COMPARISON OF DATA SAMPLING METHODS ON IRT PARAMETER ESTIMATION

by

TUGBA KARADAVUT

(Under the Direction of Ping Ma)

ABSTRACT

Data sampling methods are promising for analysis of large-scale data sets to reduce computing time and resources. These methods include uniform (random), and leverage-based sampling methods with a recent one called shrinkage leverage-based method. In this study, we compared data sampling methods for accuracy of item parameter estimates in IRT models. In addition, we introduced a new method of sampling, adjusted shrinkage leverage-based (Adj-SLEV) method. We analyzed two samples from PISA 2012 mathematics data set that were normally and non-normally distributed. Random sampling provided the most accurate Rasch item parameter estimates. The method with the highest accuracy varied depending on the type of item parameter for 2-pl and 3-pl models, if each parameter was evaluated individually. Adj-SLEV did not necessarily provide the highest accuracy for each type of item parameter individually, however, consistently provided a good trade-off when all parameters in a model were evaluated together.

INDEX WORDS: Item response theory, data sampling, PISA 2012 mathematics literacy test

COMPARISON OF DATA SAMPLING METHODS ON IRT PARAMETER ESTIMATION

by

TUGBA KARADAVUT

B.A., Balikesir University, Turkey, 2007

M.Ed., The University of Georgia, 2011

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment

of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2016

COMPARISON OF DATA SAMPLING METHODS ON IRT PARAMETER ESTIMATION

by

TUGBA KARADAVUT

Major Professor: Ping Ma
Committee: Jaxk Reeves
Allan S. Cohen

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2016

DEDICATION

Canim anneme ve canim babama…

ACKNOWLEDGEMENTS

I would like to give a special thank you to my advisors Drs. Ping Ma and Allan S Cohen for their endless help and support during this study. I would also like to thank Dr. Jaxk Reeves for his helpful comments and suggestions.

I owe a special debt of gratitude to my academic advisor Dr. Allan S. Cohen for guiding and supporting me towards a degree in Statistics from the very first day. I could not have ventured this journey without his guidance and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Background

Sampling is an important component of any research since it affects the validity of the results if the sample is not representative of the population. Data sampling, on the other hand, is a method used for analysis of large-scale data sets. The analyses of large data sets require longer time, larger data storage and CPU resources and sometimes different techniques. Recent work on sampling from a data set has suggested some techniques that may help overcome the estimation errors due to data sampling. Random sampling (also known as uniform sampling) and non-random (or leveraged-based sampling) are two general forms of sampling of the data set. Random (uniform) sampling of the data has been used largely, due to its simplicity. However, Cohen et al. (2015) has noted that random or uniform sampling is simpler but provides a weaker form of approximation of the data matrix. Even so, this method is still sufficient to approximate a large fraction of the original matrix (Cohen et al., 2015). Cohen et al. (2015) suggested an alternative method that randomly samples each row of the original data matrix with a probability that is proportional to its statistical leverage score. Although they have shown this method to be useful, leverage scores are not always easy to compute. In addition, Ma, Mahoney, and Yu (2015) suggested a new method called shrinkage leverage-based (SLEV) sampling. This method uses a combination of score probabilities from uniform and leverage-based sampling methods and has been found to provide improved conditional bias and variance estimates compared to uniform and leverage-based methods.

Item Response Theory (IRT), also known as latent trait theory, is a modern mental test paradigm which is extensively used for providing a theoretical basis for psychological measurement (Embretson, 1996), and for educational measurement (Lord & Novick, 1968). The invariance assumption is a fundamental property of IRT which makes it distinct from the classical test theory. This inherent property of IRT entails that the parameters that define the item properties be independent of the examinee sample, and the parameters that define the examinee propoerties (e.g., ability) be independent of the item sample (Hambleton, Swaminathan & Rogers, 1991). Although IRT parameter estimates are assumed to be invariant for any sample from the population, Stocking (1990) has shown that optimal samples for estimation of item parameters differ depending on the parameter being estimated. A central assumption of IRT is that examinees are randomly sampled from a population (Holland, 1990). Thus, when the distribution of ability is non-normal, for instance, errors in estimation increase in IRT models (Sass, Schmitt & Walker, 2008). This is a particular problem for statewide testing programs in that most ability distributions tend to be non-normal (Ho & Yu, 2015).

The recent work on techniques of sampling from a data set is promising for IRT estimation, because the studies provide methods other than random sampling for overcoming the estimation errors due to data sampling (e.g., Cohen et al., 2015; Ma et al., 2015). These methods may decrease the estimation errors for IRT models in the situations where data sampling is necessary such as large-scale data analyses. This study will compare the data sampling methods on IRT parameter estimation. In the following section, the objective of this study will further be introduced. In Chapters 2 and 3, a detailed background for IRT models and data sampling methods respectively will be provided. In Chapters 4 and 5, an empirical study and its results will be exhibited. Finally, in Chapter 6, the results from this study will be discussed.

1.2 Objective

Although data sampling methods have been used for estimation of regression based models (e.g., Cohen et al., 2015; Ma et al., 2015), as yet they have not been studied in estimation of Item Response Theory (IRT) models. In this study, the effects of different data sampling methods on IRT parameter estimation will be investigated. We will compare the uniform and leverage-based sampling methods, and the shrinkage leverage-based (SLEV) method for estimation of the IRT models. In addition to the SLEV method produced by Ma et al. (2015), we will introduce a new method of sampling, adjusted shrinkage leverage-based (Adj-SLEV), which provides an adjustment to the SLEV method. Two empirical examples of normally and non-normally distributed datasets from PISA 2012 will be presented for comparison of item parameter estimates from random, leverage-based, SLEV and Adj-SLEV sampling methods.

CHAPTER 2

ITEM RESPONSE THEORY

2.1 Brief Introduction

Item Response Theory (IRT) models have been extensively used in psychological measurement (Embretson, 1996) and in educational measurement (Lord & Novick, 1968). The IRT models have also been adopted for research and measurement in other fields including public health, ecology and sociology. IRT models define the relationship between an appropriate number of underlying latent traits (Embretson & Reise, 2000) and item responses through a continuous and monotonic function (Reckase, 2009).

IRT models employ parameters to describe person and item characteristics, and they vary depending on these parameters. The person parameters account for the differences between examinees regarding the underlying dimensions being measured, and the item parameters account for the differences between the items depending on the item types. The IRT models that assume only one underlying dimension and a logistic link are called unidimensional logistic IRT models. Members of these models include one-parameter logistic (1-pl), two-parameter logistic (2-pl), and three-parameter logistic (3-pl; Birnbaum,1968) models which are named depending on the number of item parameters in the models (e.g., Hambleton et al., 1991; Lord, 1980; Lord & Novick, 1968). The 1-pl, 2-pl, and 3-pl models were specifically developed for dichotomously scored item types such as multiple choice items. These are the most commonly used unidimensional IRT models and they will be the focus of this study.

