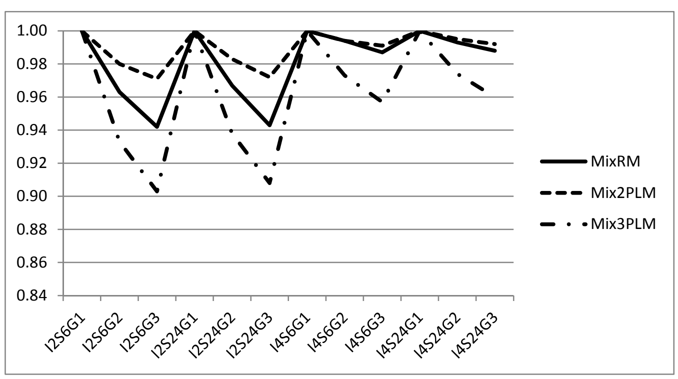


Figure 4.13: The Percentage of Correct Group Membership Identifications for Constraint 1 (Item Anchoring)

The percentages of correct group membership are plotted in Figures 4.13 to 4.15. There is a similar pattern in each of these plots. First, the Mix2PLM had the highest percentage

Table 4.6: Percent of Latent Class Membership Classification

Constraint	Item	Sample	Latent Classes	MixRM	Mix2PLM	Mix3PLM
1. Item Anchoring	20	600	1	100	100	100
			2	96	98	93
			3	94	97	90
		2400	1	100	100	100
			2	97	98	94
			3	94	97	91
	40	600	1	100	100	100
			2	99	99	97
			3	99	99	96
		2400	1	100	100	100
			2	99	100	97
			3	99	99	96
2. Person Centering	20	600	1	100	100	100
			2	96	98	93
			3	94	97	90
		2400	1	100	100	100
			2	97	98	**84**
			3	94	97	90
	40	600	1	100	100	100
			2	99	99	97
			3	99	99	96
		2400	1	100	100	100
			2	99	100	98
			3	99	99	96
3. Item Centering	20	600	1	100	100	100
			2	96	98	93
			3	94	97	90
		2400	1	100	100	100
			2	97	98	94
			3	94	97	91
	40	600	1	100	100	100
			2	99	99	97
			3	99	99	96
		2400	1	100	100	100
			2	99	100	97
			3	99	99	96
MixIRTM				98	99	96
Constraint 1				98	99	96
Constraint 2				98	99	95
Constraint 3				98	99	96
	20-items			97	98	94
	40-items			99	100	98
		$n = 600$		98	99	96
		$n = 2400$		98	99	96
			1-class	100	100	100
			2-class	98	99	95
			3-class	96	98	93

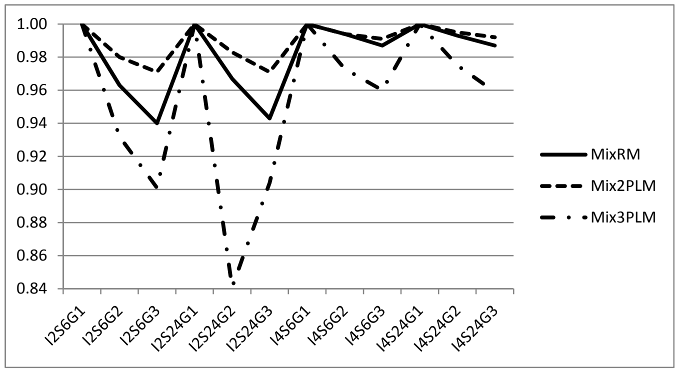


Figure 4.14: The Percentage of Correct Group Membership Identifications for Constraint 2 (Person Centering)

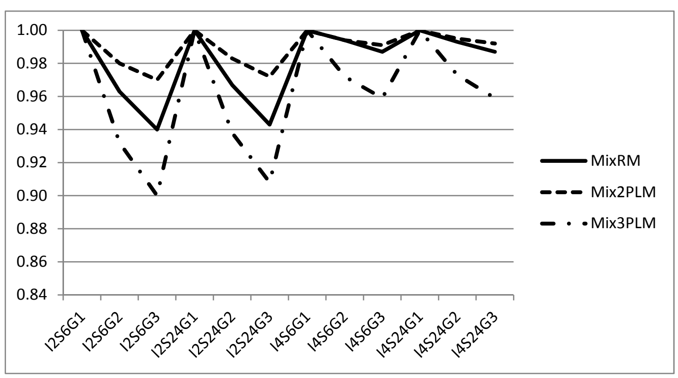


Figure 4.15: The Percentage of Correct Group Membership Identifications for Constraint 3 (Item Centering)

of recovery of correct group membership and the Mix3PLM had the lowest. The Mix2PLM appeared to recover latent group membership better than the MixRM and Mix3PLM. Second, the percentage of correct membership identifications decreased as the number of latent groups increased, that is, as the models became more complex. Third, higher percentages of correct group membership classifications were observed for the longer test length. Finally, sample size and type of constraint did not appear to affect group membership identification. These patterns are similar to those reported by Li et al. (2009).

Effects of Simulation Conditions on Recovery of Generating Parameters

Multiple regression was used to help determine the effects of the simulation conditions on the recovery of generating parameters. The simulation conditions were used as independent variables in each multiple regression analysis for bias, RMSE, and correlation. Since BIAS values for item difficulty for MixRM, Mix2PLM, and Mix3PLM were all zero (see Tables 4.3 to 4.5), BIAS was excluded from the multiple regression analysis for recovery of item difficulty. The percent of correct model selection using AIC and BIC was used for analysis of recovery of group membership.

Table 4.7 provides descriptive statistics for the dependent variables in the multiple regression analysis. Model selection using AIC and BIC and latent group membership identification were successful with higher than 92 percentage. The recovery analysis for item difficulty and discrimination parameters were performed well. All results of recovery analysis for item difficulty were lower than those for item discrimination. It may be concluded, therefore, that item difficulty parameters were recovered better than item discrimination parameters. Based on the bias and RMSE, the lower asymptote seemed to be recovered well. However, the correlations for lower asymptote were all negative. The negative correlations may be due to the very narrow range of lower asymptote values. That is, anytime the estimate of the lower asymptote was lower than the generating parameter, the correlation would be negative.

Table 4.7: Summary of Descriptive Statistics of Dependent Variables

	n	M	SD	Min	Max
Model selection					
AIC	108	92.82	14.41	25.00	100.00
BIC	108	94.07	17.67	25.00	100.00
Item difficulty					
RMSE	108	0.18	0.13	0.05	0.83
Correlation	108	0.97	0.09	0.10	1.00
Item discrimination					
Bias	72	0.07	0.09	0.00	0.51
RMSE	72	0.24	0.11	0.08	0.78
Correlation	72	0.87	0.12	0.10	0.99
Lower asymptote					
Bias	36	0.03	0.01	0.00	0.08
RMSE	36	0.09	0.04	0.05	0.24
Correlation	36	-0.22	0.22	-0.50	0.34
Correct group membership identification					
Percent in group	108	97.68	2.97	84.10	100.00

The latent group mean ($\bar{\theta}$) was not considered in the multiple regression analysis as all latent group means were very close to zero. The multiple regression analysis was only performed for bias and RMSE of the latent group means. All the regression coefficients were close to zero. Therefore, the recovery analysis using bias and RMSE of latent group means was not involved in the regression analysis.

Variable Recoding

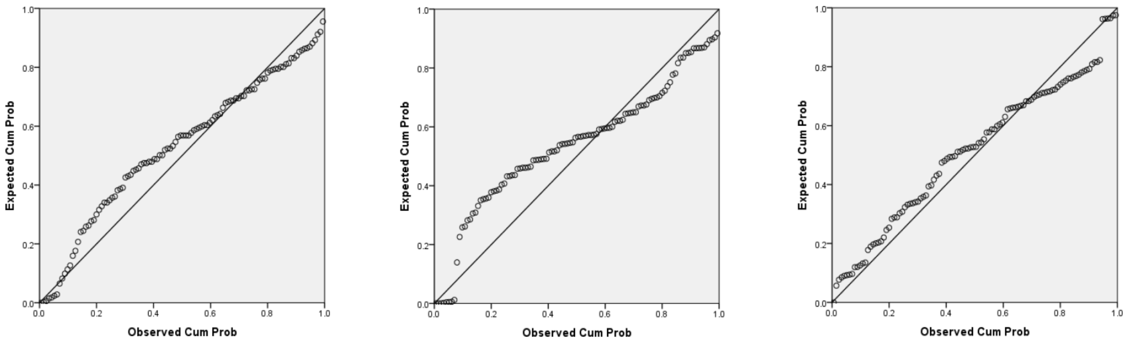
The bias values for item discrimination and lower asymptote were recoded to absolute values for ease of interpretation. Actual results as indicated in Table 4.1 were used for analyses. All independent variables involved in the multiple regression were recoded using the following dummy coding scheme. Two dummy codes were used for MixIRTMs. For one dummy code (MixIRTM_2PL), the Mix2PLM was coded to one and the MixRM and Mix3PLM were both coded as zero. For the other dummy code (MixIRTM_3PL), the Mix3PLM was coded as one and the MixRM and Mix2PLM were coded as zero. The type of identification constraints were also recoded using two dummy codes: Constraint 2 (person centering) was coded as one and Constraints 1 (item anchoring) and 3 (item centering) were coded as zero for the 'Person Centering' variable. For the other dummy code (Item Centering), Constraint 3 (item centering) was coded as one and Constraints 1 (item anchoring) and 2 (person centering) were coded as zero.

The 20-item test length coded as zero and the 40-item test length was coded as one. The small sample size (i.e., n = 600) was coded as zero and the large sample size (i.e., n = 2,400) was coded as 1. The number of latent classes was recoded to 0, 1, and 2 for the one-, two-, and three-group MixIRTMs, respectively. Using this coding scheme, the intercept can be interpreted as the expected mean of a dependent variable when for the MixRM × Constraint 1 (item anchoring) × 20-items × $n = 600$ × one-group MixIRTM condition. The total sample size for the regression analysis is 108, which is the total number of simulation conditions.

Regression Assumption

There are three assumptions for a multiple regression analysis. First, a linear relationship is assumed between independent and dependent variables. As previously noted, the latent group mean ($\bar{\theta}$) had no linear relationship with any simulation condition (i.e., with any independent variable). Therefore, the recovery analyses using BIAS and RMSE of latent group means were not included in the regression analysis.

The two assumptions are related to residuals or error terms. The expected residuals are assumed to be independent of each other. The expected value of residuals is zero and the residuals are assumed to be normally distributed. To check the assumption of the normality, p-p plots were constructed for standardized residuals from the regression analysis.



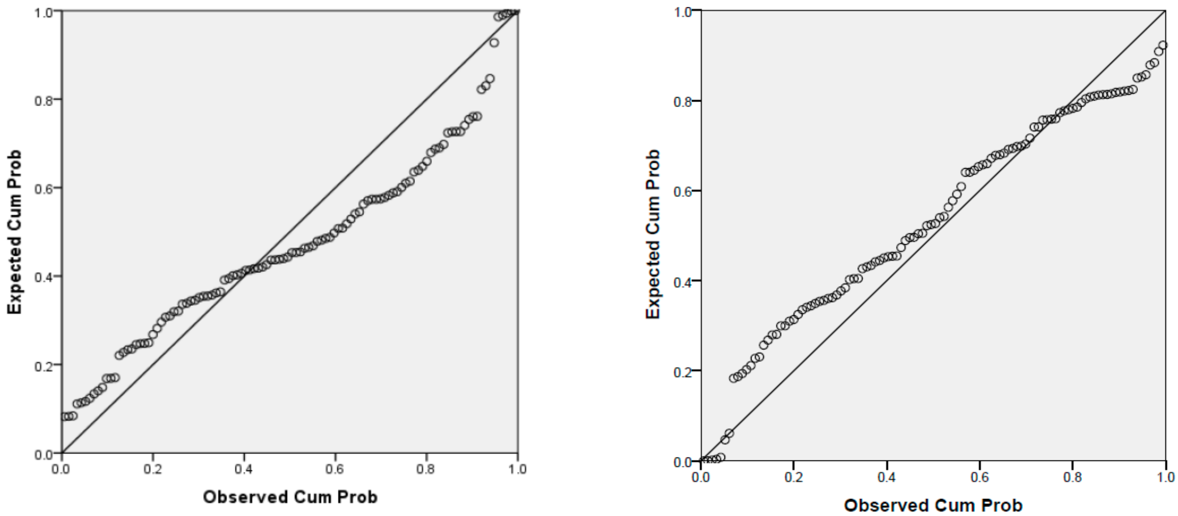
(a) AIC (b) BIC (c) Percent in Group

Figure 4.16: P-P Plots of Standardized Residuals for Model Selection and Percent Correct Classifications

These plots indicate deviations from the line indicating the normality expectation for model selection and percent correctly classified (see Figures 4.16 to 4.19). The criteria for normal distribution, in other words, is the degree to which the p-p plot for the actual values is close to the diagonal solid line, which indicates the expected values. For correlations of item parameters, Fisher's logarithmic transformation was applied to obtain better normality. Examination of the p-p plots of residuals of dependent variables, AIC, BIC, and Percent in group, for model selection and group membership identification, suggest the residuals were relatively close to the expected values for AIC and percent in group (see Figure 4.16). Bias, RMSE, and correlation results for item parameters indicated moderate normality (see Figures 4.17 to 4.19).

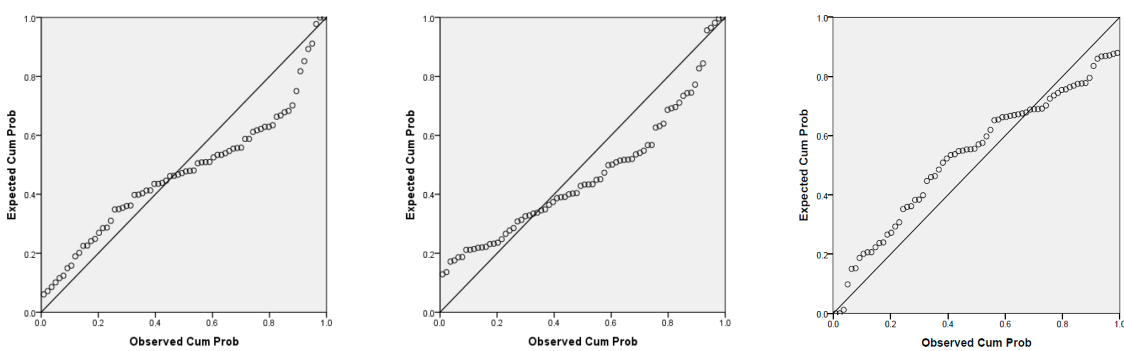
Multicollinearity Diagnostics

Multicollinearity refers to the degree to which the independent variables in the regression analysis are correlated. Complete collinearity would mean that the correlation between independent variables equals one. Collinearity diagnostics measure how much the independent



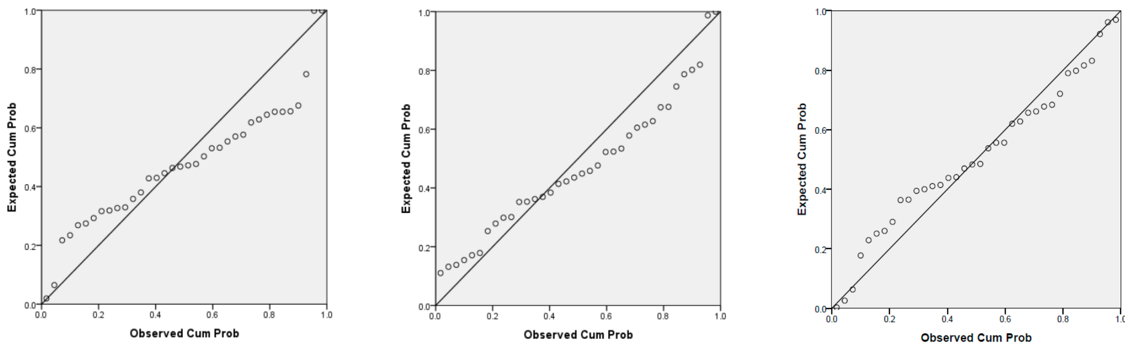
(a) RMSE (b) Correlation

Figure 4.17: P-P Plots of Standardized Residuals for Item Difficulty



(a) BIAS (b) RMSE (c) Correlation

Figure 4.18: P-P Plots of Standardized Residuals for Item Discrimination



(a) BIAS (b) RMSE (c) Correlation

Figure 4.19: P-P Plots of the Standardized Residuals for Lower Asymptote Item Parameters

variables are each related to the other independent variables and how they affect the stability and variance of the regression estimates (Belsley, Kuh, & Welsch, 2004; Pedhazur, 1997). One simple collinearity diagnostic measure is the correlation between independent variables. In this study, all independent variables were categorical and explained the simulation conditions. There was no correlation simulated among the conditions of the simulation study as the simulation conditions were designed to be independent. In addition, Phi and Cramer's V correlation statistics for categorical variables could not be computed because the asymptotic standard error equaled zero. Therefore, multicollinearity was not a concern in this regression analysis.

The Effect of Simulation Conditions on Model Selection and Group Membership Identification

Three multiple regression analyses were performed to examine the effects of simulation conditions on model selection using AIC and BIC and on group membership identification (see Table 4.8). The notation 'B' is used to indicate that the coefficient is unstandardized. Similarly, SE B indicates the standard error of unstandardized coefficient. β represents the standardized coefficient. The results of the regression analysis indicated the seven independent variables explained 31% of the variance of model selection results using AIC (R^2 = .31, $F(7, 100)$ = 6.48, p < .01), 29% of the variance of model selection results using BIC (R^2 = .29, $F(7, 100)$ = 5.86, p < .001), and 67% of the variance of latent group membership identification (R^2 = .67, $F(7, 100)$ = 29.56, p < .001).

Three simulation conditions, MixIRTM, identification constraint, and sample size, significantly predicted model selection using AIC. The Mix2PLM had a higher percent of correct model selections than either the MixRM and or the Mix3PLM (B = 7.92, p < .01). In addition, for Constraint 2 (person centering), AIC had poorer correct model selection (B = -6.67, p < .05) than for Constraints 1 (item anchoring) or 3 (item centering). AIC also had lower correct percentages as sample size increased (B = -10.83, p < .001).

For BIC, two simulation conditions, MixIRTM and constraint, significantly affected model selection. The Mix3PLM had a lower percentage of correct model selection than either the MixRM or the Mix2PLM (B = -17.50, p < .001). Constraint 2 (person centering) and Constraint 3 (item centering) had lower percentages of model selection than Constraint 1 (item anchoring) (B = -7.36, p < .05 for Constraint 2 and B = -8.33, p < .05 for Constraint 3).

Three conditions, MixIRTM, test length, and number of latent classes, significantly predicted the percent of correct group membership identification. The Mix3PLM was less effective at detecting correct group membership and the Mix2PLM was the best among the three MixIRTMs (B = .85, p < .05 for Mix2PLM, B = -2.20, p < .001 for Mix3PLM). The more items and the smaller number of latent classes appeared to get higher percent of correct group membership (B = 2.41, p < .001 for item, B = -2.04, p < .001 for latent classes).