One of the basic assumptions of IRT is the invariance property of the items and persons. The invariance property implies that the item parameters are independent of the examinee sample, and the person parameters are independent of the item sample (Hambleton, Swaminathan & Rogers, 1991). Although IRT parameter estimates are assumed to be invariant for any sample from the population, Stocking (1990) has shown that optimal samples for estimation of item parameters differ depending on the parameter being estimated.

2.2. Assumptions of Item Response Theory

Embretson and Reise (2000) have posited two basic assumptions concerning IRT. Firstly, an item characteristic curve (ICC) fits to data. Secondly, there exists an underlying latent trait (e.g., ability) which causes dependencies in examinee responses. These dependencies in the data can fully be accounted for by the model which is mathematically depicted with the fitted ICC. Ability and item difficulty are assumed to be on the same scale and in the same units. Although they can take on values changing from negative infinity to positive infinity, the range is often restricted to -3 to 3 for convenience (Baker, 2001). Ability is conventionally assumed to have a standard normal distribution (de Ayala, 2009). Item discrimination is also assumed to have a scale with a range from negative infinity to positive infinity, theoretically. However, its practical range is from 0 to 2.5 (Baker & Kim, 2004).

ICC is a monotonically increasing function of ability which presents the probability of a correct response to an item (see Figures 1-3). The function includes both person and item parameters as the variables. The person parameter is often called the ability parameter and denoted with theta (θ). The item parameters may consist of item difficulty (b), item discrimination (a) and pseudo-guessing parameters (c) depending on the IRT model (see Section 2.2). The difficulty parameter and the discrimination parameter are also referred to as location and slope parameters, respectively (Baker & Kim, 2004). The item difficulty is determined as a point on the ability score scale that corresponds to median of the ICC, and item discrimination is the slope of the ICC at this point. The pseudo-guessing parameter indicates a nonzero value of lower asymptote for ICC, which reflects the correct response to an item by chance (de Ayala, 2009).

2.3 Unidimensional Item Response Theory Models

The most commonly used unidimensional IRT models are the ones for dichotomous items that use a logistic mathematical link for defining the relationship between the latent variable (e.g., ability) and the item responses. Dichotomous items have binary response categories that correspond to either a correct response or an incorrect response. Multiple choice items are a commonly used example of the dichotomous item type. The number of item parameters in IRT models is the main decisive factor for the names given to these models.

2.3.1 One-parameter Logistic (1-pl) Item Response Theory Model or Rasch Model

The 1-pl IRT model includes only one item parameter that specifies the difficulty of an item. The model assumes item discrimination to be equal for all items. The 1-pl model that specifically fixes the discrimination parameter to one (Birnbaum, 1968, p. 402) is called the Rasch model (Rasch, 1960). The Rasch model defines the probability that an examinee *j* with ability θ answers the item i correctly ($P_i(\theta_j)$) by the following equation:

$$P_i(\theta_j) = \frac{1}{1 + e^{-(\theta_j - b_i)}}, \tag{1}$$

where b_i is the item difficulty parameter for item i. Figure 1 shows the ICCs for three different Rasch models that have different item location parameters and a fixed slope parameter of one. The ICCs are parallel to each other since they have an equal slope. The medians of the ICCs (e.g., $P_i(\theta) = 0.5$) correspond to the points -1, 0 and 1 on the ability scale, which are the measures of the item difficulties. The lower asymptotes of the ICCs are zero since the pseudo-guessing parameters do not exist, or equivalently pseudo-guessing parameters are zero in the model. The R (R Core Team, 2014) codes for creating Figures 1-3 are provided in Appendix A.

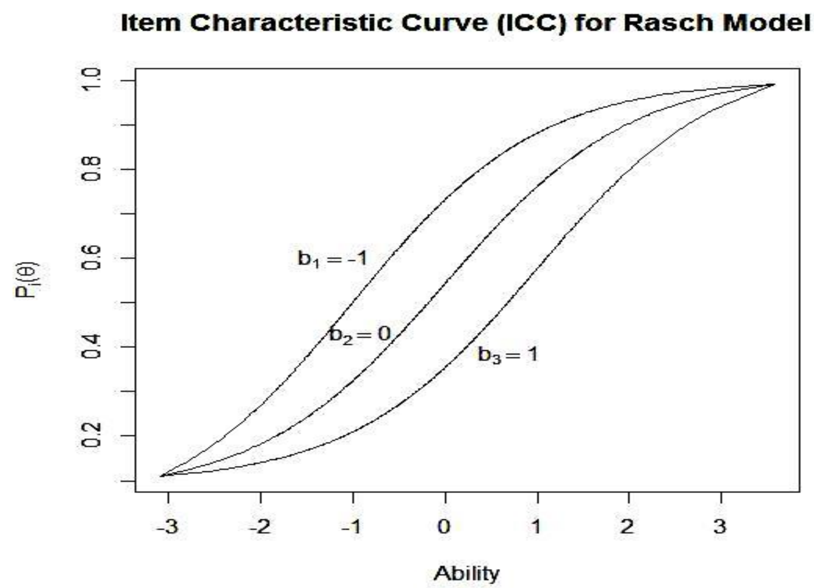

Figure 1: ICC’s for Rasch models

2.3.2 Two-parameter Logistic (2-pl) Item Response Theory Model

The 2-pl IRT model includes two item parameters which are item difficulty and item discrimination. Item difficulty and item discrimination are allowed to vary from item to item for predicting probability of correct response to an item given the ability of an examinee. The 2-pl logistic IRT model defines the probability that an examinee j with ability θ answers item i correctly ($P_i(\theta_j)$) by the following equation:

$$P_i(\theta_j) = \frac{1}{1 + e^{-a_i(\theta_j - b_i)}}, \qquad (2)$$

where b_i is the item difficulty parameter for item, and a_i is the item discrimination parameter for item i. Figure 2 shows the ICCs for three different 2-pl IRT models. Although a 2-pl model allows the item difficulty to vary from item to item, the item difficulties were fixed at zero in Figure 2 in order to demonstrate the effect of different item discrimination parameters on the ICCs. The ICCs are not parallel to each other since they have different slopes which are 0.8, 1.5 and 3. The medians of the ICCs (e.g., $P_i(\theta) = 0.5$) correspond to zero on the ability scale as the item difficulties were fixed to be zero for each of the ICCs in the figure. The pseudo-guessing parameters and, accordingly, the lower asymptotes of the ICCs are zero since the model does not incorporate a pseudo-guessing parameter.

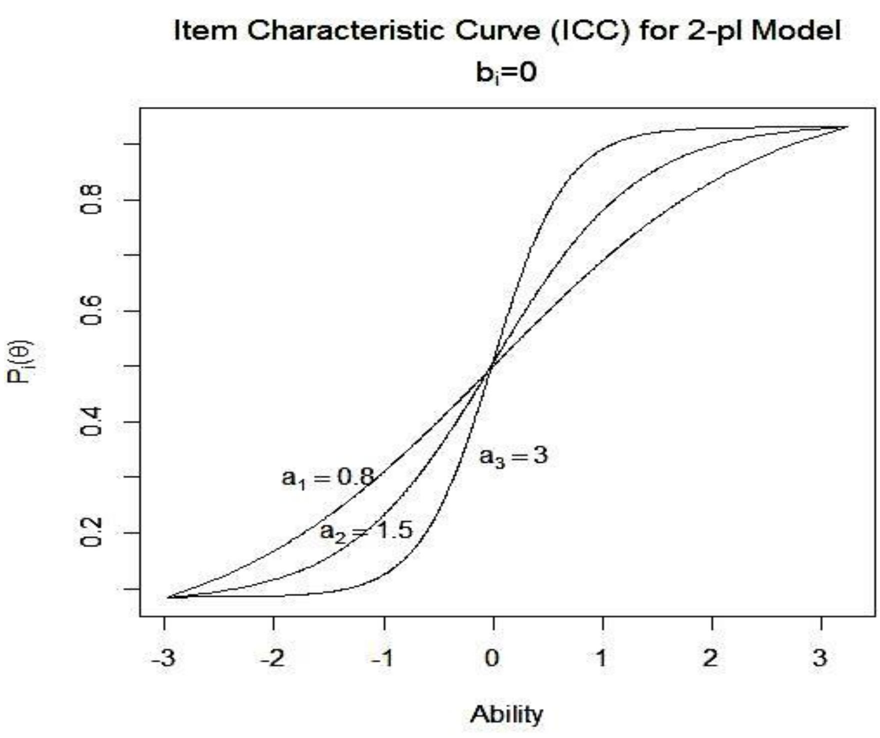

Figure 2: ICC's for 2-pl models

2.3.3 Three-parameter Logistic (3-pl) Item Response Theory Models

The 3-pl IRT model incorporates three different item parameters: item difficulty, item discrimination, and pseudo-guessing parameters. The 3-pl IRT model defines the probability that an examinee j with ability θ answers item i correctly ($P_i(\theta_j)$) by the following equation:

$$P_i(\theta_j) = c_i + (1 - c_i)\frac{1}{1 + e^{-a_i(\theta_j - b_i)}}, \qquad (3)$$

where b_i is the item difficulty parameter for item i, a_i is the item discrimination parameter for item i, and c_i is the pseudo-guessing parameter for item i. Figure 3 depicts the ICCs for three different 3-pl logistic IRT models. Although a 3-pl model allows the item difficulty and item discrimination to vary from item to item, they were fixed to be zero and one, respectively, in order to compare the effect of different pseudo-guessing parameters on the ICCs. The medians of the ICCs correspond to zero on the ability scale for each of the ICCs since the item difficulties were fixed at zero for each item. The pseudo-guessing parameters were determined to be 0.0, 0.1 and 0.2. The lower asymptotes of the ICCs in figure 3 are nonzero and they vary according to the determined pseudo-guessing parameters.

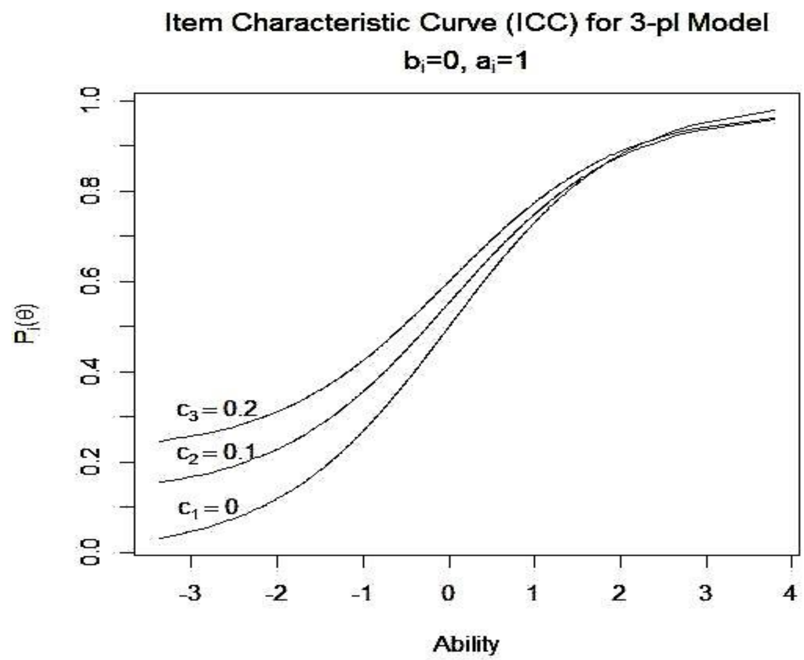

Figure 3: ICC's for 3-pl models

2.4 Scale Identification and Linking of the Scales

The scale of ability is arbitrary in the origin and in the unit. The arbitrariness of the ability scale is denoted as scale indeterminacy or the metric identification problem (de Ayala, 2009, p.41; Baker & Kim, 2004, p. 90). IRT locates item and ability parameters on the same scale. Therefore, fixing either the ability or item parameter scale solves the metric identification problem (de Ayala, 2009).Three different methods have been proposed for identifying the metric of ability in IRT models. The first method is equating via item anchoring which is particularly employed in existence of multiple examinee samples assuming that the estimates of particular item parameters are fixed across these groups (e.g., Angoff, 1971; Kolen & Brennan, 2004). The other methods include person centering and item centering (de Ayala, 2009). We employed item centering during the calibrations which implies fitting the mean of the item difficulty estimates to zero during the estimations.

The item parameter estimates obtained from different samples of examinees are not comparable since the scales of the items are different (Stocking & Lord, 1983). The estimated parameters are required to be placed on the same scale before they are compared for bias and root-mean-square error (RMSE). The indeterminacy of the scale implies that the scale of ability is unique only after a linear transformation (Lord & Novick, 1968, p. 366) since the indeterminacy is only in the origin and unit of the ability scale (Stocking & Lord, 1983). The linear transformation is achievable by the invariance property of IRT modeling (Lord, 1980). In this study, we employed the mean/sigma equating method for linear transformation of the scales when comparing estimates from sampled data sets to the estimates from the original dataset (Marco, 1977).