Table 4.8: Regression Analysis Results for Model Selection and Correct Group Membership Identification

	AIC			BIC			Identification of Correct Group		
Variable	B	SE B	β	B	SE B	β	B	SE B	β
Constant	96.62	3.47		109.84	4.32		99.09	0.49	
MixIRTM_2PL	7.92	2.91	0.26**	0.14	3.63	0.00	0.85	0.41	0.14*
MixIRTM_3PL	-3.19	2.91	-0.11	-17.50	3.63	-0.47***	-2.20	0.41	-0.35***
Person Centering	-6.67	2.91	-0.22*	-7.36	3.63	-0.20*	-0.29	0.41	-0.05
Item Centering	-3.61	2.91	-0.12	-8.33	3.63	-0.22*	-0.03	0.41	0.00
Test Length	4.17	2.38	0.15	-4.07	2.96	-0.12	2.41	0.34	0.41***
Sample Size	-10.83	2.38	0.38***	-1.11	2.96	-0.03	-0.04	0.34	-0.01
Latent Classes	1.39	1.46	0.08	-2.15	1.81	-0.10	-2.04	0.21	-0.56***
R^2		0.31			0.29			0.67	
$F(7, 100)$		6.48**			5.86***			29.56***	

Note. $N = 108$, *$p < .05$, **$p < .01$, ***$p < .001$

The Effect of Conditions on the Recovery Analysis of Item Parameters

Eight multiple regression analyses were performed to help determine whether simulation conditions significantly affected the recovery of item parameters (See Tables 4.9 to 4.11). As mentioned previously, analysis of bias for item difficulty was excluded because these values were all zero in all conditions.

Results of the regression analysis for the other variables indicated the seven predictors explained 49% of the variance of RMSE of item difficulty ($R^2 = .49$, $F(7,100) = 13.57$, $p < .001$) and 67% of the variance of correlation of item difficulty ($R^2 = .67$, $F(7,100) = 28.35$, $p < .001$). For item discrimination, the six predictors explained 31% of the variance of bias ($R^2 = .31$, $F(6,65) = 4.67$, $p < .01$), 39% of the variance of RMSE ($R^2 = .39$, $F(6,65) = 6.80$, $p < .001$), and 72% of the variance of correlation ($R^2 = .72$, $F(6,65) = 28.26$, $p < .001$). The F-test for bias and RMSE (see Table 4.11) suggested that bias and RMSE for the lower asymptote parameters could not be predicted by simulation conditions ($F(5,30) = 1.12$, $p = .37$ for Bias and $F(5,30) = 2.08$, $p = .10$ for RMSE). The results of the regression analysis indicated that the five predictors explained 58% of the variance of the correlation for the lower asymptote parameters ($R^2 = .58$, $F(5,30) = 8.15$, $p < .001$)

Table 4.9: Regression Analysis for Item Difficulty Parameters

	RMSE			Correlation		
Variable	B	SE B	β	B	SE B	β
Constant	0.07	0.03		3.40	0.12	
MixIRTM_2PL	0.01	0.02	0.06	-0.13	0.10	-0.09
MixIRTM_3PL	0.18	0.02	0.67***	-1.00	0.10	-0.70***
Person Centering	0.05	0.02	0.18*	-0.17	0.10	-0.12
Item Centering	0.02	0.02	0.07	-0.12	0.10	-0.08
Test Length	0.02	0.02	0.07	-0.12	0.08	-0.09
Sample Size	-0.02	0.02	-0.06	0.34	0.08	0.25***
Latent Classes	0.03	0.01	0.19*	-0.33	0.05	-0.39***
R^2		0.49			0.67	
$F(7,100)$		13.57***			28.35***	

Note. $N = 108$, * $p < .05$, ** $p < .01$, *** $p < .001$

RMSEs for the Mix3PLM for item difficulty appeared to be higher (B = .18, p < .001) and correlations appeared to be lower (B = -1.00, p < .001) than those for the MixRM and Mix2PLM. RMSEs for item difficulty tended to be higher for Constraint 2 (person centering) and for models with more latent classes and with larger sample size (B = .05, p < .05 for Constraint 2 (person centering), B = .03, p < .05 for number of latent classes, B = .25, p < .05 for number of latent classes). The correlations for item difficulty tended to be lower for smaller sample size and for more latent classes (B = .34, p < .001 for sample size and B = -.33, p < .001 for latent classes) (See Table 4.9).

For the recovery of item discrimination, the Mix3PLM had higher bias (B = .09, p < .001) and RMSEs (B = .12, p < .001) and lower correlations (B = -.44, p < .001) than for the MixRM or Mix2PLM. Bias for item discrimination also was higher (B =.04, p < .05) with Constraint 2 (person centering) and RMSEs for item discrimination was higher with more latent classes (B =.03, p < .05). Lower correlation was observed with smaller sample size and with more latent classes (B = .32, p < .001 for sample size and B = -.30, p < .001 for latent class) (see Table 4.10).

For larger sample size and smaller numbers of latent classes, correlations for the lower asymptote were higher (B = .23, p < .001 for sample size and B = -.15, p < .001 for latent classes) (See Table 4.11).

Table 4.10: Regression Analysis for Item Discrimination Parameters

	BIAS			RMSE			Correlation		
Variable	B	SE B	β	B	SE B	β	B	SE B	β
Constant	0.00	0.02		0.14	0.03		1.89	0.08	
MixIRT_3PL	0.09	0.02	0.50^{***}	0.12	0.02	0.54^{***}	-0.44	0.06	-0.51^{***}
Person Centering	0.04	0.02	0.21^{*}	0.05	0.03	0.22	-0.11	0.07	-0.12
Item Centering	0.02	0.02	0.08	0.02	0.03	0.10	-0.10	0.07	-0.11
Test Length	0.00	0.02	0.02	0.00	0.02	-0.01	0.00	0.06	0.00
Sample Size	0.02	0.02	0.12	-0.03	0.02	-0.11	0.32	0.06	0.36^{***}
Latent Classes	-0.01	0.01	-0.06	0.03	0.01	0.20^{*}	-0.30	0.04	-0.56^{***}
R^2		0.31			0.39			0.72	
$F(6, 65)$		4.67^{**}			6.80^{***}			28.26^{***}	

Note. $N = 72$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table 4.11: Regression Analysis for Lower Asymptote Item Parameters

	BIAS			RMSE			Correlation		
Variable	B	SE B	β	B	SE B	β	B	SE B	β
Constant	0.03	0.01		0.07	0.01		-0.15	0.07	
Person Centering	0.01	0.01	0.40^{*}	0.03	0.01	0.42^{*}	-0.12	0.07	-0.25
Item Centering	0.00	0.01	0.06	0.01	0.01	0.16	-0.06	0.07	-0.12
Test Length	0.00	0.00	0.00	0.01	0.01	0.07	0.02	0.05	0.05
Sample Size	0.00	0.00	-0.05	0.02	0.01	0.25	0.23	0.05	0.50^{***}
Latent Classes	0.00	0.00	-0.13	-0.01	0.01	-0.23	-0.15	0.03	-0.53^{***}
R^2		0.16			0.26			0.58	
$F(5,30)$		1.12			2.08			8.15^{***}	

Note. $N = 36$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Chapter 5

Discussion and Conclusions

The purpose of this study was to explore the effects of three scale identification methods for establishing a common metric between latent classes in MixIRT models. There are currently three methods commonly used for developing a common metric between latent classes. The first method is concurrent calibration in which one or more items are used to anchor the metrics between classes. The second method is to impose equality constraints by setting the mean of one latent class to zero and its standard deviation to one and a third method is setting the sum of item difficulties to zero for each latent group.

An empirical example from the TIMSS 2011 Grade 8 Science Test motivated the simulation study. The three constraints were applied to establish a common metric between latent classes for a MixRM and a Mix2PLM. Results from the MixRM suggested that each of the constraints had a somewhat different effect on item difficulty estimates, ability estimates, numbers of latent classes, classifications of examinees into latent classes, and proportions of membership in each latent class. Similar results were observed for the Mix2PLM with the exception that the same number of latent classes was extracted using all three constraints. Results based on this data set clearly differed depending on the MixIRT model and on the constraint used.

5.1 Summary of Simulation Study Results

A simulation study was designed to explore the impact of different identification constraints in the context of three MixIRT models: MixRM, Mix2PLM, and Mix3PLM. The design of the simulation study includes the three types of constraints used in the example described above: Constraint 1 (item anchoring: equating method), Constraint 2 (person centering: the mean ability of the first latent group set to zero), and Constraint 3 (item centering: the mean of item difficulties is set to zero in each class). The simulation conditions also included the following conditions: two sample sizes (600 examinees and 2,400 examinees), two test lengths (20 and 40 items), three different cases of latent groups (1-, 2-, and 3-groups) with different proportions of simulated groups for each of these three MixIRT models. There were 20 replications of the 108 conditions: Three identification constraints $\times$ two test lengths $\times$ two sample sizes $\times$ one to three latent groups $\times$ three mixture IRT models yields 108 conditions.

Three convergence diagnostics were used to monitor the convergence: Heidelberger and Welch convergence diagnostics, the ratio of the standard deviation of the parameter estimate to the MC standard error for the parameter estimate, and the credibility interval. THe Heidelberger and Welch convergence diagnostics and the ratio of the standard deviation of the parameter to the MC standard error for the parameter were used to check the convergence for MixRM and Mix2PLM. The 95% credibility interval was used to monitor the convergence of the Mix3PLM. The burn-in for the MixRM was 5,000 iterations and the post-burn-in was 5,000 iterations. For the Mix2PLM, the burn-in was 6,000 iterations and the post-burn-in was 11,000 iterations. The Mix3PLM was considered to have converged after 6,000 burn-in iterations and 11,000 post-burn-in iterations based on the 95 % credibility interval.

Exploratory MixRM, Mix2PLM, and Mix3PLM analyses were done to determine the best fitting model to the simulated data. The criterion used for model selection was BIC as suggested by Li et al. (2009). AIC was provided as a comparison index. Model selection for the MixRM and Mix2PLM using BIC was almost 100 percent correct. For the Mix3PLM, however, there were some low percentages of correct model selection. These occurred only under Constraints 2 (person centering) and 3 (item centering). Model selection under Constraint 1 (item anchoring) was better for the Mix3PLM than under Constraints 2 or 3. There was lack of correct model detection in the Mix3PLM with Constraint 2 (person centering) when the smaller sample size and 3-class model were simulated and when the larger sample size and 1-class model were simulated. Under Constraint 3 (item centering), low percentages of correct detections occurred for the longer test $\times$ larger sample size and for the longer test $\times$ smaller sample size $\times$ 3-class.

BIC had more correct model selections than AIC for the MixRM and Mix2PLM. For the Mix3PLM, BIC had equal or more correct model selections than AIC except in three conditions: the longer test length $\times$ smaller sample size $\times$ 3-class model using Constraints 2 and 3, and the longer test length $\times$ larger sample size $\times$ 3-class using Constraint 3. This agrees with results from Li et al. (2009) which found that BIC made more correct model selections for all three MixIRTMs.

A recovery analysis was performed to evaluate how the estimation algorithms based on different constraints affect recovery of the generating parameters from the simulated data. The item and latent group mean generating parameters were compared to the item and latent group mean estimates. BIAS, RMSE, and Pearson correlations between the generating parameters and estimated parameters were calculated. Before recovery analysis, all estimated parameters were placed onto the metric of the generating parameters using the mean and sigma equating method. BIAS and RMSE for latent group mean parameter estimates were close to zero for all conditions under all MixIRTMs. However, there were variations in recovery results for the different MixIRT models.

For the MixRM, all BIAS and correlations appeared to indicate that generating parameters were recovered well. RMSE results suggested recovery was negatively affected by the small sample size (i.e., n = 600) as the number of latent classes increased. The type of identification constraint did not appear to affect recovery of item difficulties as indexed by RMSEs. Recovery of the generating item and group mean parameters was generally good in the MixRM.

Recovery of item difficulty for the Mix2PLM was generally good with the possible exceptions of the 2-class and 3-class models for the small sample conditions. The type of constraint did not appear to affect recovery of item difficulty for this model. Bias statistics for item discrimination parameters were relatively small. However, RMSEs for item discrimination parameters appeared to be higher for the small sample size and 2- and 3-class conditions. Correlations for item discrimination generally moderately high to high. These results suggest sample size and number of latent classes effects on recovery of item discrimination.

Mix3PLM were recovered moderately well for Constraint 1 (item anchoring). Constraints 2 (person centering) and 3 (item centering) had worse item recovery than Constraint 1. Constraint 1 (item anchoring) had the highest RMSE values for the longer test length. Constraint 2 (person centering) had higher RMSEs for item difficulty and discrimination parameters on the 20-items × 2,400 students conditions for one group. All three constraint had high RMSE values for item difficulties and discriminations in the 40-items × 2,400 students conditions for the 1-group model. The 20-items × 2,400 students conditions for the 3-group model had higher RMSE values for item difficulties under Constraint 1 (item anchoring) and Constraint 2 (person centering). Constraint 2 (person centering) and Constraint 3 (item centering) had higher RMSE values for item difficulty in the 40-items × 2,400 students condition for the 3-group model. The RMSEs for the lower asymptote parameters clearly indicated that all conditions had lower RMSEs except for the 20-items × 2,400 students 1-group model condition for Constraint 2 (person centering). Correlations appeared to differ depending on constraint. Correlations were highest for all three item parameters for Constraint 1 (item centering). The correlations for item difficulty and item discrimination were moderately high or high for all constraints. The correlations for the lower asymptote were negative.

There was no constraint effect on the model selection and recovery analysis under the MixRM and Mix2PLM. For the Mix3PLM, Constraint 1 (item anchoring) selected the correct model well using both BIC (98.33%) and AIC (98.33%). Recovery appeared best for Constraint 1. The correct model was selected more with BIC under Constraint 2 (person centering) than Constraint 3 (item centering) but recovery of generating parameters was worse for these two constraints. Results suggest that any of the three constraints can be used for the MixRM and Mix2PLM but only Constraint 1 is recommended for the Mix3PLM.

The lack of correct model selection appeared to be related to poor recovery results. The Mix3PLM had a lower percent of correct model selection and also had poorer recovery results. Recovery of group membership was examined by calculating the percentage of correct group identifications for each condition and comparing that with the percentage of examinees who were simulated as being in that latent group. The latent class group membership was recovered well in the MixRM and Mix2PLM. For the Mix3PLM, one condition under Constraint 2 (person centering) for the 20-items $\times$ 2,400 students for the two-group model had 84% correct identifications. All other conditions had correct identifications of 90% or higher (see Table 4.6).

Multiple regression was used to help detect the effects of the simulation conditions on the recovery of generating parameters. The simulation conditions were used as independent variables in each multiple regression analysis for BIAS, RMSE, and correlation. The percent of correct model selection using AIC and BIC and the frequency of correct group membership identification were also used as independent variables for analysis of recovery of group membership. Table 5.1 summarized Tables 4.8 to 4.11. For easier understanding, all the regression coefficients are recoded in these tables by '+' for good recovery and '-' for poor recovery. When the regression coefficients for BIAS and RMSE were significantly negative,

Table 5.1: Summary of Effects of Simulation Conditions on Recovery

	Mixture 2PL Model	Mixture 3PL Model	Person Centering	Item Centering	Test Length	Sample Size	Latent Class
Model Selection							
AIC	+		−			−	
BIC		−	−	−			
Percent in Group	+	−			+		−
Item Difficulty							
RMSE		−	−				−
Correlation		−				+	−
Item Discrimination							
Bias	NA	−	−				
RMSE	NA	−					−
Correlation	NA	−				+	−
Lower Asymptote		NA					
Bias	NA	NA	−				
RMSE	NA	NA	−				
Correlation	NA	NA				+	−

Note. + indicates good recovery, − is bad recovery.

they were recoded as '+' because smaller bias and RMSE values means better recovery. The other dependent variables (i.e., AIC, BIC, percent in group, correlation for item parameters) were recoded by '+' when the coefficients were positive and by '-' when they were negative.

When the MixIRTMs were compared, the Mix2PLM had the best recovery and the Mix3PLM the worst. For the Mix3PLM, all signs were negative, indicating the Mix3PLM did not recover the generated parameters well. When the type of constraints were compared by using the 'Person Centering' and 'Item Centering' variables, Constraint 2 (person centering) had worse recovery results. Test length did not appear to have an effect on recovery for item parameters. The longer test length was associated with improved correct identification of group membership. The more sample size appeared to have better recovery results using correlation for item parameters. Percent of correct model selection using AIC was lower when larger the sample size was used. The more latent classes in the model, the poorer the recovery of group membership, item difficulty, item discrimination, and lower asymptote parameters.

5.2 Limitations and Future Studies

The Number of Replications. Although the results of Harwell et al. (1996) recommended a minimum of 25 replications for monte carlo studies in IRT-based research, 20 replications were performed in this study due to resource constraints on the campus computing cluster. At least 10 more replications would be preferable to ensure that results are somewhat more stable.

Time-Consuming Estimation using OpenBUGS. As mentioned in Table 3.5, the MixRM took up to 26 hours to run for one condition using OpenBUGS, 50 hours for the Mix2PLM, and 197 hours (about 8 days) for the Mix3PLM. For the 2,160 jobs run using OpenBUGS (= 20 replications of the 108 conditions), simulations for this research required approximately five months was taken using the GCA blades and the campus linux cluster computer.

Convergence Issue in Mix3PLM. Heidelberger and Welch diagnostic reported non-convergence for 35,000 iterations in the Mix3PLM. Instead of Heidelberger and Welch diagnostic, the 95 %credibility interval was used to monitor convergence of the Mix3PLM. Although the Mix3PLM was considered to have converged based on the 95 % credibility interval, more convergence diagnostic methods need to be examined with this model.

How to Determine Anchor Items. Among the three types of constraint, Constraint 1 (item anchoring) performed best. There is one issue when item anchoring constraint is used for metric identification with MixIRTMs. This is how to determine anchor items. This is also an important problem in IRT, particularly in studies of differential item functioning and equating. There are two methods to determine the anchor items in the MixIRTMs: deciding on the basis of theoretical background reasons which items to select and deciding based on statistical evidence. Bolt et al. (2002) studied test speededness using the MixRM. To test speededness, the first 18 of 32 items were assumed to be non-speeded and so were identified as not being speeded. These 18 items were used as anchor items. The last 8 items on the 32-item test were considered as potentially speeded items. Choi et al. (2014) used a statistical criterion based on the likelihood ratio test with the Mix3PLM to determine the anchor items. The likelihood ratio test was conducted using the MULTILOG computer program (Thissen et al., 2003). Additional methods to determine anchor items could profitably be explored.

Model Selection using BIC. Li et al. (2009) concluded that BIC was most effective among the AIC, DIC, PPMC, and PsBF. The MixIRTMs used in that simulation study did not contain any metric identification constraints. When three different types of constraint were involved in the present study, there was no problem for either the MixRM or the Mix2PLM. However, the Mix3PLM did not do as well with respect to selection of the correct model. This was particularly the case, when person centering and item centering were used.