2.5 Mean-Sigma Equating

The invariance property of IRT implies that the probability that an examinee answers an item correctly should be independent of the sample of items being calibrated (Hambleton et al., 1991). In other words, assuming θ and θ^* are ability estimates for the same examinee from two different calibrations, the probability for this examinee to answer an item i is expected to be equal over the calibrations:

$$P_i(\theta) = P_i(\theta^*), \tag{4}$$

Modeling these probabilities using a 2-pl IRT model gives the following equations:

$$\frac{1}{1 + e^{-a_i\,(\theta - b_i)}} = \frac{1}{1 + e^{-a_i^*(\theta^* - b_i^*)}}, \tag{5}$$

$$e^{a_i\,(\theta - b_i)} = e^{a_i^*(\theta^* - b_i^*)}, \tag{6}$$

$$a_i\,(\theta - b_i) = a_i^*(\theta^* - b_i^*). \tag{7}$$

Multiplying θ and b_i by a constant (e.g., A) and dividing a_i by the same constant would leave $a_i(\theta - b_i)$ unchanged. That is,

$$\frac{a_i}{A}(\theta - b_i)A = a_i^*(\theta^* - b_i^*). \tag{8}$$

This equation implies that:

$$\frac{a_i}{A} = a_i^*, \tag{9}$$

and

$$b_i A + B = b_i^*, \tag{10}$$

$$\theta A + B = \theta^*. \tag{11}$$

As a result, replacing the parameters b_i with b_i^*, θ with $\theta^*$and a_i with a_i^* does not change the initial probabilities of a correct response which was shown with Equation 4 (Hambleton et al., 1991). The constants A and B are called metric transformation coefficients. The transformations place the scale of parameters from a calibration onto the scale of parameters from a target calibration (de Ayala, 2009). The c_i parameter is not affected by the scale indeterminacy. Therefore, it is invariant across different calibrations without a need for a transformation (Lord, 1980). There are several methods for obtaining the metric transformation coefficients. In this study, we will employ Marco's (1977) mean-sigma method. The mean/sigma method uses the means and standard deviations of the parameter estimates from two calibrations to determine the coefficients. Assuming the first calibration to be the target calibration, the following constants can be calculated:

$$A = \frac{\sigma(b_{(Calib1)})}{\sigma(b_{(Calib2)})}, \tag{12}$$

$$B = \mu(b_{(Calib1)}) - A\mu(b_{(Calib2)}), \tag{13}$$

where $\sigma(b_{(Calib1)})$ is the standard deviation of the estimated b parameters from calibration one, $\sigma(b_{(Calib2)})$ is the standard deviation of the estimated b parameters from calibration two, $\mu(b_{(Calib1)})$ is the mean of the estimated b parameters from calibration one, and $\mu(b_{(Calib2)})$ is the mean of the estimated b parameters from calibration two. The scale of the estimated parameters from calibration two can be changed into the scale of the estimated parameters from scale one by employing the following transformations:

$$b_{(new)} = A(b_{(Calib2)}) + B, \tag{14}$$

$$\theta_{(new)} = A(\theta_{(Calib2)}) + B, \tag{15}$$

$$a_{(new)} = \frac{a_{(Calib2)}}{A}, \tag{16}$$

$$c_{(new)} = c_{(Calib2).} \tag{17}$$

CHAPTER 3

DATA SAMPLING METHODS

3.1 Randomization versus Statistical Adjustments

Statistical estimations and experimental designs retain some amount of uncontrolled variation (Cox, 1958). For instance, randomization is a technique that can be used to ensure that the expected error is zero when the error variable cannot be controlled. The effects of uncontrolled variation may be reduced by using the available knowledge regarding the nature of variation. Supplementary information provided by concomitant variables (a.k.a. auxiliary variables) can be used to increase precision of estimations by means of explaining some of the uncontrolled variation (Cox, 1958).

IRT models depict the relationship between the latent variable (e.g., ability) and item responses through mathematical models. This relationship is statistically adjusted by item characteristics such as item discrimination, item difficulty and guessing parameters (Van der Linden & Hambleton, 1997). Statistical adjustments can also be applied for sampling the datasets. Leverage-based sampling methods provide statistical adjustments to data sampling as opposed to the conventional method of random sampling. Leverage scores are calculated by using the concomitant variables as predictors in a linear regression model (Ma et al., 2015).

3.2 Data Sampling Methods

In this section, data sampling methods will be introduced. The original data will be sampled by preserving the data rows because each row corresponds to observations of one individual. The data rows to be sampled in this study are the dichotomously scored examinee responses to mathematics items.

3.2.1 Uniform (Random) Sampling Method

Uniform sampling draws the data rows uniformly at random, which means each row of the original data has the same probability of being sampled. That is,

$$\pi_i^{Uni} = 1/n \tag{18}$$

for each $i \epsilon n$ where n is the number of rows in the original data matrix (equivalently, the size of the original sample), and π_i is the probability that data raw i will be sampled (Ma et al., 2015).

3.2.2 Leverage-based Sampling Method

Leverage scores are commonly measured as hat-values (h_{ii}) which are the elements in the diagonal of hat matrix. The hat matrix is calculated by the following equation:

$$H = X(X'X)^{-1}X', \tag{19}$$

where the hat matrix is denoted by H, and X' is the transpose of the design matrix in matrix formation of linear regression (Hoaglin & Welsch, 1978). In simple regression, the observations that are far from the mean of predictor variable have high leverage scores. Observations with high leverage scores have substantial impact on the fitted values. The hat-value measure for simple regression can be restated as:

$$h_{ii} = \frac{1}{n} + \frac{(x_{i\_}\bar{x})^2}{\sum_{j=1}^{n}(x_{j\_}\bar{x})^2}. \tag{20}$$

with x_i as the values of the predictor variable, and $\bar{x}$ as the mean of the predictor variable (Fox, 1991). In this study, we used a single predictor for calculating the leverage scores to be used for sampling.

Leverage-based sampling method draws the data according to "an importance sampling distribution that is proportional to the normalized leverage scores". The probability that data row i will be sampled is calculated as:

$$\pi_i^{Lev} = \frac{h_{ii}}{\sum_{i=1}^{n}(h_{ii})}, \tag{21}$$

where h_{ii} is the leverage score for data row i (Ma et al., 2015).

In this study, there was a necessity to determine a dependent and a predictor variable in order to calculate the leverage scores. Traditional mathematics total scores were calculated by summing up the item scores to be used as the independent variable. We preferred the total raw scores to the scores after IRT calibration, because the purpose of this study is to achieve the IRT parameter estimates of the original dataset from the analysis of the subsamples without analyzing the original dataset. We will refer to the predictor variable as a *covariate* following the literature on IRT (e.g., Dai, 2013; Tay, Vermunt & Wang, 2013). The covariate can be selected from among the concomitant or auxiliary variables if such are available. In this study, we generated a covariate that has a high correlation with the dependent variable (e.g., r =.90). A covariate having a higher correlation with the dependent variable is assumed to produce more accurate leverage scores, because it explains a higher variance in the dependent variable. The selection of the dependent variable does not have a direct effect on the leverage scores, because the leverage scores are calculated based on the X matrix (Hoaglin & Welsch, 1978). However, we generated the covariate to have a high correlation with the dependent variable. Therefore, the dependent variable had an impact on the leverage scores by this means.