Results from this study extend those by Li et al. (2009) by showing the differential effects of identification constraints for each of the MixIRT models considered. Comparison using different information indices might be useful for further extending this kind of research.

Bibliography

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Caski (Eds.), Proceedings of the 2nd International Symposium on Information Theory (pp. 267-281), Akademia Kiado, Budapest.

[2] Baker, F. B. & Kim, S.-H. (2004). *Item response theory: Parameter estimation techniques*. New York, NY: Marcel Dekker.

[3] Belsley, D. A., Kuh, E., & Welsch, R. E. (2004) *Regression diagnostics: Identifying influential data and sources of collinearity*. New York: John Wiley and Sons.

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

[5] 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.

[6] Carlin, B. P. & Louis, T. A. (2001). *Bayes and emperical bayes methods for data analysis (2nd ed.)*. New York, NY: Chapman & Hall.

[7] Celeux, G., Hurn, M., & Robert, C. P. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association, 95*, 957-970.

[8] Cho, S.-J., & Cohen, A. S. (2010). A multilevel mixture IRT model with applications to DIF. *Journal of Educational and Behavioral Statistics, 35*, 336-370.

[9] Cho, S.-J., Cohen, A. S., & Kim, S.-H. (2006, June). *An investigation of priors on the probabilities of mixtures in the mixture Rasch model*. Paper presented at the annual meeting of the Psychometric Society, Montreal, Canada.

[10] Cho, S.-J., Cohen, A. S., & Kim, S.-H. (2013). Markov Chain Monte Carlo estimation of a mixture item response theory model. *Journal of Statistical Computation and Simulation, 83,* 278-306.

[11] Cho, S.-J., Cohen, A. S., & Kim, S.-H. (in press). A mixture group bi-factor model for binary responses. *Structural Equation Modeling.*

[12] Cho, S.-J., Cohen, A. S., & Templin, J. (2008, March). *A multidimensional mixture IRT model for DIF analysis*. Paper presented at the annual meeting of the National Council on Measurement in Education, New York, NY.

[13] Choi, Y.-J., Alexeev, N., Cohen, A. S., & Kim, S.-H. (2010, July). Identifying a Mixture 3-Parameter Model. Paper presented at the International Meeting of the Psychometric Society: The 75th annual meeting of the Psychometric Society, Athens, GA

[14] Choi, Y.-J., Alexeev, N., & Cohen, A. S. (2014). DIF analysis using a mixture 3PL Model with a Covariate on the TIMSS 2007 mathematics test. *KAERA Research Forum 1*(1), 4-14.

[15] Choi, Y.-J., Cohen, A. S., Lu, Z., & Kim, S. (2014, April). The impact of equating on detection of treatment effects. Paper presented at the annual meeting of the National Council on Measurement in Education, Philadelphia, PA.

[16] Claeskens, G. & Hjort, N.L. (2008). *Model Selection and Model Averaging*, Cambridge, UK: Cambridge University Press.

[17] Cohen, A. S., Wollack, J. A., Bolt, D. M., & Mroch, A. A. (2002, April). *A mixture Rasch model analysis of test speededness.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

[18] Congdon, P. (2003). *Applied Bayesian modelling.* New York: Wiley.

[19] Dai, Y., & Mislevy, R. (2006, April). *Using structured mixture IRT models to study differentiating item functioning.* Paper presented at the annual meeting of the National Council on Measurement in Education. San Francisco, CA.

[20] de Ayala, R. J. (2009). *The theory and practice of item response theory.* New York, NY : Guilford Press.

[21] De Boeck, P., Cho, S.-J., & Wilson, M. (2011). Explanatory secondary dimension modelling of latent DIF. *Applied Psychological Measurement, 35*, 583-603.

[22] du Toit, M. (2003). *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT.* Lincolnwood, IL: Scientific Software International.

[23] Finch, W. H. & Pierson, E. E. (2011) A mixture IRT analysis of risky youth behavior. *Front. Psychology 2.* doi: 10.3389/fpsyg.2011.00098

[24] Geisser, S., & Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association, 74*, 153-160.

[25] Gelfand, A. E., & Dey, D. K. (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society, Series B, 56*, 501-514.

[26] Gelman, A. E., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2003), *Bayesian Data Analysis*(2nd ed.), Boca Raton, FL: Chapman & Hall.

[27] Gelman, A. & Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*, 457-511.

[28] Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith (Eds.), *Bayesian statistics 4* (pp. 169-194). Oxford: Oxford University Press.

[29] Gilks, W. R., Richardson, S., & Spiegelhalter, D. J.(1996). *Markov Chain Monte Carlo in practice*. Boca Raton, FL: Chapman & Hall.

[30] Gill, J. (2002). *Bayesian methods: A social and behavioral science approach*. Boca Raton, FL: Chapman & Hall.

[31] Hambleton, R. K., & Swaminathan, H.(1985) *Item response theroy: Principles and applications*. Hingham, MA: Kluwer-Nijhoff Publishing Company.

[32] Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage Publications.

[33] Hannan, E. J., & B. G. Quinn (1979). The determination of the order of an autoregression, *Journal of the Royal Statistical Society, Series B, 41*, 190195.

[34] Harwell, M., Stone, C.A., Hsu, T.-C., & Kirisci, L. (1996). Monte Carlo studies in item response theory. *Applied Psychological Measurement, 20*(2), 101-125.

[35] Heidelberger, P. & Welch, P. (1983). Simulation run length control in the presence of an initial transient. *Operations Research, 31*, 1109-1144.

[36] Izsák, A., Orrill, C. H., Cohen, A. S., & Brown, R. E. (2010). Measuring middle grades teachers understanding of rational numbers with the mixture Rasch model. *Elementary School Journal, 110*(3), 279-300.

[37] Jasra, A., Holmes, C. C. & Stephens, D. A. (2005). Markov Chain Monte Carlo Methods and the Label Switching Problem in Bayesian Mixture Modeling. *Statistical Science, 20*, 50-67.

[38] Kass, R. E. (1993). Bayes factors in practice, *Statistician, 42*, 551-560.

[39] Kass, R. E. & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*, 773-795.

[40] Kim, S.-H. (2001). An evaluation of a Markov chain Monte Carlo method for the Rasch model. *Applied Psychological Measurement, 25*, 163-176.

[41] Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices.* Newbury Park, NY: Springer.

[42] Li, F., Cohen, A.S., Kim, S.-H., & Cho, S.-J. (2009). Model selection methods for dichotomous mixture IRT models. *Applied Psychological Measurement, 33*, 353-373.

[43] Linacre, J. M. (2001a). *A user's guide to WINSTEPS/MINISTEPS*. Chicago, IL: Winsteps.com.

[44] Linacre, J. M. (2001b). Facets Rasch measurement software. Chicago: Winsteps.com.

[45] Linacre, J. M. & Wright, B. D. (2001). *A user's guide to BIGSTEPS*. Chicago, IL: MESA Press.

[46] Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* NJ : Lawrence Erlbaum Associates, Inc.

[47] Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores.* MA: Addison-Wesley.

[48] Meiser, T. & Machunsky, M. (2008). The personal structure of personal need for structure: A mixture-distribution Rasch analysis. *European Journal of Psychological Assessment 24*(1), 27-34.

[49] Muraki, E., & Bock, R. D. (2003). PARSCALE (Version 4.1) [Computer program]. Mooresville, IN: Scientific Software.

[50] Muthén, L.K. & Muthén, B.O. (2012). *Mplus User's Guide. Sixth Edition.* Los Angeles, CA: Muthén & Muthén

[51] Meyer, J. P. (2008, March). *A mixture Rasch model with item response time components.* Paper presented at the annual meeting of the National Council on Measurement in Education, New York, NY.

[52] Patz, R. J., & Junker,B . W. (1999a). A straight forward approach to Markov chain monte carlo methods for item response models. *Journal of Educational and Behavioral Statistics, 24*, 146-178.

[53] Patz, R. J., & Junker, B. W. (1999b). Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses. *Journal of Educational and Behavioral Statistics, 24*, 342-366.

[54] Plummer, M., Best, N., Cowles, K., Vine, K., Sarkar, D., & Almond, R. (2012). Package coda [Computer software]. Retrieved from http://cran.r-project.org/web/packages/coda/coda.pdf

[55] Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction (3rd ed.).* New York: Holt, Rinehart and Winston.

[56] Raftery, A. L. & Lewis, S. (1992). How many iterations in the Gibbs sampler? In J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith (Eds.), *Bayesian statistics 4*. Oxford: Oxford University Press.

[57] Render, R. A. & Walker, H. F. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review., 26*, 195-239.

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

[59] Samuelsen, K. M. (2008). Examining differential item functioning from a latent class perspective. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 67-113). Charlotte, NC: Information Age Publishing.

[60] Sawa, T (1978). Information criteria for discrimination among alternative regression models. *Econometrica, 46*, 1273-1291.

[61] Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461-464.

[62] Smith, B. J. (2007). boa: An R package for MCMC output convergence assessment and posterior interence. *Journal of Statistical Software, 21*(11), 1-37.

[63] Sperrin, M., Jaki, T., Wit, E. (2010) Probabilistic relabelling strategies for the label switching problem in Bayesian mixture models. *Statistics and Computing, 20*, 357-366.

[64] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Satistical Society, Series B, 64*, 1-34.

[65] Spiegelhalter, D. J., Thomas, A., Best, N. G., & Lunn, D. J. (2007). *OpenBUGS User Manual, Version 3.0.2.*, MRC Biostatistics Unit, Institute of Public Health and De-

partment of Epidemiology and Public Health, Imperial College School of Medicine, UK, available at http://mathstat.helsinki.fi/openbugs/Manuals/Contents.html.

[66] Stephens, M. (2000). Dealing with Label Switching in Mixture Models. *Journal of the Royal Statistical Society, Series B, 62*, 795-809.

[67] Stephens, M. (1997). Contribution to the discussion of Richardson and Green (1997), On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society, series B, 59*, 768-769.

[68] Thissen, D., Chen, W-H., & Bock, R. D. (2003). MULTILOG 7 for Windows: Multiple-category item analysis and test scoring using item response theory [Computer software]. Skokie, IL: Scientific Software International, Inc.

[69] Thissen, D., Steinberg, L., & Wainer, H. (1993). DIF detection using IRT models. In P. W. Holland, & H. Wainer (Eds.), *Differential item functioning*(pp. 67-113). Hillsdale, NJ: Lawrence Erlbaum Associates.

[70] von Davier, M. (2001). *WINMIRA 2001 user's manual.* available at http://208.76.84.140/ svfklumu/wmira/winmiramanual.pdf

[71] von Davier, M., & Yamamoto, K. (2004). Partially observed mixtures of IRT models: An extension of the generalized partial-credit model. *Applied Psychological Measurement, 28*, 389-406.

[72] Wingersky, M. S., Barton, M. A., & Lord, F. M. (1982). *LOGIST user's guide*. Princeton, NJ: Educational Testing Service.

[73] Zimowski, M., Muraki, E., Mislevy, R. J., & Bock, R. D. (2003). BILOG-MG (Version 3.0) [Computer program]. Mooresville, IN: Scientific Software.

Appendix A

R Code for Generating Data: Mix3PLM with Three Latent Class Groups

```
# Fix the working directory
setwd("c:/yj/DataThree")
getwd()

mainFolder <- "c:/yj/DataThree"

# Number of items
nItem<-20
#nItem<-40

# Number of subjects
nSub<-600
#nSub<-2400

# Item parameter information for 20 items
# Item difficulty parameters
b1<-c(-0.50, -0.50, 0, 0.50, 0.50, -2.00, -1.75, -1.50, -1.25, -1.00,
      -0.50, -0.25, 0, 0.25, 0.50, 1.00, 1.25, 1.50, 1.75, 2.00)
b2<-c(-0.50, -0.50, 0, 0.50, 0.50, -0.50, -0.25, 0, 0.25, 0.50,
       1.00, 1.25, 1.50, 1.75, 2.00, -2.00, -1.75, -1.50, -1.25, -1.00)
```

```
b3<-c(-0.50, -0.50, 0, 0.50, 0.50, 1.00, 1.25, 1.50, 1.75, 2.00,
      -2.00, -1.75, -1.50, -1.25, -1.00, -0.50, -0.25, 0, 0.25, 0.50)

# Item discrimination parameters
a1<-c(1,1,1,1,1,2,2,2,2,2,1,1,1,1,1,1,1,1,1,1)
a2<-c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2)
a3<-c(1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,1,1,1,1,1)

# Lower asymptote parameters
c1<-c(0.20, 0.20, 0.20, 0.20, 0.20, 0.10, 0.10, 0.10, 0.10, 0.10,
      0.20, 0.20, 0.20, 0.20, 0.20, 0.25, 0.25, 0.25, 0.25, 0.25)
c2<-c(0.20, 0.20, 0.20, 0.20, 0.20, 0.20, 0.20, 0.20, 0.20, 0.20,
      0.25, 0.25, 0.25, 0.25, 0.25, 0.10, 0.10, 0.10, 0.10, 0.10)
c3<-c(0.20, 0.20, 0.20, 0.20, 0.20, 0.25, 0.25, 0.25, 0.25, 0.25,
      0.10, 0.10, 0.10, 0.10, 0.10, 0.20, 0.20, 0.20, 0.20, 0.20)

# Item parameter information for 40 items
# Item difficulty parameters
#b1<-c(-0.50, -0.50, 0, 0.50, 0.50, -2.00, -1.75, -1.50, -1.25, -1.00,
#      -0.50, -0.25, 0, 0.25, 0.50, 1.00, 1.25, 1.50, 1.75, 2.00,
#      -0.50, -0.50, 0, 0.50, 0.50, -2.00, -1.75, -1.50, -1.25, -1.00,
#      -0.50, -0.25, 0, 0.25, 0.50, 1.00, 1.25, 1.50, 1.75, 2.00)
#b2<-c(-0.50, -0.50, 0, 0.50, 0.50, -0.50, -0.25, 0, 0.25, 0.50,
#       1.00, 1.25, 1.50, 1.75, 2.00, -2.00, -1.75, -1.50, -1.25, -1.00,
#      -0.50, -0.50, 0, 0.50, 0.50, -0.50, -0.25, 0, 0.25, 0.50,
#       1.00, 1.25, 1.50, 1.75, 2.00, -2.00, -1.75, -1.50, -1.25, -1.00)
#b3<-c(-0.50, -0.50, 0, 0.50, 0.50, 1.00, 1.25, 1.50, 1.75, 2.00,
#      -2.00, -1.75, -1.50, -1.25, -1.00, -0.50, -0.25, 0, 0.25, 0.50,
#      -0.50, -0.50, 0, 0.50, 0.50, 1.00, 1.25, 1.50, 1.75, 2.00,
#      -2.00, -1.75, -1.50, -1.25, -1.00, -0.50, -0.25, 0, 0.25, 0.50)

# Item discrimination parameters
#a1<-c(1,1,1,1,1,2,2,2,2,2,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,1,1,1,1,1,1,1,1,1,1)
#a2<-c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2)
#a3<-c(1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,1,1,1,1,1)

# Lower asymptote parameters
#c1<-c(0.20, 0.20, 0.20, 0.20, 0.20, 0.10, 0.10, 0.10, 0.10, 0.10,
#      0.20, 0.20, 0.20, 0.20, 0.20, 0.25, 0.25, 0.25, 0.25, 0.25,
#      0.20, 0.20, 0.20, 0.20, 0.20, 0.10, 0.10, 0.10, 0.10, 0.10,
#      0.20, 0.20, 0.20, 0.20, 0.20, 0.25, 0.25, 0.25, 0.25, 0.25)
#c2<-c(0.20, 0.20, 0.20, 0.20, 0.20, 0.20, 0.20, 0.20, 0.20, 0.20,
```

```
#      0.25, 0.25, 0.25, 0.25, 0.25, 0.10, 0.10, 0.10, 0.10, 0.10,
#      0.20, 0.20, 0.20, 0.20, 0.20, 0.20, 0.20, 0.20, 0.20, 0.20,
#      0.25, 0.25, 0.25, 0.25, 0.25, 0.10, 0.10, 0.10, 0.10, 0.10)
#c3<-c(0.20, 0.20, 0.20, 0.20, 0.20, 0.25, 0.25, 0.25, 0.25, 0.25,
#      0.10, 0.10, 0.10, 0.10, 0.10, 0.20, 0.20, 0.20, 0.20, 0.20,
#      0.20, 0.20, 0.20, 0.20, 0.20, 0.25, 0.25, 0.25, 0.25, 0.25,
#      0.10, 0.10, 0.10, 0.10, 0.10, 0.20, 0.20, 0.20, 0.20, 0.20)

# To make file names
if (nItem==20) {
    nItemm<-2
}
if (nItem==40) {
    nItemm<-4
}

if (nSub==600) {
    nSubb<-6
}
if (nSub==1200) {
    nSubb<-12
}
if (nSub==2400) {
    nSubb<-24
}

# Mean of normal distribution for class 1, 2, and 3
meanTheta1<-0
meanTheta2<-0
meanTheta3<-0

# SD of normal distribution for class 1, 2, and 3
sdTheta1<-1
sdTheta2<-1
sdTheta3<-1

# Random seeds for three groups and 20 replications
seed3<-c(14834, 37776, 60718, 83659, 06602, 29544, 52485, 75427, 98369, 21311,
         44253, 63389, 74860, 86330, 97801, 09273, 20744, 32215, 43686, 55156,
         10756, 22227, 33697, 45168, 56639, 68110, 79581, 91052, 02523, 13994,
         25465, 36936, 48407, 59878, 71348, 82819, 94290, 05762, 17233, 28703,
         97476, 53212, 08948, 64683, 20419, 76154, 31890, 87625, 43361, 99096,
```

```
        54831, 10567, 66302, 22038, 77773, 33509, 89244, 44980, 00716, 56451)

seed3<-matrix(seed3,20,3)

for (iter in 1:20) {
    theta1<-mat.or.vec(nSub*.6,1)
    theta2<-mat.or.vec(nSub*.3,1)
    theta3<-mat.or.vec(nSub*.1,1)
    set.seed(seed3[iter,1])
    theta1<-rnorm(nSub*.6,mean=meanTheta1,sd=sdTheta1)
    set.seed(seed3[iter,2])
    theta2<-rnorm(nSub*.3,mean=meanTheta2,sd=sdTheta2)
    set.seed(seed3[iter,3])
    theta3<-rnorm(nSub*.1,mean=meanTheta3,sd=sdTheta3)
    thetatemp<-c(theta1,theta2,theta3)

### Generating complete data set ###

response<-mat.or.vec(nSub,nItem+1)
for (row in 1:nSub) {
if (row<=nSub*.6) {
            class<-1
            b<-b1
            a<-a1
            c<-c1
            theta<-thetatemp[row]
}
    if (row>nSub*.6 & row<=nSub*.9) {
            class<-2
            b<-b2
            a<-a2
            c<-c2
            theta<-thetatemp[row]
}
if (row>nSub*.9) {
            class<-3
            b<-b3
            a<-a3
            c<-c3
            theta<-thetatemp[row]
}
for (column in 1:nItem) {
```

```
            p=c[column] + (1-c[column])*(exp(a[column]*(theta-b[column]))
                /(1+exp(a[column]*(theta-b[column]))))
            r<-runif(1,0,1)
            if (r<=p) {
                response[row,column]<-1
            }
            else {
                response[row,column]<-0
            }
        }
        response[row,nItem+1]<-class
    }

#### The End of generating complete data set ####

# generating output file for complete data
# response : item + 1 = class
comp<-response[,1:nItem]

# File for item calibration
filename1<-paste("ThreeI",nItemm,"S",nSubb,"G3B_",iter,".txt",sep="")

# Original file containing class information
# ThreeI2S6G3_1 : Mix3PLM & 20 items & 600 subjects & 3 latent groups & 1st replication
filename2<-paste("ThreeI",nItemm,"S",nSubb,"G3B_C",iter,".txt",sep="")

# Data with generated class information
write.table(response,file=filename2,sep=',',row.names=FALSE,col.names=FALSE)

# Data for OpenBUGS script
write(paste("list(NE=",nSub,", NI=",nItem,",G2=2,G3=3,G4=4,alpha2=c(.5,.5),
        alpha3=c(.5,.5,.5),alpha4=c(.5,.5,.5,.5),",sep=""),
        file=file.path(mainFolder,filename1), append=T)
write("resp=structure(.Data=c(",file=file.path(mainFolder,filename1), append=T)

    for (j in 1:nSub) {
        write(comp[j,],sep=',',ncolumns=nItem,
            file=file.path(mainFolder,filename1), append=T)
}

    write(paste("), .Dim=c(",nSub,",",nItem,")))",sep=""),
        file=file.path(mainFolder,filename1), append=T)
```

```
    temp <- readLines(file.path(mainFolder,filename1),n=-1)

    for (i in 3:(3+nSub-2)) {
        temp[i] <- paste(temp[i],",",sep="")
    }
    write(temp,file=file.path(mainFolder,filename1), append=F)
}
```

Appendix B

OpenBUGS Code Used for Mixture Rasch Model with Item Anchoring

```
# NE: the number of examinees
# NI: the number of items
# theta: ability parameter
# b: item difficulty parameter
# p: probability of correct response
# mut: latent group mean
# gmem: group membership
# G1 to G4: one to four latent groups
# par: the number of estimated parameters

model
{
    for (j in 1:NE) {
        for (i in 1:NI) {
            r1[j, i] <- resp[j, i]
            r2[j, i] <- resp[j, i]
            r3[j, i] <- resp[j, i]
            r4[j, i] <- resp[j, i]
        }
    }
# one group
    for (j in 1:NE) {
        for (i in 1:NI) {
            logit(p1[j, i]) <- theta1[j] - b1[i]
            r1[j, i] ~ dbern(p1[j, i])
            l1[j, i] <- log(p1[j, i]) * r1[j, i] + log(1 - p1[j,i]) * (1 - r1[j, i])
        }
    }
```

```
    for (j in 1:NE) {
        theta1[j] ~ dnorm(mut1, 1)
    }
    mut1 ~ dnorm(0.00000E+00, 1)
    for (i in 6:NI) {
        b1[i] ~ dnorm(0.00000E+00, 1)
    }
# five anchor items
    b1[1] <- -0.5
    b1[2] <- -0.5
    b1[3] <- 0.00000E+00
    b1[4] <- 0.5
    b1[5] <- 0.5

# two groups
    for (j in 1:NE) {
        for (i in 1:NI) {
            logit(p2[j, i]) <- theta2[j] - b2[gmem2[j], i]
            r2[j, i] ~ dbern(p2[j, i])
            l2[j, i] <- log(p2[j, i]) * r2[j, i] + log(1 - p2[j,i]) * (1 - r2[j, i])
        }
    }
    for (j in 1:NE) {
        theta2[j] ~ dnorm(mut2[gmem2[j]], 1)
        gmem2[j] ~ dcat(pi2[1:G2])
    }
    for (k in 1:G2) {
        mut2[k] ~ dnorm(0.00000E+00, 1)
    }
    pi2[1:G2] ~ ddirch(alpha2[])
    for (i in 1:NI) {
        b2[1, i] ~ dnorm(0.00000E+00, 1)
    }
    for (i in 6:NI) {
        b2[2, i] ~ dnorm(0.00000E+00, 1)
    }
# five anchor items
    b2[2, 1] <- b2[1, 1]
    b2[2, 2] <- b2[1, 2]
    b2[2, 3] <- b2[1, 3]
    b2[2, 4] <- b2[1, 4]
    b2[2, 5] <- b2[1, 5]
```

```
# three groups
    for (j in 1:NE) {
        for (i in 1:NI) {
            logit(p3[j, i]) <- theta3[j] - b3[gmem3[j], i]
            r3[j, i] ~ dbern(p3[j, i])
            l3[j, i] <- log(p3[j, i]) * r3[j, i] + log(1 - p3[j,i]) * (1 - r3[j, i])
        }
    }
    for (j in 1:NE) {
        theta3[j] ~ dnorm(mut3[gmem3[j]], 1)
        gmem3[j] ~ dcat(pi3[1:G3])
    }
    for (k in 1:G3) {
        mut3[k] ~ dnorm(0.00000E+00, 1)
    }
    pi3[1:G3] ~ ddirch(alpha3[])
    for (i in 1:NI) {
        b3[1, i] ~ dnorm(0.00000E+00, 1)
    }
    for (k in 2:G3) {
        for (i in 6:NI) {
            b3[k, i] ~ dnorm(0.00000E+00, 1)
        }
    }
# five anchor items
    b3[2, 1] <- b3[1, 1]
    b3[2, 2] <- b3[1, 2]
    b3[2, 3] <- b3[1, 3]
    b3[2, 4] <- b3[1, 4]
    b3[2, 5] <- b3[1, 5]
    b3[3, 1] <- b3[1, 1]
    b3[3, 2] <- b3[1, 2]
    b3[3, 3] <- b3[1, 3]
    b3[3, 4] <- b3[1, 4]
    b3[3, 5] <- b3[1, 5]

# four groups
    for (j in 1:NE) {
        for (i in 1:NI) {
            logit(p4[j, i]) <- theta4[j] - b4[gmem4[j], i]
            r4[j, i] ~ dbern(p4[j, i])
```

```
            l4[j, i] <- log(p4[j, i]) * r4[j, i] + log(1 - p4[j,i]) * (1 - r4[j, i])
        }
    }
    for (j in 1:NE) {
        theta4[j] ~ dnorm(mut4[gmem4[j]], 1)
        gmem4[j] ~ dcat(pi4[1:G4])
    }
    for (k in 1:G4) {
        mut4[k] ~ dnorm(0.00000E+00, 1)
    }
    pi4[1:G4] ~ ddirch(alpha4[])
    for (i in 1:NI) {
        b4[1, i] ~ dnorm(0.00000E+00, 1)
    }
    for (k in 2:G4) {
        for (i in 6:NI) {
            b4[k, i] ~ dnorm(0.00000E+00, 1)
        }
    }
# five anchor items
    b4[2, 1] <- b4[1, 1]
    b4[2, 2] <- b4[1, 2]
    b4[2, 3] <- b4[1, 3]
    b4[2, 4] <- b4[1, 4]
    b4[2, 5] <- b4[1, 5]
    b4[3, 1] <- b4[1, 1]
    b4[3, 2] <- b4[1, 2]
    b4[3, 3] <- b4[1, 3]
    b4[3, 4] <- b4[1, 4]
    b4[3, 5] <- b4[1, 5]
    b4[4, 1] <- b4[1, 1]
    b4[4, 2] <- b4[1, 2]
    b4[4, 3] <- b4[1, 3]
    b4[4, 4] <- b4[1, 4]
    b4[4, 5] <- b4[1, 5]

# model selection using AIC and BIC
    loglik[1] <- sum(l1[1:NE, 1:NI])
    loglik[2] <- sum(l2[1:NE, 1:NI])
    loglik[3] <- sum(l3[1:NE, 1:NI])
    loglik[4] <- sum(l4[1:NE, 1:NI])
```

```
    par[1] <- (NI - 5) + 1

    for (k in 2:G4) {
        par[k] <- (k * NI) - 5 * (k - 1) + (2 * k - 1)
    }

    for (k in 1:G4) {
        AIC[k] <- -2 * loglik[k] + 2 * par[k]
        BIC[k] <- -2 * loglik[k] + par[k] * log(NE)
    }
}
```

Appendix C

OpenBUGS Code Used for Mixture 2PL Model with Person Centering

```
# NE: the number of examinees
# NI: the number of items
# theta: ability parameter
# a: item discrimination parameter
# b: item difficulty parameter
# p: probability of correct response
# mut: latent group mean
# gmem: group membership
# G1 to G4: one to four latent groups
# par: the number of estimated parameters

model
{
    for (j in 1:NE) {
        for (i in 1:NI) {
            r1[j, i] <- resp[j, i]
            r2[j, i] <- resp[j, i]
            r3[j, i] <- resp[j, i]
            r4[j, i] <- resp[j, i]
        }
    }
# one group
    for (j in 1:NE) {
        for (i in 1:NI) {
            logit(p1[j, i]) <- a1[i] * (theta1[j] - b1[i])
            r1[j, i] ~ dbern(p1[j, i])
            l1[j, i] <- log(p1[j, i]) * r1[j, i] + log(1 - p1[j, i]) * (1 - r1[j, i])
        }
```

```
    }
    for (j in 1:NE) {
        theta1[j] ~ dnorm(mut1, 1)
    }
    mut1 <- 0.00000E+00
    for (i in 1:NI) {
        a1[i] ~ dnorm(0.00000E+00, 1)T(0,)
        b1[i] ~ dnorm(0.00000E+00, 1)
    }

# two groups
    for (j in 1:NE) {
        for (i in 1:NI) {
            logit(p2[j, i]) <- a2[gmem2[j], i] * (theta2[j] - b2[gmem2[j], i])
            r2[j, i] ~ dbern(p2[j, i])
            l2[j, i] <- log(p2[j, i]) * r2[j, i] + log(1 - p2[j, i]) * (1 - r2[j, i])
        }
    }
    for (j in 1:NE) {
        theta2[j] ~ dnorm(mut2[gmem2[j]], 1)
        gmem2[j] ~ dcat(pi2[1:G2])
    }
    mut2[1] <- 0.00000E+00
    mut2[2] ~ dnorm(0.00000E+00, 1)
    pi2[1:G2] ~ ddirch(alpha2[])
    for (k in 1:G2) {
        for (i in 1:NI) {
            a2[k, i] ~ dnorm(0.00000E+00, 1)T(0,)
            b2[k, i] ~ dnorm(0.00000E+00, 1)
        }
    }

# three groups
    for (j in 1:NE) {
        for (i in 1:NI) {
            logit(p3[j, i]) <- a3[gmem3[j], i] * (theta3[j] - b3[gmem3[j], i])
            r3[j, i] ~ dbern(p3[j, i])
            l3[j, i] <- log(p3[j, i]) * r3[j, i] + log(1 - p3[j, i]) * (1 - r3[j, i])
        }
    }
    for (j in 1:NE) {
        theta3[j] ~ dnorm(mut3[gmem3[j]], 1)
```

```
        gmem3[j] ~ dcat(pi3[1:G3])
    }
    for (k in 2:G3) {
        mut3[k] ~ dnorm(0.00000E+00, 1)
    }
    mut3[1] <- 0.00000E+00
    pi3[1:G3] ~ ddirch(alpha3[])
    for (k in 1:G3) {
        for (i in 1:NI) {
            a3[k, i] ~ dnorm(0.00000E+00, 1)T(0,)
            b3[k, i] ~ dnorm(0.00000E+00, 1)
        }
    }

# four groups
    for (j in 1:NE) {
        for (i in 1:NI) {
            logit(p4[j, i]) <- a4[gmem4[j], i] * (theta4[j] - b4[gmem4[j], i])
            r4[j, i] ~ dbern(p4[j, i])
            l4[j, i] <- log(p4[j, i]) * r4[j, i] + log(1 - p4[j, i]) * (1 - r4[j, i])
        }
    }
    for (j in 1:NE) {
        theta4[j] ~ dnorm(mut4[gmem4[j]], 1)
        gmem4[j] ~ dcat(pi4[1:G4])
    }
    for (k in 2:G4) {
        mut4[k] ~ dnorm(0.00000E+00, 1)
    }
    mut4[1] <- 0.00000E+00
    pi4[1:G4] ~ ddirch(alpha4[])
    for (k in 1:G4) {
        for (i in 1:NI) {
            a4[k, i] ~ dnorm(0.00000E+00, 1)T(0,)
            b4[k, i] ~ dnorm(0.00000E+00, 1)
        }
    }
# model selection using AIC and BIC
    loglik[1] <- sum(l1[1:NE, 1:NI])
    loglik[2] <- sum(l2[1:NE, 1:NI])
    loglik[3] <- sum(l3[1:NE, 1:NI])
    loglik[4] <- sum(l4[1:NE, 1:NI])
```

```
    for (k in 1:4) {
        par[k] <- (NI * 2 * k) + 2 * (k - 1)
        AIC[k] <- -2 * loglik[k] + 2 * par[k]
        BIC[k] <- -2 * loglik[k] + par[k] * log(NE)
    }
}
```

Appendix D

OpenBUGS Code Used for Mixture 3PL Model with Item Centering

```
# NE: the number of examinees
# NI: the number of items
# theta: ability parameter
# a: item discrimination parameter
# b: item difficulty parameter
# c: lower asymptote parameter
# p: probability of correct response
# mut: latent group mean
# gmem: group membership
# G1 to G4: one to four latent groups
# par: the number of estimated parameters

model
{
    for (j in 1:NE) {
        for (i in 1:NI) {
            r1[j, i] <- resp[j, i]
            r2[j, i] <- resp[j, i]
            r3[j, i] <- resp[j, i]
            r4[j, i] <- resp[j, i]
        }
    }
# one group
    for (j in 1:NE) {
        for (i in 1:NI) {
            logit(tt1[j,i]) <- a1[i]*(theta1[j]-b1[i])
            p1[j,i] <- c1[i]+(1-c1[i])* tt1[j,i]
            r1[j, i] ~ dbern(p1[j, i])
```

```
            l1[j, i] <- log(p1[j, i]) * r1[j, i] + log(1 - p1[j, i]) * (1 - r1[j, i])
        }
    }
    for (j in 1:NE) {
        theta1[j] ~ dnorm(mut1, 1)
    }
    mut1 ~ dnorm(0.00000E+00, 1)
    for (i in 1:NI) {
        a1[i] ~ dnorm(0.00000E+00, 1)T(0,)
        beta1[i] ~ dnorm(0.00000E+00, 1)
        b1[i] <- beta1[i] - mean(beta1[1:NI])
        c1[i] ~ dbeta(5, 17)
    }

# two groups
    for (j in 1:NE) {
        for (i in 1:NI) {
            logit(tt2[j,i]) <- a2[gmem2[j],i]*(theta2[j]-b2[gmem2[j],i])
            p2[j,i] <- c2[gmem2[j],i]+(1-c2[gmem2[j],i])* tt2[j,i]
            r2[j, i] ~ dbern(p2[j, i])
            l2[j, i] <- log(p2[j, i]) * r2[j, i] + log(1 - p2[j, i]) * (1 - r2[j, i])
        }
    }
    for (j in 1:NE) {
        theta2[j] ~ dnorm(mut2[gmem2[j]], 1)
        gmem2[j] ~ dcat(pi2[1:G2])
    }
    for (k in 1:G2) {
        mut2[k] ~ dnorm(0.00000E+00, 1)
    }
    pi2[1:G2] ~ ddirch(alpha2[])
    for (k in 1:G2) {
        for (i in 1:NI) {
            a2[k, i] ~ dnorm(0.00000E+00, 1)T(0,)
            beta2[k, i] ~ dnorm(0.00000E+00, 1)
            b2[k, i] <- beta2[k, i] - mean(beta2[k, 1:NI])
            c2[k, i] ~ dbeta(5, 17)
        }
    }

# three groups
    for (j in 1:NE) {
```

```
        for (i in 1:NI) {
            logit(tt3[j,i]) <- a3[gmem3[j],i]*(theta3[j]-b3[gmem3[j],i])
            p3[j,i] <- c3[gmem3[j],i]+(1-c3[gmem3[j],i])* tt3[j,i]
            r3[j, i] ~ dbern(p3[j, i])
            l3[j, i] <- log(p3[j, i]) * r3[j, i] + log(1 - p3[j, i]) * (1 - r3[j, i])
        }
    }
    for (j in 1:NE) {
        theta3[j] ~ dnorm(mut3[gmem3[j]], 1)
        gmem3[j] ~ dcat(pi3[1:G3])
    }
    for (k in 1:G3) {
        mut3[k] ~ dnorm(0.00000E+00, 1)
    }
    pi3[1:G3] ~ ddirch(alpha3[])
    for (k in 1:G3) {
        for (i in 1:NI) {
            a3[k, i] ~ dnorm(0.00000E+00, 1)T(0,)
            beta3[k, i] ~ dnorm(0.00000E+00, 1)
            b3[k, i] <- beta3[k, i] - mean(beta3[k, 1:NI])
            c3[k, i] ~ dbeta(5, 17)
        }
    }
# four groups
    for (j in 1:NE) {
        for (i in 1:NI) {
            logit(tt4[j,i]) <- a4[gmem4[j],i]*(theta4[j]-b4[gmem4[j],i])
            p4[j,i] <- c4[gmem4[j],i]+(1-c4[gmem4[j],i])* tt4[j,i]
            r4[j, i] ~ dbern(p4[j, i])
            l4[j, i] <- log(p4[j, i]) * r4[j, i] + log(1 - p4[j, i]) * (1 - r4[j, i])
        }
    }
    for (j in 1:NE) {
        theta4[j] ~ dnorm(mut4[gmem4[j]], 1)
        gmem4[j] ~ dcat(pi4[1:G4])
    }
    for (k in 1:G4) {
        mut4[k] ~ dnorm(0.00000E+00, 1)
    }
    pi4[1:G4] ~ ddirch(alpha4[])
    for (k in 1:G4) {
        for (i in 1:NI) {
```

```
            a4[k, i] ~ dnorm(0.00000E+00, 1)T(0,)
            beta4[k, i] ~ dnorm(0.00000E+00, 1)
            b4[k, i] <- beta4[k, i] - mean(beta4[k, 1:NI])
            c4[k, i] ~ dbeta(5, 17)
        }
    }
# model selection using AIC and BIC
    loglik[1] <- sum(l1[1:NE, 1:NI])
    loglik[2] <- sum(l2[1:NE, 1:NI])
    loglik[3] <- sum(l3[1:NE, 1:NI])
    loglik[4] <- sum(l4[1:NE, 1:NI])

    for (k in 1:4) {
        par[k] <- (NI * 3 * k) + 2 * k - 1
        AIC[k] <- -2 * loglik[k] + 2 * par[k]
        BIC[k] <- -2 * loglik[k] + par[k] * log(NE)
    }
}
```