3.2.3 Shrinkage Leverage-based (SLEV) Sampling Method

Shrinkage leverage-based (SLEV) sampling method is a new data sampling method introduced by Ma et al. (2015). The method combines the benefits of uniform and leverage-based sampling methods by employing a convex combination of the probability distributions from two methods. The probability that the data row i will be sampled is determined by:

$$\pi_i^{Shr} = \alpha\pi_i^{Lev} + (1-\alpha)\pi_i^{Unif}, \tag{22}$$

where $\alpha \epsilon (0,1)$, π_i^{Lev} is the probability that data row i will be sampled based on the leverage-based sampling method, and π_i^{Unif} is the probability that data row i will be sampled based on the uniform (random) sampling method (Ma et al., 2015). Based on the simulation studies, Ma et al. recommended using $\alpha = 0.9$ as a rule of thumb in order to account for the subsample size and variance inflation trade-off in parameter estimation.

3.2.4 Adjusted Shrinkage Leverage-based (Adj-SLEV) Sampling Method

In this study, we propose an adjustment to the shrinkage leverage-based (SLEV) method. We propose setting $\alpha = 1$ when the leverage score is higher than the uniform probability. Similarly, we set $\alpha = 0$ when the uniform probability is higher than the leverage score (see Equation 23). Adj-SLEV sampling method ensures that a data row *i* has at least an equal probability of being selected as it would have if the sampling distribution of the population was uniform.

$$\pi_i{}^{AdjShr} = \alpha\pi_i{}^{Lev} + (1-\alpha)\pi_i{}^{Unif},$$
$$if\ \pi_i{}^{Lev} > \pi_i{}^{Unif}\ then\ \alpha = 1,$$
$$if\ \pi_i{}^{Lev} < \pi_i{}^{Unif}\ then\ \alpha = 0. \tag{23}$$

CHAPTER 4

EMPRICAL STUDY

4.1 Normal and Non-normal Ability Distributions

We considered two empirical data sets where one of them had an approximately normal distribution of raw scores (e.g., total score) and the other had a non-normal distribution of raw scores. The non-normality of the raw score distribution may also indicate non-normality in distribution of the latent trait (e.g., ability). We considered both normal and non-normal distribution of ability, because errors in estimation increase in IRT models when the distribution of ability is non-normal (Sass, Schmitt & Walker, 2008).

The data sets are two different samples from the 2012 cycle of the Program for International Student Assessment (PISA), which belongs to the Organization for Economic Cooperation and Development (OECD). PISA assesses students in an international context for their readiness to become a member of the society when they are near their end of the compulsory education (OECD, 2013, p.13-18). It measures reading literacy, science literacy, and mathematics literacy by determining one of these domains as the major domain in each cycle. The mathematics literacy was the main domain in PISA 2012.

PISA 2012 administered 13 booklets for assessing literacy in mathematics, science, and reading. The booklets included clusters of items in a rotation design. That is, students were administered different sets of items. In order to have a sample of students who were administered the same set of items, we determined to analyze the mathematics data from Booklet 10 for both normally and non-normally distributed data sets. The mathematics clusters contained multiple choice and constructed response items (OECD, 2014).

There were 36 mathematics items in Booklet 10 and four of these items were partial credit items. We dropped these four items from the data set which resulted in 32 items. As a result, we used the same items to result in normal and non-normal datasets. However, we sampled students from different sets of countries to end with normal and non-normal distributions. More information about the data sets is provided in the following section. PISA provides four types of missing data. The invalid and missing data were recoded as an incorrect response, while N/A and unreached items were kept as missing values. We dropped the missing values from the data set listwise. This resulted in a sample size of 2,058 for an approximately normal data set and a sample size of 1,906 for the non-normal data set.

4.2 Data set with Normal Ability Distribution (Empirical Study 1)

4.2.1 Distribution of Raw Scores

The data was from four countries including United Kingdom ($n_1 = 959$), Germany ($n_2 = 343$), Belgium($n_3 = 495$), and Latvia ($n_4 = 261$) with a total sample size of ($N = 2{,}058$). The distribution of the mathematics raw scores was as shown in Figure 4. The range of the scores was 32 with a minimum score of 0 and a maximum score of 32. The mean, median and mode of the distribution were 16.44, 16.00 and 16.00, respectively.

The skewness and kurtosis of the raw score distribution were estimated and tested for significance by using R moments package (Komsta & Novomestky, 2015). The moments package (Komsta & Novomestky, 2015) provides the Anscombe-Glynn test of kurtosis (Anscombe & Glynn, 1983) and the D’Agostino test of skewness (D’Agostino, 1970) for normal samples. Both methods assume a null hypothesis of normality and an alternative hypothesis of deviation from normality. Under the null hypothesis of normality, the distribution of the data should have a kurtosis of three and a skewness of zero. The two-sided tests of skewness and kurtosis indicated an approximately normal distribution for the five countries with an insignificant estimate of -0.002 (p = .964) for skewness and a significant estimate of 2.23 (p < .001) for kurtosis. The Q-Q plot exhibited heavy tails for this distribution (see Figure 2.5). The normality of the latent ability distribution is estimated in the next section.