Appendix E

Monitoring Convergence using Heidelberger and Welch's (1983) Convergence Diagnostics and Bayesian Credible Intervals

Table E.1: Heidelberger and Welch's (1983) Convergence Diagnostics for MixRM

	Post	20-items		40-items		
Burn-in	Burn-in	n = 600	n = 2,400	n = 600	n = 2,400	Total
0	15,000	**200**	**200**	**400**	398	1,198
1,000	7,000	**200**	199	**400**	379	1,178
1,000	8,000	199	**200**	392	397	1,188
1,000	9,000	**200**	196	**400**	398	1,194
2,000	9,000	199	198	396	**400**	1,193

Continued on next page

Table E.1 – continued from previous page

Burn-in	Post Burn-in	20-items n = 600	20-items n = 2,400	40-items n = 600	40-items n = 2,400	Total
2,000	10,000	199	199	399	395	1,192
2,000	11,000	**200**	**200**	**400**	395	1,195
2,000	12,000	198	**200**	**400**	**400**	1,198
2,000	13,000	**200**	**200**	393	**400**	1,193
3,000	7,000	199	198	**400**	**400**	1,197
3,000	8,000	199	**200**	399	399	1,197
3,000	9,000	199	199	**400**	395	1,193
3,000	10,000	**200**	**200**	**400**	398	1,198
3,000	11,000	197	199	398	**400**	1,194
3,000	12,000	199	**200**	361	**400**	1,160
4,000	6,000	199	**200**	**400**	332	1,131
4,000	7,000	199	**200**	**399**	399	1,197
4,000	8,000	**200**	199	**400**	391	1,190
4,000	9,000	**200**	**200**	**400**	399	1,199
4,000	10,000	199	199	363	**400**	1,161
4,000	11,000	199	**200**	359	388	1,146
5,000	**5,000**	**200**	**200**	**400**	**400**	**1,200**
5,000	6,000	186	199	**400**	398	1,183
5,000	7,000	186	194	**400**	398	1,178
5,000	8,000	**200**	**200**	**400**	**400**	1,200
5,000	9,000	198	**200**	360	**400**	1,158
5,000	10,000	199	**200**	357	392	1,148

Continued on next page

Table E.1 – continued from previous page

Burn-in	Post Burn-in	20-items n = 600	20-items n = 2,400	40-items n = 600	40-items n = 2,400	Total
6,000	4,000	179	**200**	**400**	**400**	1,179
6,000	5,000	199	199	399	366	1,163
6,000	6,000	179	199	**400**	398	1,176
6,000	7,000	199	**200**	398	**400**	1,197
6,000	8,000	195	198	360	**400**	1,153
6,000	9,000	198	**200**	358	397	1,153

Table E.2: Heidelberger and Welch's (1983) Convergence Diagnostics for Mix2PLM

	Post	20-items		40-items		
Burn-in	Burn-in	n = 600	n = 2,400	n = 600	n = 2,400	Total
0	15,000	**200**	190	348	393	1,131
1,000	7,000	**200**	182	386	351	1,119
1,000	8,000	196	178	382	385	1,141
1,000	9,000	187	198	391	370	1,146
2,000	9,000	193	194	382	390	1,159
2,000	10,000	196	192	367	391	1,146
2,000	11,000	197	174	358	391	1,120
2,000	12,000	198	194	352	361	1,105
2,000	13,000	**200**	189	344	357	1,090
3,000	7,000	181	193	391	370	1,135
3,000	8,000	196	195	385	389	1,165
3,000	9,000	197	191	362	394	1,144
3,000	10,000	198	175	356	372	1,101
3,000	11,000	197	193	351	357	1,098
3,000	12,000	**200**	189	346	375	1,110
4,000	6,000	181	197	383	342	1,103
4,000	7,000	191	196	380	396	1,163
4,000	8,000	196	195	360	386	1,137
4,000	9,000	196	181	357	358	1,092
4,000	10,000	196	194	355	380	1,125
4,000	11,000	198	192	353	381	1,124

Continued on next page

Table E.2 – continued from previous page

	Post	20-items		40-items		
Burn-in	Burn-in	n = 600	n = 2,400	n = 600	n = 2,400	Total
4,000	12,000	**200**	189	356	397	1,142
4,000	13,000	**200**	191	376	396	1,163
4,000	14,000	199	195	375	389	1,158
4,000	15,000	198	195	377	391	1,161
4,000	16,000	199	198	387	392	1,176
5,000	5,000	182	197	381	374	1,134
5,000	6,000	198	194	375	386	1,153
5,000	7,000	196	193	359	379	1,127
5,000	8,000	196	184	363	396	1,139
5,000	9,000	195	196	364	395	1,150
5,000	10,000	199	195	358	382	1,134
5,000	11,000	**200**	184	361	396	1,141
5,000	12,000	197	196	375	394	1,162
5,000	13,000	**200**	195	380	389	1,164
5,000	14,000	199	196	388	391	1,174
5,000	15,000	199	197	388	392	1,176
6,000	4,000	187	197	367	388	1,139
6,000	5,000	196	192	376	390	1,154
6,000	6,000	195	188	367	394	1,144
6,000	7,000	198	183	376	399	1,156
6,000	8,000	195	195	379	369	1,138
6,000	9,000	199	183	359	393	1,134

Continued on next page

	Post	20-items		40-items		
Burn-in	Burn-in	n = 600	n = 2,400	n = 600	n = 2,400	Total
6,000	10,000	**200**	193	364	396	1,153
6,000	**11,000**	199	**200**	391	395	**1,185**
6,000	12,000	**200**	196	385	386	1,167
6,000	13,000	199	196	385	391	1,171
6,000	14,000	199	196	390	393	1,178
7,000	9,000	**200**	193	371	397	1,161
7,000	10,000	196	199	396	392	1,183
7,000	11,000	199	196	387	387	1,169
7,000	12,000	199	194	389	391	1,173
7,000	13,000	198	195	389	391	1,173
8,000	8,000	197	193	391	399	1,180
8,000	9,000	195	198	398	391	1,182
8,000	10,000	199	195	387	388	1,169
8,000	11,000	198	193	389	390	1,170
8,000	12,000	199	195	391	389	1,174
9,000	7,000	197	192	392	**400**	1,181
9,000	8,000	194	196	396	391	1,177
9,000	9,000	198	197	385	385	1,165
9,000	10,000	199	193	386	389	1,167
9,000	11,000	197	191	392	389	1,169
10,000	6,000	195	193	395	393	1,176
10,000	7,000	194	196	394	387	1,171

Continued on next page

Table E.2 – continued from previous page

	Post	20-items		40-items		
Burn-in	Burn-in	n = 600	n = 2,400	n = 600	n = 2,400	Total
10,000	8,000	198	194	380	376	1,148
10,000	9,000	198	185	383	388	1,154
10,000	10,000	197	175	392	390	1,154
11,000	5,000	193	188	391	396	1,168
11,000	6,000	195	194	379	364	1,132
11,000	7,000	197	190	371	378	1,136
11,000	8,000	199	174	380	388	1,141
11,000	9,000	196	177	390	391	1,154
12,000	4,000	195	180	394	359	1,128
12,000	5,000	195	173	350	379	1,097
12,000	6,000	198	172	376	388	1,134
12,000	7,000	198	171	381	387	1,137
12,000	8,000	199	195	386	393	1,173
13,000	3,000	196	168	374	366	1,104
13,000	4,000	199	177	369	375	1,120
13,000	5,000	199	176	372	387	1,134
13,000	6,000	**200**	184	385	387	1,156
13,000	7,000	199	**196**	391	390	1,176
14,000	3,000	183	176	365	369	1,093
14,000	4,000	195	186	388	374	1,143
14,000	5,000	199	197	387	386	1,169
14,000	6,000	**200**	196	388	386	1,170

Continued on next page

Table E.2 – continued from previous page

	Post	20-items		40-items		
Burn-in	Burn-in	n = 600	n = 2,400	n = 600	n = 2,400	Total
15,000	3,000	198	176	377	395	1,146
15,000	4,000	199	**200**	391	363	1,153
15,000	5,000	198	190	390	390	1,168
16,000	3,000	198	193	391	386	1,168
16,000	4,000	195	193	388	390	1,166
17,000	3,000	**200**	192	393	392	1,177
18,000	2,000	192	190	387	373	1,142

Table E.3: Bayesian Credible Intervals for Mix3PLM with 20-items and 1-group

Latent Class	Item	n = 600			n = 2,400		
No.	No.	Mean	2.5%	97.5%	Mean	2.5%	97.5%
1	1	-0.28	-0.79	0.30	-0.65	-1.12	0.38
1	2	0.18	-0.37	0.68	0.00	-0.38	0.39
1	3	0.37	-0.31	1.09	-0.24	-0.72	0.28
1	4	2.49	2.09	2.89	-2.07	-2.70	-1.49
1	5	0.42	0.02	0.88	-0.09	-1.01	0.55
1	6	0.42	-0.25	1.04	-0.56	-1.13	0.02
1	7	1.44	0.87	2.25	0.05	-0.40	0.85
1	8	-0.97	-1.55	0.06	-1.07	-2.21	-0.47
1	9	0.26	-0.40	0.94	-0.22	-0.74	0.32
1	10	-0.63	-1.53	0.27	-0.54	-1.33	0.18
1	11	0.17	-0.48	0.97	1.05	0.52	1.69
1	12	0.55	-0.55	1.59	1.47	0.75	2.07
1	13	-0.20	-1.03	0.85	1.41	0.86	2.54
1	14	0.55	-0.95	1.44	-0.26	-0.97	0.43
1	15	1.39	-0.17	2.47	1.22	0.50	1.63
1	16	-0.78	-1.18	0.27	1.00	0.22	1.86
1	17	1.12	0.14	1.99	-0.79	-1.61	-0.34
1	18	0.87	-0.30	2.21	-0.06	-0.70	0.71
1	19	1.08	0.48	1.67	1.79	1.13	2.68
1	20	0.89	0.20	1.67	-1.31	-2.49	-0.22

Table E.4: Bayesian Credible Intervals for Mix3PLM with 20-items and 2-group

Latent Class	Item	n = 600			n = 2,400		
No.	No.	Mean	2.5%	97.5%	Mean	2.5%	97.5%
1	1	-1.55	-2.02	-0.85	-0.97	-1.99	-0.31
1	2	-0.21	-1.10	0.44	1.81	0.98	2.55
1	3	-0.08	-0.54	0.54	0.31	-0.86	1.16
1	4	0.06	-0.41	0.92	1.25	0.82	1.77
1	5	-1.30	-1.64	-0.83	0.56	0.26	0.93
1	6	-2.76	-3.28	-2.31	-2.45	-3.11	-2.06
1	7	-2.61	-2.94	-2.22	1.84	1.42	2.13
1	8	-2.61	-2.99	-2.06	-2.57	-2.99	-2.15
1	9	-2.08	-2.70	-1.51	-2.78	-3.81	-2.12
1	10	-2.09	-2.65	-1.56	-1.85	-2.71	-0.48
1	11	0.42	-0.03	1.02	-1.50	-2.08	-0.96
1	12	0.07	-0.38	0.61	0.56	-0.83	1.09
1	13	1.14	0.13	1.78	-0.18	-0.83	0.35
1	14	0.26	-0.22	0.71	-0.06	-0.48	0.44
1	15	-0.21	-1.34	0.81	0.09	-0.61	0.90
1	16	-0.04	-0.72	0.34	0.03	-0.65	0.71
1	17	-0.25	-0.70	0.20	-0.46	-1.12	0.12
1	18	-0.58	-1.16	0.07	1.04	-0.19	1.73
1	19	2.10	1.60	2.76	-0.97	-1.64	-0.44
1	20	1.21	0.58	2.11	0.31	-1.20	2.17
2	1	0.30	-0.01	0.73	0.73	-0.03	1.80

Continued on next page

Table E.4 – continued from previous page

Latent Class	Item	n = 600			n = 2,400		
No.	No.	Mean	2.5%	97.5%	Mean	2.5%	97.5%
2	2	0.69	0.24	1.08	-0.28	-0.59	0.00
2	3	0.27	-0.45	0.81	1.71	1.04	2.40
2	4	2.09	1.74	2.41	1.97	1.48	2.40
2	5	-0.11	-0.99	1.25	-0.43	-1.83	1.97
2	6	0.19	-0.20	0.70	0.03	-0.32	0.52
2	7	-2.03	-2.55	-1.28	-0.21	-1.15	0.79
2	8	1.19	0.12	1.99	0.78	-0.29	1.34
2	9	0.26	-0.20	0.73	1.33	0.67	2.39
2	10	1.10	-0.05	1.69	2.08	1.34	2.99
2	11	-0.15	-0.80	0.61	0.32	-0.62	1.18
2	12	0.82	-0.44	1.56	1.54	1.12	1.90
2	13	0.21	-1.11	1.27	-1.98	-2.83	-0.86
2	14	-0.22	-0.92	0.52	1.76	-0.74	3.20
2	15	2.18	1.75	2.72	2.65	1.76	2.99
2	16	1.37	0.62	2.08	0.54	0.07	0.90
2	17	-0.87	-1.56	-0.29	0.37	-0.27	0.85
2	18	0.50	0.07	0.99	-0.11	-0.75	0.59
2	19	0.03	-0.45	0.57	-0.74	-1.38	-0.11
2	20	1.10	0.08	1.94	0.41	-0.17	1.08

Table E.5: Bayesian Credible Intervals for Mix3PLM with 20-items and 3-group

Latent Class	Item	n = 600			n = 2,400		
No.	No.	Mean	2.5%	97.5%	Mean	2.5%	97.5%
1	1	-1.41	-1.79	-1.15	-0.16	-0.96	0.56
1	2	0.38	-0.68	1.66	-1.07	-1.64	-0.29
1	3	-1.37	-2.01	-0.63	-1.12	-1.82	-0.41
1	4	1.10	0.34	1.81	-0.80	-1.45	-0.03
1	5	-0.59	-1.15	-0.15	-0.10	-1.64	0.92
1	6	-3.47	-3.98	-2.63	-0.87	-3.48	1.00
1	7	1.30	0.85	1.85	-3.24	-3.63	-2.94
1	8	-2.31	-2.81	-1.60	-2.65	-3.37	-1.74
1	9	-2.15	-2.80	-1.73	-0.12	-1.04	0.71
1	10	-0.63	-1.73	0.45	-2.12	-2.75	-1.64
1	11	0.70	0.31	1.29	-1.00	-1.62	-0.65
1	12	-0.11	-0.64	0.61	-1.36	-2.19	-0.28
1	13	1.55	0.81	2.15	-0.68	-1.51	0.10
1	14	-0.61	-1.15	-0.06	-0.14	-0.76	0.62
1	15	-0.29	-0.97	0.25	-0.16	-0.79	0.68
1	16	1.64	1.19	2.11	-1.70	-2.57	-0.89
1	17	-0.38	-1.01	0.08	0.89	0.32	1.52
1	18	1.05	0.72	1.59	0.48	-0.09	0.90
1	19	-2.47	-3.30	-1.89	2.49	1.93	3.07
1	20	1.58	1.16	2.03	2.78	2.17	3.94
2	1	0.14	-0.49	0.69	1.03	0.41	1.66

Continued on next page

Table E.5 – continued from previous page

Latent Class	Item	n = 600			n = 2,400		
No.	No.	Mean	2.5%	97.5%	Mean	2.5%	97.5%
2	2	-0.38	-0.74	-0.03	1.86	1.07	2.77
2	3	-2.40	-2.77	-2.02	2.71	2.16	3.35
2	4	-1.07	-1.55	-0.61	2.70	1.94	3.18
2	5	-0.17	-0.51	0.42	2.18	1.06	2.87
2	6	-0.75	-1.20	-0.05	-0.18	-0.71	0.39
2	7	-0.96	-1.67	-0.25	-1.33	-1.88	-0.74
2	8	-1.47	-1.87	-1.07	-2.06	-2.82	-1.23
2	9	0.30	-0.29	0.90	1.68	1.30	2.02
2	10	-0.35	-1.00	0.34	2.15	1.85	2.47
2	11	-0.46	-0.94	-0.01	-0.12	-2.03	1.47
2	12	-0.18	-1.42	0.68	0.48	-0.16	1.44
2	13	0.29	-0.29	0.87	0.33	-0.76	1.77
2	14	1.53	1.09	1.93	-0.86	-1.67	0.98
2	15	0.68	-0.08	1.36	2.73	2.05	3.39
2	16	-2.66	-2.99	-2.39	0.62	-0.06	1.04
2	17	-0.90	-2.03	0.13	1.15	-0.30	1.89
2	18	-1.99	-2.68	-1.34	2.36	1.86	3.13
2	19	-1.15	-2.14	0.49	1.41	-0.32	2.33
2	20	-1.29	-2.43	-0.65	2.34	1.85	2.91
3	1	-0.51	-1.24	0.00	-0.59	-0.85	-0.21
3	2	-1.09	-1.81	-0.39	-0.29	-0.82	0.37
3	3	1.25	0.71	1.83	-0.35	-0.86	0.12

Continued on next page

Table E.5 – continued from previous page

Latent Class	Item	n = 600			n = 2,400		
No.	No.	Mean	2.5%	97.5%	Mean	2.5%	97.5%
3	4	1.77	1.21	2.26	-1.52	-2.44	-0.58
3	5	1.86	1.23	2.60	0.09	-0.24	0.53
3	6	0.67	-0.29	1.53	-0.50	-0.93	0.07
3	7	0.02	-0.62	0.94	-0.34	-1.21	0.12
3	8	1.73	1.19	2.33	-1.15	-1.94	-0.34
3	9	-0.91	-1.77	-0.24	1.18	0.25	2.10
3	10	0.53	-0.56	1.63	1.23	0.62	1.90
3	11	1.53	0.89	1.93	0.32	-0.43	1.09
3	12	0.58	0.10	1.12	-0.21	-0.76	0.76
3	13	2.01	1.64	2.42	-1.19	-2.05	0.60
3	14	2.42	2.00	2.84	0.87	0.17	1.45
3	15	0.91	-0.40	1.92	0.88	-0.34	1.87
3	16	1.48	0.90	2.22	-2.99	-3.44	-2.20
3	17	1.07	0.08	1.74	-2.44	-2.85	-2.09
3	18	1.15	0.62	1.61	-2.38	-2.95	-1.47
3	19	1.83	1.44	2.34	-1.79	-2.38	-0.88
3	20	2.16	1.60	2.61	-1.18	-2.83	0.33

Table E.6: Bayesian Credible Intervals for Mix3PLM with 20-items and 4-group

Latent Class	Item	n = 600			n = 2,400		
No.	No.	Mean	2.5%	97.5%	Mean	2.5%	97.5%
1	1	-0.27	-0.67	0.24	-1.18	-1.74	-0.63
1	2	-1.29	-2.20	-0.23	-0.57	-1.12	0.41
1	3	-0.19	-0.74	0.14	-0.01	-0.57	0.66
1	4	1.01	0.62	1.36	0.49	-0.73	1.36
1	5	-0.06	-0.51	0.93	-1.67	-2.42	-1.07
1	6	0.16	-0.29	0.57	-0.32	-0.70	0.15
1	7	1.92	1.03	2.60	-2.42	-3.29	-1.93
1	8	0.38	-0.37	0.83	0.69	0.19	1.31
1	9	0.40	-0.18	1.01	0.28	-0.28	1.18
1	10	-0.02	-0.36	0.38	-0.11	-0.79	0.78
1	11	-2.15	-2.62	-1.71	0.36	-0.24	1.22
1	12	-1.49	-2.27	-0.93	0.65	0.32	1.16
1	13	-1.36	-1.98	-0.94	0.21	-1.05	1.38
1	14	-0.86	-1.75	0.09	2.09	1.62	2.77
1	15	1.43	1.10	1.91	1.11	-0.25	2.53
1	16	0.70	0.02	1.44	-2.77	-4.03	-0.69
1	17	-1.29	-1.84	-0.58	-0.42	-1.29	0.21
1	18	-0.16	-1.27	0.72	-0.40	-1.19	0.11
1	19	-0.32	-1.39	1.29	-1.67	-2.29	-1.19
1	20	0.59	-0.09	1.34	0.77	0.07	1.21
2	1	0.07	-0.26	0.35	-1.40	-1.89	-0.62

Continued on next page

Table E.6 – continued from previous page

Latent Class No.	Item No.	n = 600 Mean	2.5%	97.5%	n = 2,400 Mean	2.5%	97.5%
2	2	0.14	-0.16	0.52	-0.23	-0.95	0.30
2	3	1.43	0.92	2.06	-0.83	-1.54	-0.21
2	4	1.32	0.80	1.75	0.88	-0.07	1.68
2	5	1.95	1.57	2.31	-0.07	-0.49	0.23
2	6	-2.01	-2.38	-1.44	-3.60	-4.10	-3.10
2	7	0.20	-0.53	0.93	-3.56	-4.30	-2.23
2	8	-0.24	-0.58	0.12	-3.27	-3.76	-2.91
2	9	-0.44	-0.80	-0.08	-2.99	-3.54	-2.25
2	10	0.06	-0.42	0.45	0.96	0.61	1.37
2	11	0.63	0.22	1.02	0.18	-0.19	0.60
2	12	-0.65	-2.13	0.71	0.38	-0.70	0.97
2	13	0.91	0.37	1.52	-1.46	-1.88	-1.01
2	14	1.88	1.11	2.40	1.17	-0.58	2.11
2	15	1.48	1.05	1.87	0.30	-0.41	1.09
2	16	0.08	-1.44	1.26	0.47	-0.09	1.25
2	17	1.23	0.65	1.90	-2.07	-2.54	-1.58
2	18	1.21	0.73	1.64	-0.07	-1.01	0.87
2	19	1.38	0.60	2.13	0.44	-0.08	0.82
2	20	1.52	0.90	2.23	0.65	-0.50	1.22
3	1	-2.00	-2.56	-1.33	2.56	1.65	3.16
3	2	0.22	-0.66	1.42	1.61	1.07	2.10
3	3	-1.04	-1.53	-0.52	1.17	0.86	1.58

Continued on next page

Table E.6 – continued from previous page

Latent Class	Item	n = 600			n = 2,400		
No.	No.	Mean	2.5%	97.5%	Mean	2.5%	97.5%
3	4	0.31	-0.19	0.93	1.71	0.89	2.19
3	5	0.42	-0.21	1.09	2.02	1.57	2.56
3	6	-2.91	-3.37	-2.52	0.65	-0.34	1.50
3	7	-0.03	-0.63	0.47	-0.48	-1.30	0.29
3	8	-0.33	-2.09	0.86	-1.10	-1.91	-0.54
3	9	-2.66	-3.33	-2.11	0.14	-0.43	0.65
3	10	-2.68	-3.19	-2.17	0.86	0.00	2.09
3	11	-0.80	-1.38	-0.38	-1.14	-1.84	-0.28
3	12	-0.35	-1.18	0.19	-0.87	-1.93	0.20
3	13	0.94	0.40	1.41	1.32	0.34	2.10
3	14	-1.02	-2.09	-0.30	1.12	-0.38	2.07
3	15	-0.17	-0.79	0.28	1.92	1.26	2.74
3	16	0.54	-0.02	0.97	1.36	0.80	1.95
3	17	-0.18	-0.51	0.19	1.15	0.60	1.71
3	18	-0.22	-0.67	0.69	2.73	1.65	3.26
3	19	0.05	-0.73	0.64	3.29	2.48	3.98
3	20	-1.10	-1.70	-0.46	2.30	1.88	2.85
4	1	-0.10	-0.85	0.41	0.32	-0.62	1.11
4	2	-0.08	-0.49	0.31	0.17	-0.54	0.85
4	3	-0.54	-0.91	-0.15	0.11	-0.38	0.49
4	4	1.08	0.60	1.33	0.36	-0.14	0.85
4	5	0.05	-0.39	0.60	0.63	0.13	1.20

Continued on next page

Table E.6 – continued from previous page

Latent Class	Item	n = 600			n = 2,400		
No.	No.	Mean	2.5%	97.5%	Mean	2.5%	97.5%
4	6	-0.67	-1.19	0.51	0.52	-0.03	1.01
4	7	0.44	-0.11	1.21	1.22	0.68	1.81
4	8	0.38	-0.20	1.04	1.68	1.19	2.35
4	9	-0.91	-1.37	-0.57	1.04	0.00	1.67
4	10	1.18	0.65	1.92	0.98	0.38	1.47
4	11	0.22	-0.20	0.90	-0.79	-1.10	-0.19
4	12	-1.15	-2.26	0.02	-0.75	-1.08	-0.29
4	13	0.79	0.44	1.30	-0.67	-1.06	-0.12
4	14	1.84	0.88	2.56	-0.61	-1.02	-0.11
4	15	0.70	0.33	1.34	-0.57	-0.84	-0.12
4	16	-1.88	-2.58	-1.23	0.99	0.42	1.79
4	17	-2.96	-3.48	-2.36	0.27	-0.61	1.21
4	18	0.42	-0.73	1.27	-1.40	-2.50	-0.76
4	19	-1.36	-1.88	-0.76	-0.30	-0.84	0.10
4	20	1.14	0.57	1.93	1.07	0.62	1.49

Table E.7: Bayesian Credible Intervals for Mix3PLM with 40-items and 1-group

Latent Class	Item	n = 600			n = 2,400		
No.	No.	Mean	2.5%	97.5%	Mean	2.5%	97.5%
1	1	-1.07	-1.74	-0.44	1.53	0.96	1.93
1	2	-0.71	-1.35	-0.19	-2.02	-2.42	-1.25
1	3	-0.16	-0.58	0.28	-1.73	-2.24	-1.19
1	4	0.82	0.26	1.39	0.67	0.29	1.21
1	5	-0.54	-1.81	0.49	-1.40	-1.96	-0.92
1	6	1.97	1.04	2.86	-1.12	-1.56	-0.59
1	7	-1.19	-2.13	-0.23	-3.17	-3.79	-2.42
1	8	-1.41	-2.07	-0.39	-0.78	-1.29	-0.47
1	9	0.63	-0.56	1.32	0.10	-0.46	0.61
1	10	0.19	-0.65	1.02	0.46	-0.10	1.08
1	11	0.10	-0.39	0.51	1.17	-0.29	2.22
1	12	0.52	0.01	0.88	1.88	1.41	2.35
1	13	0.46	-0.15	0.80	0.98	0.37	1.44
1	14	0.36	-0.37	1.07	-0.07	-0.44	0.40
1	15	1.20	0.42	1.64	1.64	0.92	2.97
1	16	1.02	0.30	1.57	-1.25	-2.25	-0.69
1	17	0.58	-0.39	1.20	0.37	-0.09	0.76
1	18	0.32	0.02	0.77	0.51	0.07	1.05
1	19	0.00	-0.79	0.77	-0.99	-1.48	-0.34
1	20	-0.65	-1.17	0.09	-1.74	-2.25	-1.03
1	21	-0.35	-1.00	0.38	1.01	0.23	1.58

Continued on next page

Table E.7 – continued from previous page

Latent Class	Item	n = 600			n = 2,400		
No.	No.	Mean	2.5%	97.5%	Mean	2.5%	97.5%
1	22	-0.16	-1.01	0.30	0.08	-0.44	0.88
1	23	0.40	-0.33	1.01	-0.32	-1.00	0.55
1	24	3.00	2.21	3.66	0.58	0.08	1.06
1	25	0.81	0.16	1.50	0.60	0.03	1.02
1	26	-1.47	-1.85	-1.01	0.13	-1.11	1.08
1	27	-1.40	-2.29	-0.66	0.36	-0.97	1.72
1	28	-0.54	-1.03	0.01	-0.10	-0.41	0.26
1	29	-0.44	-1.30	0.14	-1.53	-2.14	-0.83
1	30	-1.15	-1.92	-0.35	1.20	0.41	2.03
1	31	0.87	0.42	1.36	0.70	0.30	1.20
1	32	1.00	0.39	1.39	0.51	-0.33	1.48
1	33	0.16	-0.19	0.46	1.42	0.64	1.90
1	34	1.29	0.65	1.92	1.02	0.37	1.72
1	35	-0.70	-1.49	0.42	0.84	-0.07	1.74
1	36	0.11	-0.29	0.50	0.13	-1.05	1.28
1	37	0.28	-0.55	1.17	-1.16	-1.60	-0.66
1	38	0.70	0.08	2.07	0.97	0.44	1.58
1	39	1.95	1.18	2.81	-0.38	-0.88	0.00
1	40	1.44	0.84	2.35	0.21	-0.48	0.97

Table E.8: Bayesian Credible Intervals for Mix3PLM with 40-items and 2-group

Latent Class	Item	n = 600			n = 2,400		
No.	No.	Mean	2.5%	97.5%	Mean	2.5%	97.5%
1	1	0.08	-0.37	0.86	-0.18	-0.82	0.48
1	2	0.06	-0.87	1.00	-0.02	-0.32	0.29
1	3	0.31	-0.09	0.62	0.85	0.20	1.33
1	4	0.82	0.44	1.21	1.28	0.81	2.01
1	5	1.24	0.53	1.95	1.49	0.73	2.73
1	6	-0.40	-0.79	-0.11	-1.05	-1.47	-0.26
1	7	-0.24	-0.57	0.22	0.17	-0.56	0.66
1	8	0.05	-0.35	0.32	0.70	0.16	1.08
1	9	0.30	-0.12	1.06	1.32	0.78	1.72
1	10	0.70	0.22	1.17	2.72	1.10	3.81
1	11	0.72	0.35	1.10	0.84	0.48	1.29
1	12	1.13	0.25	1.65	0.56	-0.10	1.27
1	13	0.88	0.41	1.38	1.56	0.99	2.14
1	14	1.60	1.06	2.14	1.43	0.58	2.08
1	15	1.41	0.86	1.83	1.81	1.17	2.33
1	16	0.20	-0.18	0.79	-1.05	-1.73	-0.42
1	17	0.11	-0.44	0.69	0.38	-0.63	1.32
1	18	1.06	0.44	1.94	0.01	-0.54	0.73
1	19	-0.12	-0.64	0.45	-0.37	-1.01	0.02
1	20	-0.26	-0.62	0.13	0.09	-0.38	0.70
1	21	-0.42	-0.78	-0.06	-0.88	-1.66	0.16

Continued on next page

Table E.8 – continued from previous page

Latent Class	Item	n = 600			n = 2,400		
No.	No.	Mean	2.5%	97.5%	Mean	2.5%	97.5%
1	22	0.29	-0.38	0.87	0.85	-0.38	2.38
1	23	1.20	0.50	1.72	0.56	0.16	0.89
1	24	0.71	-0.36	1.71	2.12	1.18	3.01
1	25	1.89	1.35	2.58	-0.43	-1.79	1.39
1	26	-0.38	-0.78	0.08	-0.04	-0.46	0.42
1	27	-0.38	-1.22	0.78	-0.21	-0.64	0.42
1	28	-0.31	-0.89	0.37	0.61	0.17	1.01
1	29	0.24	-0.09	0.53	0.92	0.42	1.34
1	30	0.29	-0.19	0.76	-1.33	-1.97	-0.67
1	31	1.83	1.15	3.16	0.24	-0.38	0.83
1	32	0.31	-0.18	0.67	1.44	-0.58	2.33
1	33	0.85	-0.57	1.53	-1.52	-1.92	-1.13
1	34	1.78	1.03	2.34	2.06	1.44	2.77
1	35	1.27	0.49	2.40	1.94	1.49	2.43
1	36	0.66	0.18	1.12	-0.36	-0.79	0.61
1	37	0.90	-0.30	1.65	-0.31	-0.73	0.28
1	38	-1.25	-1.69	-0.61	-0.44	-0.89	-0.08
1	39	-0.89	-1.53	-0.03	0.18	-0.59	0.91
1	40	0.52	0.19	0.98	-0.50	-1.38	0.10
2	1	-1.08	-2.06	-0.22	-0.17	-0.67	0.28
2	2	-2.32	-2.79	-1.91	-0.48	-0.96	0.15
2	3	1.40	0.77	1.94	0.08	-0.29	0.60

Continued on next page

Table E.8 – continued from previous page

Latent Class	Item	n = 600			n = 2,400		
No.	No.	Mean	2.5%	97.5%	Mean	2.5%	97.5%
2	4	0.74	0.12	1.28	0.59	-0.04	1.14
2	5	-0.27	-1.28	0.79	1.36	0.17	2.42
2	6	1.41	0.66	2.23	-4.08	-4.57	-3.34
2	7	-0.49	-0.84	0.04	0.18	-1.19	1.83
2	8	-0.33	-1.24	0.45	-2.45	-2.91	-1.48
2	9	-0.71	-1.73	0.20	-2.09	-2.92	-1.10
2	10	-1.89	-2.69	-1.13	-1.50	-2.49	-0.66
2	11	-0.99	-2.06	0.10	-0.98	-1.60	0.08
2	12	-1.41	-1.74	-1.06	0.16	-0.42	0.67
2	13	0.07	-0.47	0.57	0.05	-0.80	0.92
2	14	-0.52	-0.93	0.16	0.38	-0.58	1.10
2	15	1.44	1.02	1.80	0.67	0.02	1.43
2	16	-1.04	-1.66	-0.39	0.94	0.45	1.36
2	17	-1.04	-1.37	-0.71	0.47	-0.59	1.13
2	18	-0.54	-0.90	-0.05	1.40	1.08	1.79
2	19	-0.50	-0.87	-0.17	1.44	-0.04	2.17
2	20	-0.52	-1.20	0.54	2.61	2.07	3.26
2	21	1.59	1.00	1.99	-1.16	-1.70	-0.45
2	22	-0.72	-1.32	-0.23	-0.72	-1.27	-0.22
2	23	0.17	-0.57	0.67	-0.53	-1.00	0.34
2	24	0.23	-0.17	0.63	0.94	0.43	1.78
2	25	-0.78	-1.21	-0.38	0.04	-0.35	0.62

Continued on next page

Table E.8 – continued from previous page

Latent Class	Item	n = 600			n = 2,400		
No.	No.	Mean	2.5%	97.5%	Mean	2.5%	97.5%
2	26	0.39	-0.27	1.42	-2.96	-3.58	-2.20
2	27	0.24	-0.19	0.75	-3.31	-3.76	-2.98
2	28	0.29	-0.28	0.69	-2.37	-2.95	-1.08
2	29	-1.65	-2.41	-0.66	-1.72	-2.32	-1.12
2	30	-2.51	-3.48	-1.77	-0.99	-1.73	-0.26
2	31	-1.44	-2.12	-0.96	0.66	-0.12	1.28
2	32	-0.63	-1.12	-0.14	0.40	-0.13	0.87
2	33	-0.17	-0.79	0.48	-0.34	-1.07	0.24
2	34	-1.29	-2.06	-0.76	-1.07	-1.80	-0.24
2	35	-0.84	-1.44	-0.43	0.38	-0.19	0.94
2	36	0.42	-0.10	1.02	0.11	-0.50	0.69
2	37	0.29	-0.33	1.05	1.04	0.29	1.71
2	38	-0.14	-0.83	0.61	1.50	0.87	2.29
2	39	0.32	-1.16	1.28	2.42	1.33	3.01
2	40	0.29	-0.30	1.03	2.33	1.86	3.01

Table E.9: Bayesian Credible Intervals for Mix3PLM with 40-items and 3-group

Latent Class	Item	n = 600			n = 2,400		
No.	No.	Mean	2.5%	97.5%	Mean	2.5%	97.5%
1	1	1.80	1.12	2.37	0.29	-0.15	0.70
1	2	0.51	0.16	0.89	0.92	0.52	1.44
1	3	0.43	-0.09	1.01	1.72	1.26	2.17
1	4	2.26	1.37	2.69	1.90	1.43	2.51
1	5	1.24	0.84	1.54	1.80	1.47	2.24
1	6	0.11	-0.81	0.78	-0.16	-0.54	0.25
1	7	0.77	0.23	1.38	-1.31	-1.95	-0.69
1	8	0.78	-0.35	1.64	-0.83	-1.52	0.05
1	9	1.37	1.02	1.93	1.36	0.89	1.82
1	10	0.81	0.39	1.28	1.86	1.46	2.28
1	11	0.39	0.00	0.68	0.59	0.22	0.91
1	12	1.35	0.91	1.77	1.32	0.75	1.71
1	13	1.06	0.80	1.55	0.58	-0.48	1.04
1	14	-0.36	-1.45	1.23	1.39	0.91	1.86
1	15	0.64	0.17	1.38	2.38	1.27	3.47
1	16	0.71	-0.20	1.52	1.04	-0.06	1.75
1	17	1.41	0.91	2.05	0.87	0.40	2.05
1	18	0.59	-0.26	1.61	0.38	-1.95	1.96
1	19	2.02	1.56	2.55	0.03	-1.26	1.47
1	20	2.40	1.74	2.92	2.31	1.95	2.68
1	21	0.24	-0.43	0.73	-1.93	-2.21	-1.56