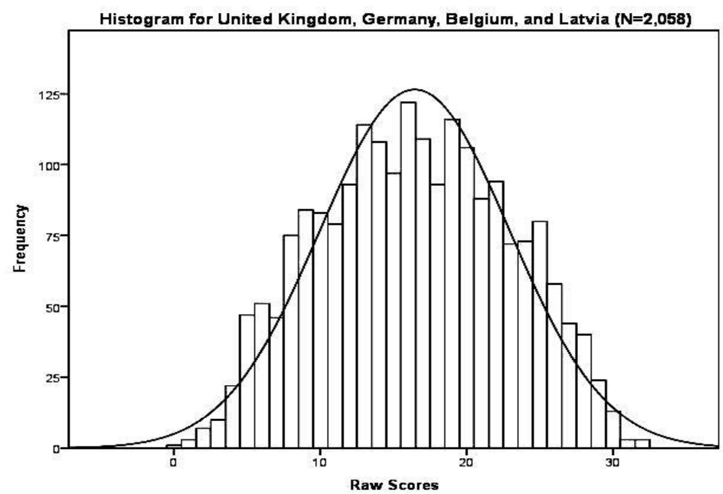

Figure 4: Distribution of total scores

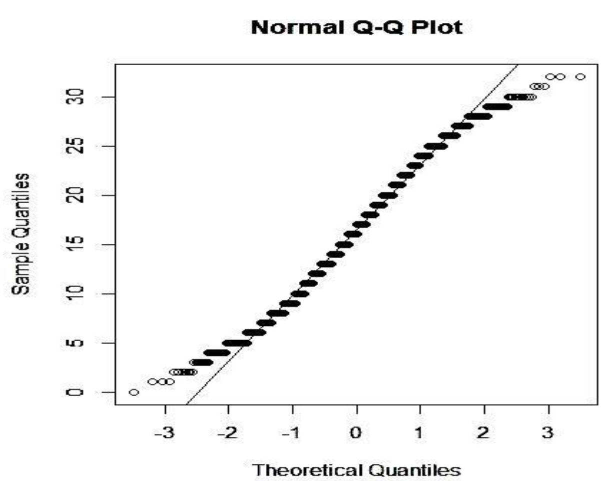

Figure 5: Q-Q plot for checking normality of total score distribution

4.2.2 Distribution of Latent Ability

The distribution of the latent ability was investigated for non-normality. We used the *sirt* package (Robitzsch, 2013) as implemented in R for estimation of the latent density. The *sirt* package allows for semiparametric marginal maximum likelihood (MML) estimation. That is, a non-parametric estimation of the latent density and the estimation of item parameters using MML estimation were conducted simultaneously (e.g., Lindsay, Clogg, & Grego, 1991; Wellner, 1986). Log-linear smoothing up to third and fourth moments were fitted to the data for estimating the latent density (e.g., Xu & von Davier, 2008). The first four moments of a distribution are mean, variance, skewness and kurtosis, respectively. Taking the third moment of the distribution into account for smoothing captures the non-normality in the distribution (Xu & von Davier, 2008).

The best fitting models were determined based on the AIC, BIC and CAIC information criteria for Rasch, 2-pl and 3-pl models (see Table 1). For each model, the best fitting model was either smoothed up to third or four moments which indicated non-normal distribution for the latent ability. A model with log-linear smoothing up to the first moment yielded a relatively best fit for the Rasch model, a model with log-linear smoothing up to the second moment yielded a relatively best fit for the 2PL, and a model with log-linear smoothing up to the third moment yielded a relatively best fit for the 3PL model. The estimated distribution of latent ability from different models is shown in Figure 6.

The *sirt* package provides the rasch.mmle2 function which estimates skewness for the latent ability distribution. However, it does not provide an estimate of kurtosis for the latent ability distribution. The R sirt package uses equation 24 for estimating skewness, however, it does not provide an estimate of kurtosis. We have written an R function for estimating the kurtosis by using the equation 25. Theta.k, pi.k, and mean.trait are reported by the rasch.mmle2 function and they describe the latent distribution of ability. Theta.k is a vector of the grid points over which the ability should be evaluated, pi.k is the distribution of ability on theta.k, and mean.trait is the estimated mean of ability (Robitzsch, 2015). The estimated skewness for the best fitting models were 0.00, -0.294, and -2.258 for Rasch, 2PL and 3PL models, respectively. The estimated kurtosis was 3.00, 2.745, and 8.254 for these models, respectively.

$$Skewness = \frac{\sum pi.k * (theta.k - mean.trait)^3}{\sum (pi.k * (theta.k - mean.trait)^2)^{(\frac{3}{2})}}, \tag{24}$$

$$Kurtosis = \frac{\sum pi.k * (theta.k - mean.trait)^4}{\sum (pi.k * (theta.k - mean.trait)^2)^{(\frac{4}{2})}}, \tag{25}$$

Table 1: Model Fit Information for Models with Log-linear Smoothing up to the Specified Moments

Moments		Mean			Variance		
Model Fit Indices		AIC	BIC	CAIC	AIC	BIC	CAIC
Models	Rasch	68651.86	**68837.64**	**68870.64**	68651.95	68837.72	68870.72
	2-pl	67941.79	68302.07	68366.07	**67918.70**	**68278.99**	**68342.99**
	3-pl	67911.97	68452.41	68548.41	67851.95	68392.38	68488.38

Table 1 Continued: Model Fit Information for Models with Log-linear Smoothing up to the Specified Moments

Moments		Skewness			Kurtosis		
Model Fit Indices		AIC	BIC	CAIC	AIC	BIC	CAIC
Models	Rasch	68646.93	68838.33	68872.33	**68643.74**	68840.78	68875.78
	2-pl	67920.40	68286.32	68351.32	67919.23	68290.77	68356.77
	3-pl	**67804.57**	**68350.63**	**68447.63**	67805.08	68356.77	68454.77

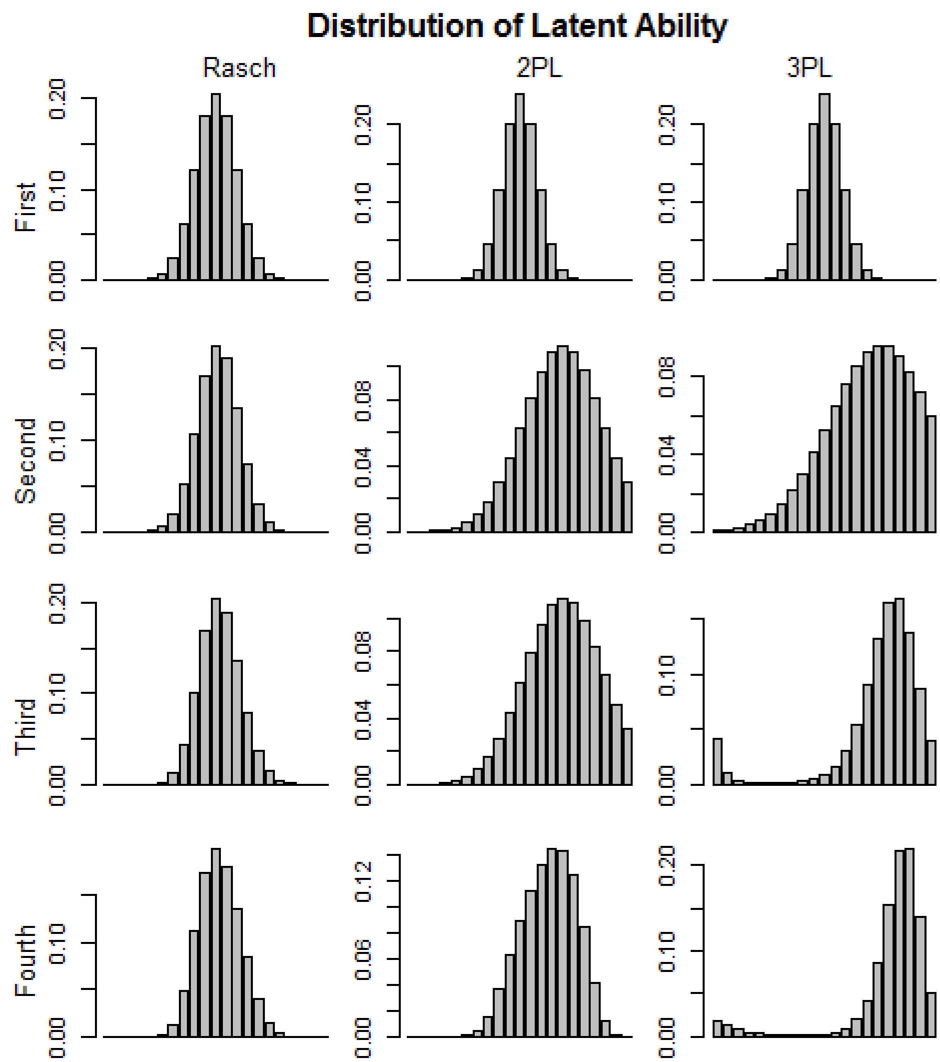

Figure 6: Estimated distribution of latent ability from different models

4.3 Data set with Non-normal Ability Distribution (Empirical Study 2)

4.3.1 Distribution of Raw Scores

The data from the five countries with the lowest mathematics average scores among 31 participating countries in PISA 2012 were selected for this dataset. The mathematics average for these counties varied from 368 to 386, while the average was 494 for all participating countries (OECD, 2014). The distribution of the mathematics raw scores exhibited a positively skewed distribution for this sample (see Figure 2.7). The dataset included 240 students from Peru, 356 students from Indonesia, 609 students from Qatar, 282 students from Colombia and 419 students from Jordan (N=1,906). Remembering that the number of items was 32, the range of the scores was observed to be 31 with a minimum score of 0 and a maximum score of 31. The mean, median and mode of the distribution were 9.13, 8.00 and 6.00, respectively. Among the students in the population, 62.2% scored below 9 and 69.9% scored below 10. Only 5.9 % of the students scored 20 and higher.

The skewness and kurtosis of the raw score distribution were estimated and tested for significance by using the R moments package (Komsta & Novomestky, 2015). The Anscombe-Glynn test of kurtosis (Anscombe & Glynn, 1983) and D'Agostino test of skewness (D'Agostino, 1970) for normal samples both indicated deviation from normality. The two-sided tests of skewness and kurtosis exhibited a significant non-normality for the distribution of the five countries with the estimates of 1.099 (p < .001) and 4.124 (p < .001), respectively. The non-normality of the observed scores may indicate a non-normal latent ability distribution as well. The Q-Q plot indicates a right skew in the distribution (see Figure 8). In the next section, we estimated the distribution of latent ability.

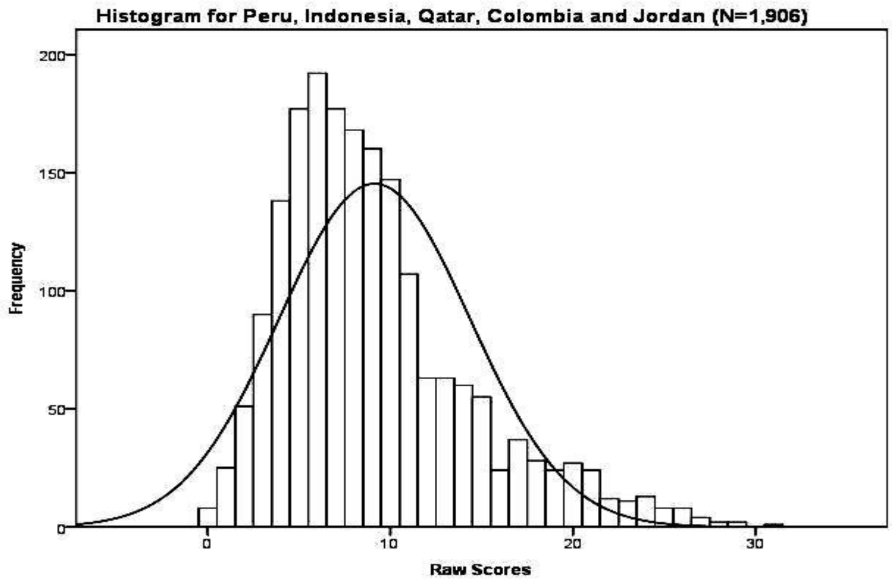

Figure 7: Distribution of total scores

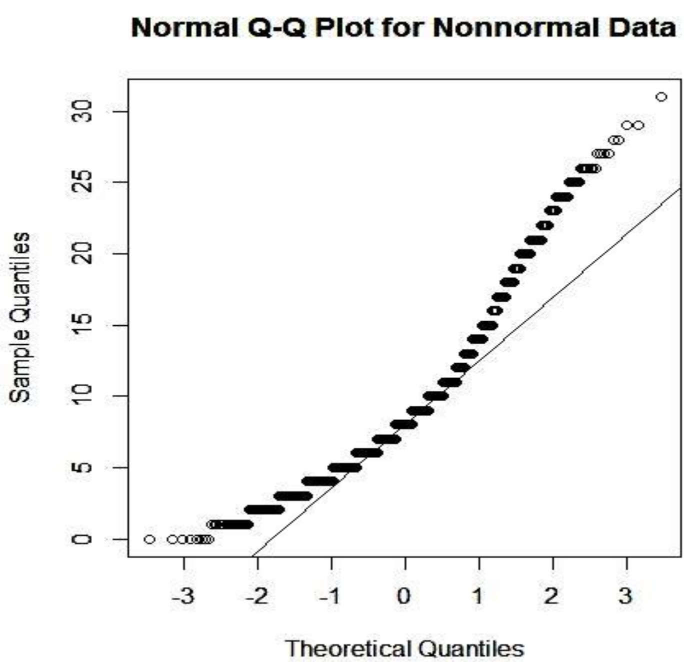

Figure 8: Q-Q plot for checking normality of total score distribution

4.3.2 Distribution of Latent Ability

The observed distribution of the scores exhibited non-normality for the five countries. However, the distribution of the latent ability yet needs to be investigated for the non-normality. We used the *sirt* package (Robitzsch, 2013) as implemented in R for estimation of the latent density. Log-linear smoothing up to third and fourth moments were fitted to the data for estimating the latent density (e.g., Xu & von Davier, 2008). The best fitting models were determined based on the AIC, BIC and CAIC information criteria for Rasch, 2-pl and 3-pl models (see Table 2). For each model, the best fitting model was either smoothed up to third or four moments which indicated non-normal distribution for the latent ability. A model with log-linear smoothing up to the fourth moments yielded a relatively best fit for the Rasch model, while a model with log-linear smoothing up to the third moments yielded a relatively best fits for 2-pl and 3-pl models. The estimated skewnesses for the best fitting models were 0.958, 0.289, and -2.099 for Rasch, 2-pl and 3-pl models, respectively. Similarly, the estimated kurtoses were 4.062, 2.922, and 6.029 for these models, respectively. The estimated distribution of latent ability from different models is shown in Figure 9.