Continued on next page

Table E.9 – continued from previous page

Latent Class No.	Item No.	n = 600 Mean	2.5%	97.5%	n = 2,400 Mean	2.5%	97.5%
1	22	0.86	0.49	1.27	1.65	1.18	2.00
1	23	0.77	-0.74	2.13	1.83	1.50	2.31
1	24	0.70	0.23	1.40	2.36	1.75	2.99
1	25	1.95	1.22	2.60	2.31	1.72	2.80
1	26	-0.54	-0.98	-0.03	0.02	-0.80	0.82
1	27	0.60	-0.16	1.22	0.77	0.38	1.39
1	28	-0.63	-1.58	-0.02	0.07	-0.72	0.86
1	29	-1.35	-1.96	-0.59	1.43	0.81	2.04
1	30	0.82	0.53	1.09	-1.21	-1.83	-0.06
1	31	1.43	0.89	2.20	1.34	0.47	2.00
1	32	0.44	-0.24	1.16	1.22	0.13	1.83
1	33	0.52	0.23	0.92	1.53	1.02	2.11
1	34	0.96	0.57	1.79	1.30	0.88	1.81
1	35	2.21	1.88	2.48	2.10	1.27	2.92
1	36	-0.99	-1.74	-0.19	1.28	0.68	2.16
1	37	1.08	0.54	1.91	1.52	1.07	1.92
1	38	1.13	0.62	1.73	-0.84	-1.40	0.25
1	39	2.07	1.07	2.76	1.54	0.94	2.07
1	40	2.06	1.64	2.49	2.05	1.71	2.38
2	1	0.52	-0.33	1.01	-1.13	-2.27	-0.57
2	2	-0.50	-1.38	-0.03	-0.19	-0.87	0.38
2	3	-0.65	-1.11	-0.22	-0.43	-1.26	0.28

Continued on next page

Table E.9 – continued from previous page

Latent Class	Item	n = 600			n = 2,400		
No.	No.	Mean	2.5%	97.5%	Mean	2.5%	97.5%
2	4	0.53	-0.02	0.97	0.37	-0.26	0.96
2	5	-0.27	-0.80	0.28	-0.67	-1.40	-0.03
2	6	-2.98	-3.68	-1.72	0.81	0.28	1.28
2	7	-2.72	-3.42	-1.59	-1.20	-1.85	-0.50
2	8	-2.66	-3.01	-2.22	-0.88	-1.61	-0.44
2	9	-2.83	-3.28	-2.35	-0.10	-0.63	0.18
2	10	-2.03	-2.60	-1.49	-0.34	-0.73	0.12
2	11	1.01	0.45	1.50	1.30	0.95	1.76
2	12	0.06	-0.98	1.27	-0.41	-1.30	0.42
2	13	-1.07	-1.64	-0.45	0.80	-0.05	1.48
2	14	-0.12	-0.43	0.28	0.17	-0.59	0.81
2	15	-0.03	-0.66	0.52	-0.75	-2.41	1.33
2	16	-0.88	-1.44	-0.44	-2.35	-3.39	-0.76
2	17	1.03	0.58	1.51	1.13	0.00	1.99
2	18	0.90	0.56	1.35	-3.39	-3.78	-2.95
2	19	0.10	-1.38	1.18	-2.14	-2.73	-1.46
2	20	-0.13	-1.38	1.17	-2.16	-2.97	-0.89
2	21	-0.30	-1.34	0.89	0.06	-1.21	0.77
2	22	-1.49	-2.02	-0.99	0.07	-1.35	1.16
2	23	0.10	-0.70	0.77	1.89	0.78	2.76
2	24	1.00	-0.04	1.93	1.34	0.91	1.78
2	25	1.62	1.20	2.34	-2.39	-2.77	-2.08

Continued on next page

Table E.9 – continued from previous page

Latent Class	Item	n = 600			n = 2,400		
No.	No.	Mean	2.5%	97.5%	Mean	2.5%	97.5%
2	26	-2.94	-3.40	-2.25	-0.72	-1.76	0.67
2	27	-3.04	-3.82	-2.46	1.44	0.74	2.13
2	28	-2.17	-3.63	-0.72	0.06	-0.66	0.83
2	29	1.36	0.66	2.16	0.23	-0.17	0.62
2	30	-0.56	-1.35	0.24	1.13	0.29	1.70
2	31	0.07	-0.70	1.03	0.72	0.02	1.46
2	32	-0.56	-1.02	0.28	-0.11	-0.54	0.50
2	33	-0.16	-1.40	0.55	0.77	0.25	1.21
2	34	-1.16	-2.09	0.12	0.64	0.08	1.25
2	35	0.63	0.10	1.32	-0.16	-0.59	0.33
2	36	-0.39	-1.37	0.40	-2.39	-3.00	-1.10
2	37	0.76	0.45	1.22	-2.95	-3.38	-2.34
2	38	-0.48	-2.03	0.64	-1.98	-3.51	-0.38
2	39	0.88	0.36	1.35	-3.15	-3.66	-2.76
2	40	1.52	1.06	1.99	-0.67	-1.83	0.58
3	1	0.08	-0.88	1.03	-0.65	-1.75	0.56
3	2	-1.44	-1.99	-0.86	0.23	-0.93	1.16
3	3	-0.39	-1.58	0.39	-2.46	-3.34	-1.27
3	4	2.09	1.58	2.57	-0.73	-1.56	0.02
3	5	1.19	0.04	2.10	-0.64	-0.94	-0.18
3	6	-0.46	-1.22	0.14	-1.77	-3.53	0.60
3	7	-0.89	-1.31	-0.38	-3.68	-4.04	-3.23

Continued on next page

Table E.9 – continued from previous page

Latent Class	Item	n = 600			n = 2,400		
No.	No.	Mean	2.5%	97.5%	Mean	2.5%	97.5%
3	8	-0.16	-0.99	0.52	-1.94	-3.31	-0.65
3	9	0.10	-0.59	0.67	-3.40	-3.73	-3.00
3	10	0.75	-0.09	1.47	-1.85	-2.65	-1.04
3	11	0.77	0.13	1.50	-0.15	-0.63	0.50
3	12	0.43	-0.24	0.94	0.29	-0.10	0.61
3	13	0.23	-0.86	1.22	0.69	0.26	1.16
3	14	-0.21	-0.76	0.32	0.35	0.01	0.96
3	15	2.41	1.51	3.51	1.06	-0.27	1.89
3	16	-1.50	-2.02	0.12	0.28	-0.35	0.85
3	17	-2.15	-2.61	-1.39	-0.82	-1.74	-0.05
3	18	-1.51	-1.82	-1.18	-0.34	-2.17	0.88
3	19	-1.32	-1.65	-1.03	2.00	1.56	2.49
3	20	-1.02	-1.53	-0.37	2.08	1.35	2.35
3	21	-0.61	-1.08	-0.01	-0.71	-1.61	1.00
3	22	-1.35	-1.84	-0.92	0.36	-0.97	1.16
3	23	-0.45	-0.83	0.13	-0.56	-0.99	0.00
3	24	0.68	0.10	1.31	0.18	-0.58	0.78
3	25	-0.40	-0.73	-0.16	-0.48	-0.98	-0.17
3	26	-0.65	-1.15	-0.24	-3.70	-3.99	-3.42
3	27	0.29	-0.08	0.83	-3.26	-3.84	-1.82
3	28	-0.35	-0.83	0.15	-3.69	-4.21	-3.21
3	29	-0.38	-1.35	0.43	-3.42	-4.23	-1.75

Continued on next page

Table E.9 – continued from previous page

Latent Class	Item	n = 600			n = 2,400		
No.	No.	Mean	2.5%	97.5%	Mean	2.5%	97.5%
3	30	-0.03	-0.27	0.32	-2.05	-3.03	-0.92
3	31	-0.05	-0.57	0.52	0.78	-0.13	1.52
3	32	0.59	-0.04	1.22	0.70	0.05	1.20
3	33	0.41	0.07	0.85	-0.94	-1.64	-0.24
3	34	1.46	0.98	1.92	-0.16	-1.43	0.77
3	35	2.10	1.60	2.55	-0.94	-1.38	-0.49
3	36	-2.76	-3.02	-2.48	0.23	-0.29	0.73
3	37	-1.91	-2.40	-1.37	1.64	1.25	2.29
3	38	-0.41	-2.05	1.19	0.12	-1.65	1.40
3	39	-1.75	-2.22	-1.27	2.30	1.46	2.77
3	40	-1.29	-1.68	-0.64	2.20	1.63	2.93

Table E.10: Bayesian Credible Intervals for Mix3PLM with 40-items and 4-group

Latent Class	Item	n = 600			n = 2,400		
No.	No.	Mean	2.5%	97.5%	Mean	2.5%	97.5%
1	1	1.26	0.88	1.61	0.20	-0.46	1.01
1	2	0.29	-1.20	0.97	-0.61	-1.29	0.10
1	3	1.87	1.41	2.40	-0.85	-1.41	0.06
1	4	2.60	1.95	3.07	-0.42	-1.54	0.65
1	5	1.86	1.48	2.29	-0.53	-1.21	0.11
1	6	0.72	0.25	1.14	1.30	0.00	2.52
1	7	0.76	0.31	1.41	-0.53	-1.64	0.23
1	8	0.64	-0.20	1.51	0.19	-0.11	0.44
1	9	1.75	1.07	2.45	-0.37	-0.61	-0.05
1	10	1.94	1.47	2.61	0.73	0.15	1.43
1	11	0.84	-0.08	1.36	-0.67	-1.58	0.23
1	12	0.17	-0.76	0.84	0.72	0.10	1.30
1	13	1.35	0.85	1.80	0.42	-0.61	1.28
1	14	2.60	1.72	3.32	1.58	0.81	2.35
1	15	1.68	1.16	2.12	1.04	0.52	1.44
1	16	0.05	-0.55	0.78	-1.96	-2.72	-0.67
1	17	0.67	-0.29	1.54	-2.60	-2.98	-2.14
1	18	0.92	0.17	1.31	-2.09	-2.64	-1.52
1	19	1.17	-0.41	2.34	-1.48	-2.45	-0.44
1	20	1.31	0.85	1.88	-1.75	-2.24	-0.71
1	21	0.16	-0.54	0.92	-0.36	-0.93	0.21

Continued on next page

Table E.10 – continued from previous page

Latent Class	Item	n = 600			n = 2,400		
No.	No.	Mean	2.5%	97.5%	Mean	2.5%	97.5%
1	22	0.76	-0.37	1.38	-0.96	-1.53	-0.30
1	23	1.39	0.59	1.89	0.01	-1.19	0.78
1	24	1.70	0.94	2.30	0.80	-0.02	2.06
1	25	1.23	0.25	1.76	-0.47	-0.98	0.21
1	26	1.12	0.29	1.68	-0.70	-1.03	-0.18
1	27	-0.82	-1.25	-0.27	-0.09	-0.62	0.51
1	28	0.00	-0.91	1.12	0.46	-0.63	1.39
1	29	1.26	0.82	1.87	0.42	-0.42	1.42
1	30	1.11	-0.19	2.03	0.18	-0.10	0.52
1	31	0.93	-0.29	1.48	-0.25	-0.64	0.18
1	32	1.00	-0.15	1.65	-0.62	-1.20	0.31
1	33	1.72	1.31	2.20	0.44	0.05	0.90
1	34	1.52	1.20	1.85	-0.91	-1.79	0.38
1	35	-1.33	-1.89	-0.44	-2.34	-2.99	-1.61
1	36	-0.62	-0.98	-0.11	-2.78	-3.16	-2.38
1	37	0.90	0.31	1.30	-2.28	-2.59	-1.81
1	38	1.32	0.57	2.10	-2.25	-2.63	-1.88
1	39	0.53	-0.93	1.81	-0.51	-1.85	1.19
1	40	1.77	1.49	2.05	-1.52	-2.00	-1.09
2	1	-1.80	-2.34	-1.39	-1.16	-1.67	-0.76
2	2	-1.06	-1.69	-0.37	0.37	-0.03	0.74
2	3	0.43	-0.06	0.86	0.92	0.36	1.23

Continued on next page

Table E.10 – continued from previous page

Latent Class	Item	n = 600			n = 2,400		
No.	No.	Mean	2.5%	97.5%	Mean	2.5%	97.5%
2	4	0.75	0.37	1.10	1.36	1.04	1.88
2	5	0.33	-0.50	1.28	0.49	-1.54	1.73
2	6	-0.49	-1.05	-0.08	-1.19	-2.01	0.64
2	7	0.46	-0.87	1.60	-2.19	-2.53	-1.46
2	8	0.99	0.33	1.50	-2.17	-2.94	-1.45
2	9	0.40	-1.00	1.11	-0.92	-1.42	-0.40
2	10	0.66	0.26	1.05	-0.38	-0.67	-0.04
2	11	1.10	0.40	1.89	0.24	-0.02	0.66
2	12	-0.49	-1.88	1.03	0.55	0.10	0.98
2	13	-1.33	-2.05	-0.72	0.90	0.19	1.55
2	14	-1.66	-2.15	-1.03	-0.16	-1.67	1.39
2	15	-0.16	-1.47	0.49	-0.67	-1.51	0.12
2	16	-0.90	-1.53	-0.05	2.06	1.60	2.49
2	17	0.91	0.19	1.54	-0.32	-1.08	0.31
2	18	-0.46	-0.99	0.22	2.36	1.98	2.73
2	19	-0.68	-1.66	0.18	2.07	1.76	2.37
2	20	-0.38	-1.06	0.17	3.08	2.76	3.45
2	21	-0.60	-1.07	0.24	0.32	-0.10	0.76
2	22	0.24	-0.06	0.67	-1.91	-2.52	-1.43
2	23	0.26	-0.46	1.04	0.90	0.37	1.53
2	24	-0.81	-1.53	-0.43	1.54	0.99	1.98
2	25	1.24	0.63	1.86	1.73	1.36	2.30

Continued on next page

Table E.10 – continued from previous page

Latent Class	Item	n = 600			n = 2,400		
No.	No.	Mean	2.5%	97.5%	Mean	2.5%	97.5%
2	26	0.64	0.19	1.17	-2.14	-2.76	-1.55
2	27	-1.54	-2.36	-0.85	-1.74	-2.45	-1.25
2	28	0.65	-0.42	1.49	-1.31	-1.89	-0.84
2	29	-0.14	-0.59	0.45	-0.67	-0.92	-0.41
2	30	-1.61	-2.02	-1.04	-0.15	-0.44	0.22
2	31	-1.48	-2.60	0.43	0.23	-0.40	0.97
2	32	0.25	-1.03	1.32	0.63	0.31	0.99
2	33	-1.56	-2.09	-0.88	1.16	0.65	1.59
2	34	-1.75	-2.23	-1.13	1.20	0.80	1.58
2	35	-1.04	-1.54	-0.54	1.82	1.31	2.38
2	36	-0.16	-0.76	0.37	1.76	1.27	2.18
2	37	0.54	0.07	1.06	2.70	2.17	3.34
2	38	-0.72	-1.28	0.04	1.83	1.44	2.10
2	39	-0.76	-1.43	-0.18	-0.36	-1.14	0.82
2	40	0.85	0.11	1.56	2.92	2.46	3.50
3	1	-1.13	-1.76	-0.56	-1.44	-2.06	-0.96
3	2	-0.42	-0.75	0.03	-2.37	-2.82	-2.00
3	3	0.42	-0.01	0.88	-1.18	-1.55	-0.80
3	4	-0.54	-0.94	0.13	-0.91	-1.39	-0.44
3	5	1.00	0.34	1.80	-0.53	-1.20	0.55
3	6	-0.13	-0.69	0.43	-2.88	-3.73	-2.44
3	7	-0.45	-1.31	0.83	-3.04	-3.81	-2.49

Continued on next page

Table E.10 – continued from previous page

Latent Class	Item	n = 600			n = 2,400		
No.	No.	Mean	2.5%	97.5%	Mean	2.5%	97.5%
3	8	0.25	-0.28	0.74	-2.45	-2.86	-1.33
3	9	0.63	-0.01	1.26	1.24	0.00	2.09
3	10	-0.39	-0.82	0.49	-1.72	-2.35	-1.19
3	11	0.54	-0.52	1.22	-2.19	-2.55	-1.75
3	12	2.99	2.29	3.40	0.18	-0.34	0.72
3	13	0.49	-0.38	0.95	-0.63	-1.04	-0.23
3	14	1.47	1.05	1.81	-0.89	-1.85	1.16
3	15	1.00	-0.25	1.71	-0.34	-1.15	0.17
3	16	-1.66	-2.04	-1.34	-0.48	-0.72	-0.15
3	17	-1.88	-2.17	-1.48	-0.11	-0.51	0.28
3	18	-1.71	-2.21	-1.17	0.34	-0.14	0.79
3	19	-1.54	-2.11	-1.13	0.65	0.04	1.04
3	20	-1.78	-2.20	-1.44	0.74	-0.11	1.26
3	21	0.91	0.17	1.69	-0.44	-1.24	0.10
3	22	-0.85	-1.41	-0.32	-1.15	-1.53	-0.81
3	23	0.08	-1.04	0.90	-0.77	-1.56	-0.04
3	24	0.97	0.29	1.67	-1.00	-1.53	-0.56
3	25	-0.32	-1.05	0.36	0.21	-0.43	0.63
3	26	0.16	-0.22	0.45	-3.24	-3.79	-2.80
3	27	-0.72	-1.30	0.05	-2.42	-2.79	-2.03
3	28	0.00	-0.40	0.50	-1.88	-3.33	0.02
3	29	-0.18	-0.61	0.18	-2.95	-3.36	-2.47

Continued on next page

Table E.10 – continued from previous page

Latent Class	Item	n = 600			n = 2,400		
No.	No.	Mean	2.5%	97.5%	Mean	2.5%	97.5%
3	30	-1.35	-2.74	-0.62	-2.02	-2.65	-1.21
3	31	-0.07	-0.77	0.58	-1.36	-1.72	-1.03
3	32	-0.11	-0.98	0.60	-1.24	-1.65	-0.94
3	33	0.28	-0.52	0.96	-0.47	-1.66	0.23
3	34	2.34	1.38	2.93	-1.05	-1.43	-0.55
3	35	0.20	-0.90	1.29	-0.40	-0.69	0.15
3	36	-0.94	-2.33	0.94	-0.56	-0.90	-0.21
3	37	-1.64	-2.31	-1.09	0.78	0.01	1.55
3	38	-0.61	-1.62	1.09	0.10	-0.77	0.79
3	39	-1.23	-1.68	-0.79	1.24	0.17	1.78
3	40	-1.81	-3.02	-0.89	0.66	0.24	1.02
4	1	-1.07	-1.52	-0.56	1.31	0.46	2.07
4	2	-0.14	-1.02	1.10	1.48	0.98	2.15
4	3	-1.83	-2.52	-0.92	0.48	-0.52	1.83
4	4	-0.40	-1.09	0.39	2.82	2.47	3.43
4	5	-0.03	-0.62	0.53	2.24	1.87	2.54
4	6	-2.11	-2.51	-1.62	1.48	0.99	2.19
4	7	-2.49	-2.90	-1.85	1.05	0.36	1.79
4	8	-2.07	-2.53	-1.75	1.59	1.29	1.86
4	9	-1.43	-2.04	-0.58	2.07	1.71	2.47
4	10	-1.75	-2.36	-0.88	2.62	1.87	3.36
4	11	-1.93	-2.44	-1.32	0.35	-0.44	0.97

Continued on next page

Table E.10 – continued from previous page

Latent Class No.	Item No.	n = 600 Mean	n = 600 2.5%	n = 600 97.5%	n = 2,400 Mean	n = 2,400 2.5%	n = 2,400 97.5%
4	12	0.14	-0.16	0.50	1.99	0.67	3.29
4	13	1.44	0.61	1.87	0.99	0.52	1.36
4	14	1.07	0.58	1.61	1.47	1.05	1.97
4	15	0.98	0.58	1.50	1.64	-0.39	2.76
4	16	0.36	-0.25	1.20	0.41	-0.11	0.87
4	17	1.17	0.49	1.52	0.86	0.29	1.24
4	18	2.61	2.00	3.18	1.10	0.31	1.75
4	19	0.72	-0.53	1.56	1.95	1.50	2.41
4	20	0.29	-0.46	1.33	1.29	-0.22	2.37
4	21	-0.54	-1.04	0.31	0.05	-0.31	0.49
4	22	-1.89	-2.19	-1.46	1.72	1.25	2.34
4	23	0.93	0.42	1.34	0.77	-0.80	1.81
4	24	0.11	-0.52	1.06	2.42	0.94	3.14
4	25	0.76	0.18	1.28	2.67	2.25	3.14
4	26	-2.62	-3.54	-1.96	1.23	0.92	1.67
4	27	-2.29	-3.14	-1.68	1.90	1.44	2.33
4	28	-2.43	-3.25	-1.67	1.99	1.61	2.53
4	29	-0.96	-2.01	0.53	1.96	1.60	2.39
4	30	-1.36	-1.89	-0.69	2.52	2.16	3.34
4	31	-1.50	-1.93	-1.03	0.72	-0.24	1.27
4	32	-0.15	-0.81	0.84	1.21	0.69	1.66
4	33	0.46	-0.25	1.29	0.07	-0.33	0.86

Continued on next page

Table E.10 – continued from previous page

Latent Class	Item	n = 600			n = 2,400		
No.	No.	Mean	2.5%	97.5%	Mean	2.5%	97.5%
4	34	0.18	-0.68	1.19	0.76	0.12	1.59
4	35	-0.35	-1.18	0.60	0.22	-0.88	1.54
4	36	0.49	0.07	1.14	1.63	1.20	2.25
4	37	1.60	0.89	2.19	-1.58	-2.33	-0.88
4	38	0.80	-0.23	1.75	-0.04	-1.02	1.51
4	39	1.53	-0.03	2.43	-0.31	-1.47	1.36
4	40	-1.91	-2.95	-1.27	-1.29	-2.05	-0.80

Appendix F

Convergence figures for the condition with Mix2PLM, person centering, 40-items, 600 examinees, 2 latent groups, and 11th replication

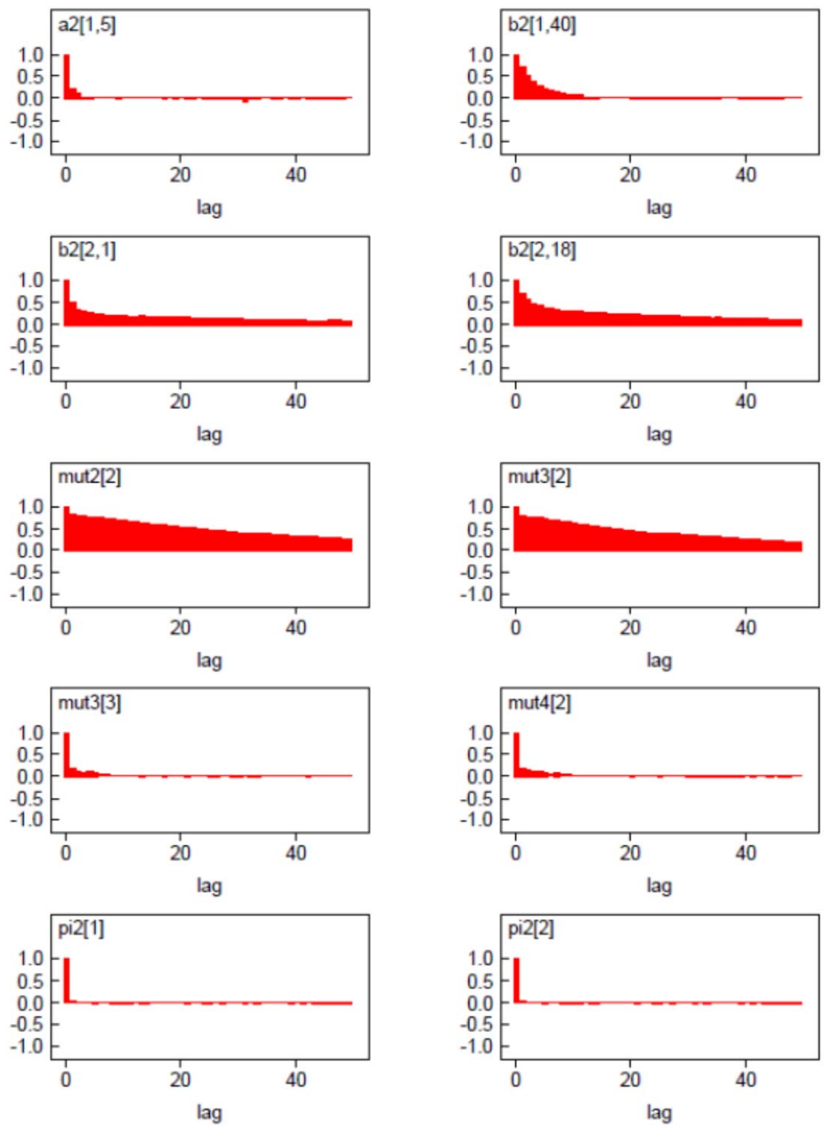


Figure F.1: The Autocorrelation Graphs for Mix2PLM

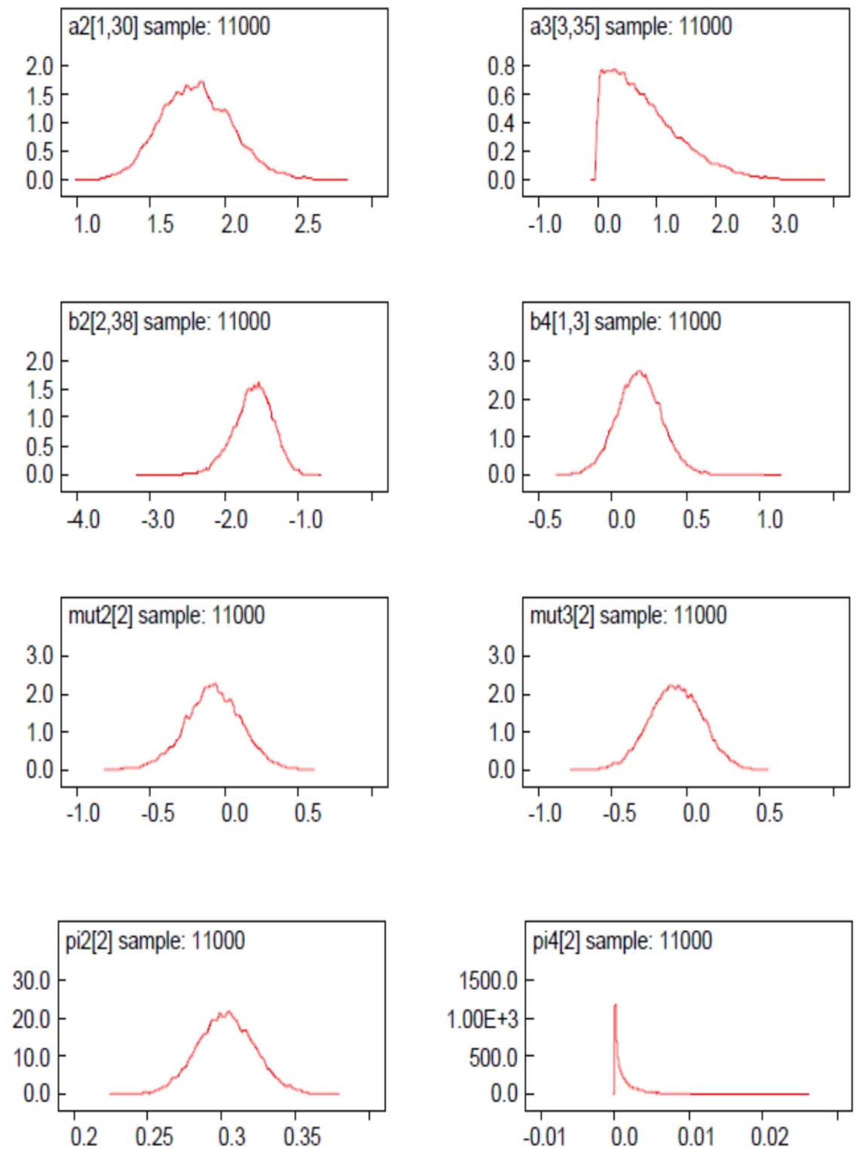


Figure F.2: The Density Plots for Mix2PLM

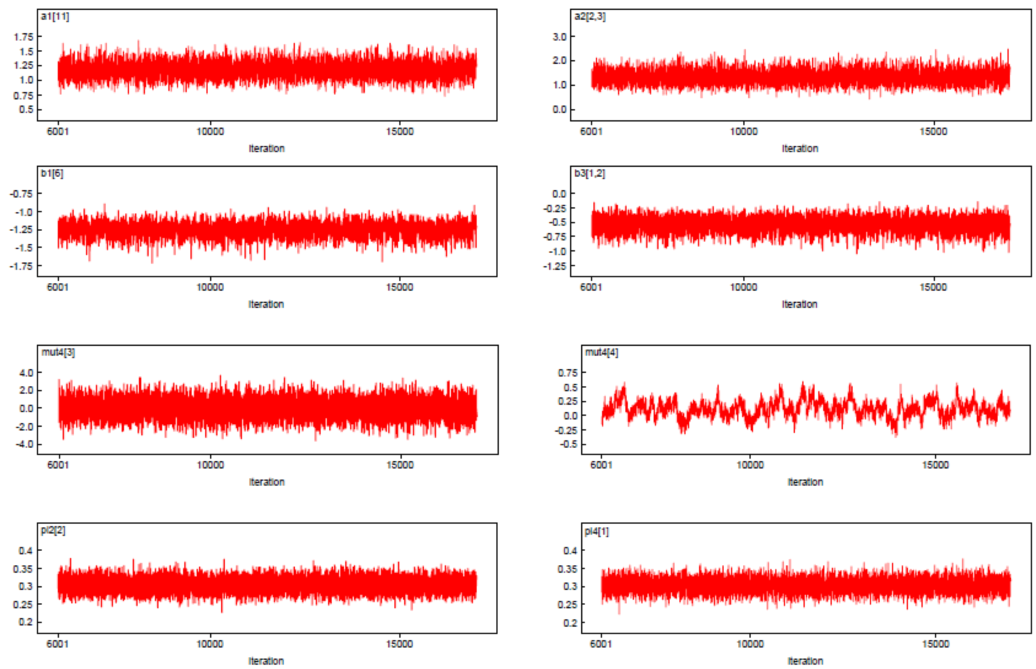


Figure F.3: The History Plots for Mix2PLM

Appendix G

Convergence figures for the condition with Mix3PLM, person centering, 20-items, 600 examinees, 3 latent groups, and 1st replication

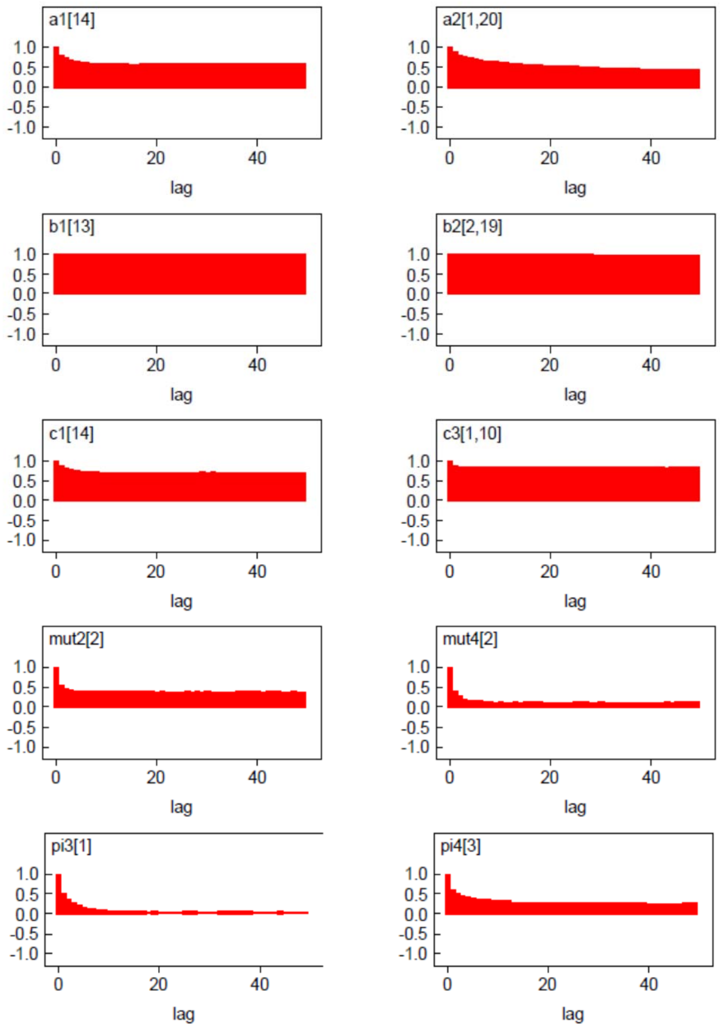


Figure G.1: The Autocorrelation Graphs for Mix3PLM

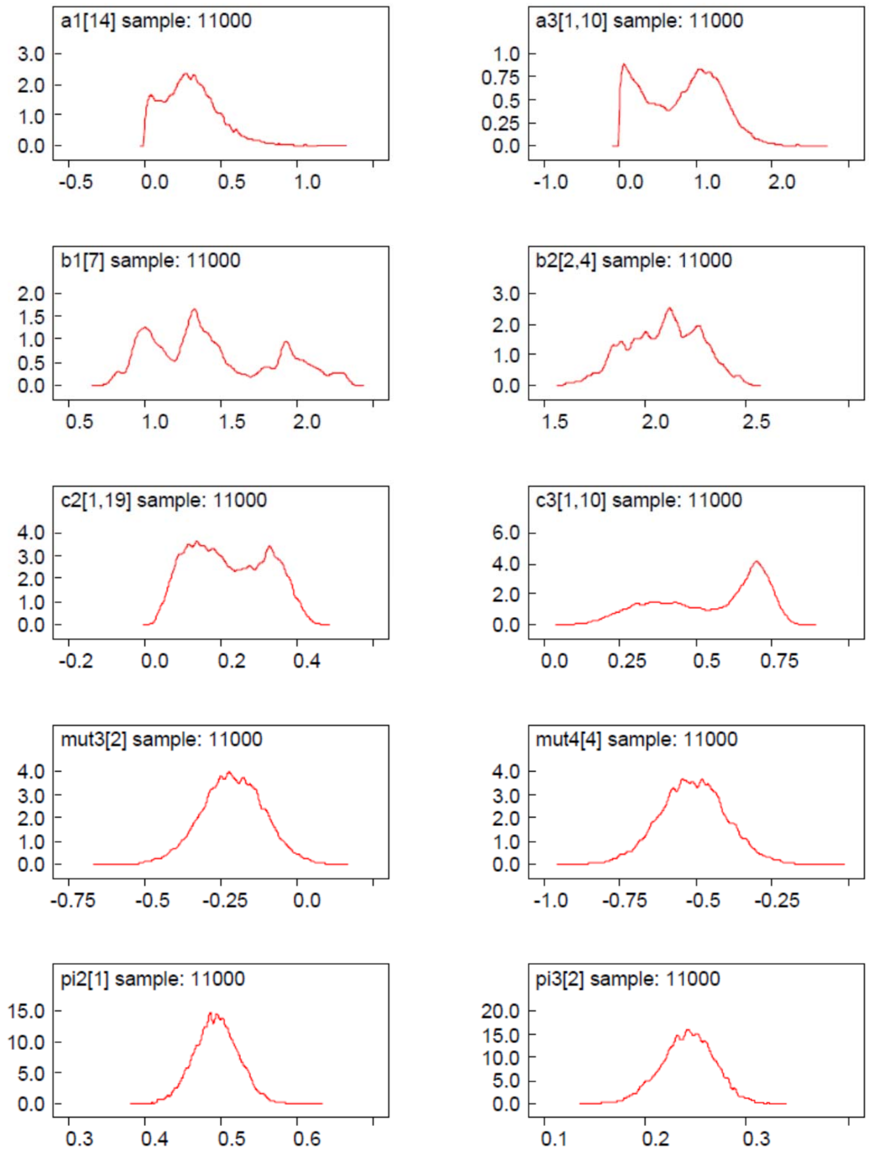


Figure G.2: The Density Plots for Mix3PLM

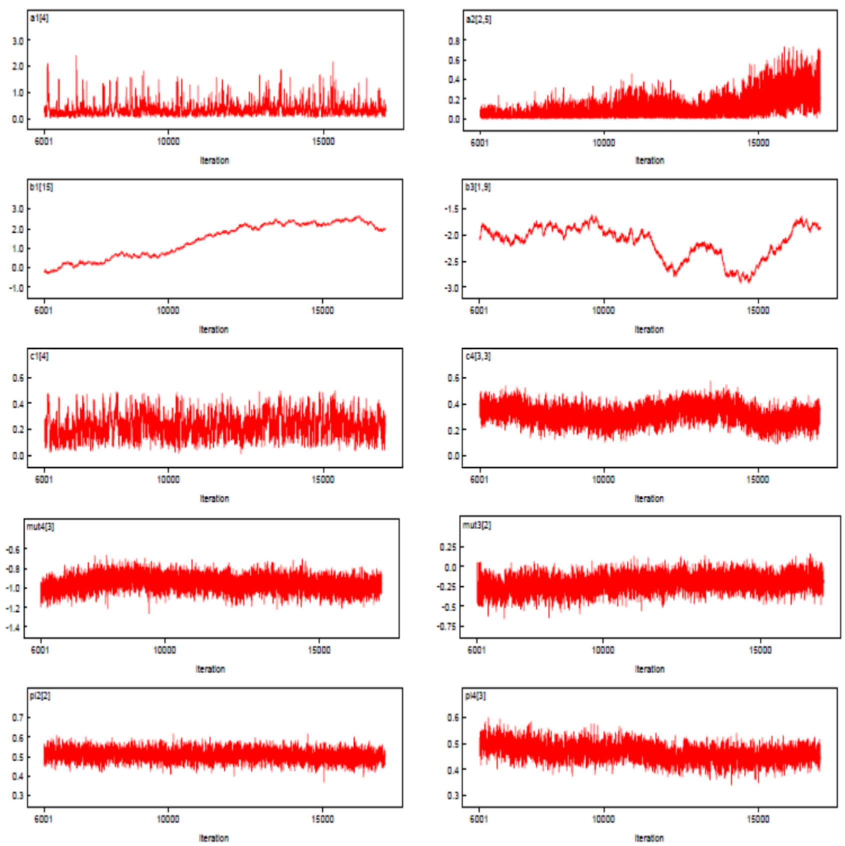


Figure G.3: The History Plots for Mix3PLM