Table 2: Model Fit Information for Models with Log-linear Smoothing up to the Specified Moments

Moments		Mean			Variance		
Model Fit Indices		AIC	BIC	CAIC	AIC	BIC	CAIC
Models	Rasch	54468.97	54652.21	54685.2	54471.64	54654.88	54687.88
	2-pl	53775.51	54130.89	54194.9	53777.13	54132.51	54196.51
	3-pl	53628.69	54161.75	54257.8	53616.56	54149.63	54245.63

Table 2 Continued: Model Fit Information for Models with Log-linear Smoothing up to the Specified Moments

Moments		Skewness			Kurtosis		
Model Fit Indices		AIC	BIC	CAIC	AIC	BIC	CAIC
	Rasch	54391.11	54579.91	54613.90	**54350.41**	**54544.75**	**54579.75**
Models	2-pl	**53767.32**	**54128.25**	**54193.30**	53769.12	54135.60	54201.60
	3-pl	53540.68	**54079.30**	**54176.30**	**53540.31**	54084.48	54182.48

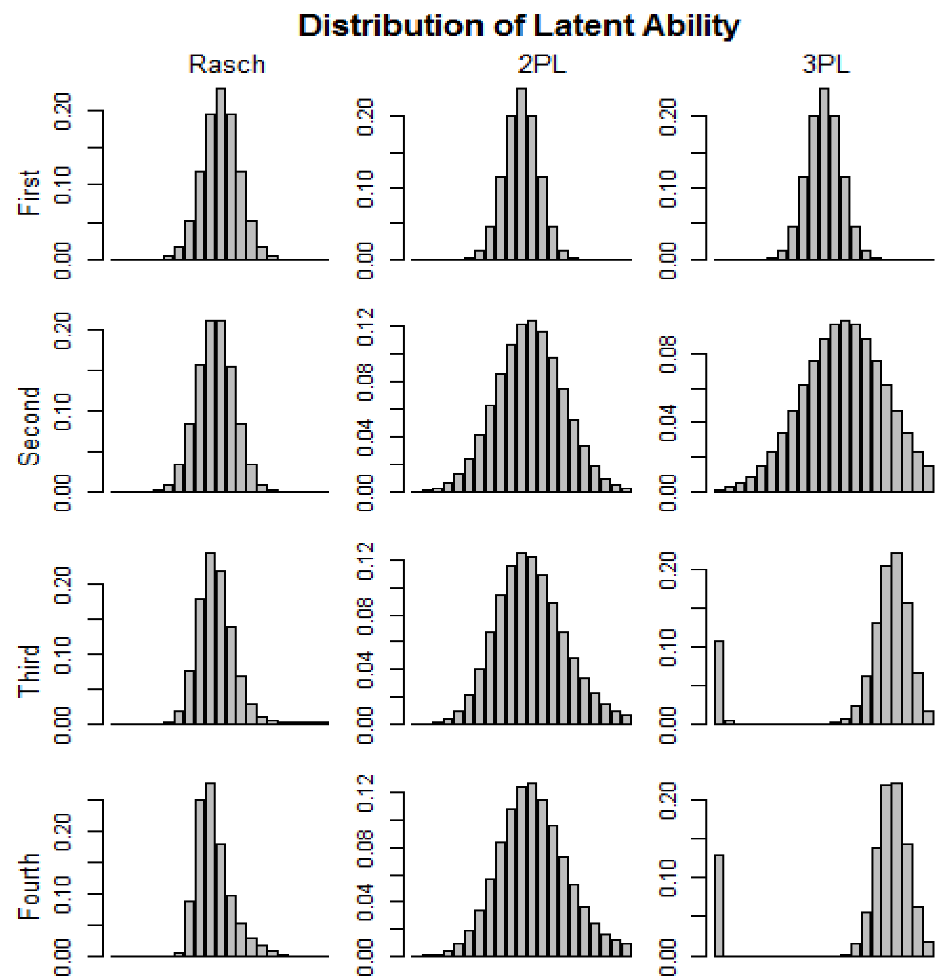

Figure 9: Estimated distribution of latent ability from different models

4.4 Sampling the Empirical Data Sets

Data sampling methods included in this study were random, leverage-based, SLEV, and Adj-SLEV sampling methods. These methods were described broadly in section 3.2. Sampling of data sets based on data sampling methods were done using the R (R Core Team, 2014) software. The R code used for data sampling is given in Appendix C.3. From each data set, 548 students were sampled, which corresponds to 27% of the students for dataset with an approximately normal raw score distribution (N=2,058) and 29% of the students for the dataset with a non-normal raw score distribution (N=1,906). Fifty data sets were sampled for each data sampling method. For leverage-based sampling methods, we generated a covariate that has a high correlation with the dependent variable (e.g., r =.90). This covariate was used to predict total mathematics score in a univariate linear regression model in order to calculate the leverage scores. The leverage scores were later normalized to create an importance sampling distribution for sampling data rows from the full datasets (Ma, 2015). R (R Core Team, 2014) code for generating the covariate and calculating leverage scores are given in Appendix C.3.

4.5 Parameter Estimation

4.5.1 Estimation of Parameters from Full Data Sets

Our interest in this study was Bayesian estimation of IRT models (i.e. marginalized Bayesian estimation; Mislevy, 1986). Estimation of item parameters can be done by using the Markov Chain Monte Carlo (MCMC) method as implemented in the computer software OpenBUGS (Lunn, Spiegelhalter, Thomas & Best, 2009). Bayesian estimation specifies a prior distribution for the parameters to be estimated (Baker & Kim, 2004). The convention is assuming a normal prior distribution for the latent ability. However, the ability distribution may not be normal necessarily (Hambleton & Swaminathan, 1985). In this study, we presented two empirical studies: one with approximately normal total score distribution and one with non-normal total score distribution. The non-normality of the total score distribution may also indicate non-normality of the ability distribution, although it does not guarantee the non-normality for ability. Similarly, the normality of the total scores does not guarantee normality of the latent score distribution. The semi-parametric analyses can be used to estimate the latent density. We used the R *sirt* package (Robitzsch, 2013) for estimation of the latent density. The *sirt* package allows for semiparametric marginal maximum likelihood estimation with log-linear smoothing. The log-linear smoothing up to third or fourth moments can be used to address the non-normality in the latent density (Xu & von Davier, 2008). The analyses of full datasets could be done using MCMC estimation if the ability distribution is normal. Alternatively, semi-parametric estimation with log-linear smoothing could be used when the ability distribution is either normal or non-normal.

The semi-parametric analyses of the approximately normal dataset using the R *sirt* package (Robitzsch, 2013) resulted with approximately normal distributions of ability for Rasch and 2-pl models, and a skewed distribution for the 3-pl model (see Table 1). Therefore, either MCMC estimation or semi-parametric estimation with log-linear smoothing can be used for item parameter estimation of Rasch and 2-pl models, and semi-parametric analyses can be used for item parameter estimation of the 3-pl model. On the other hand, semi-parametric analyses of a non-normal dataset with log-linear smoothing up to fourth moments yielded a relatively best fit for the Rasch model, while the analyses with log-linear smoothing up to the third moments yielded relatively best fits for 2-pl and 3-pl models (see Table 2). Therefore, item parameter estimation for the non-normal dataset can be done using semi-parametric analyses with log-linear smoothing. In this paper, we used semi-parametric marginal maximum likelihood estimation with log-linear smoothing as implemented in the R *sirt* package (Robitzsch, 2013) for estimation of the full datasets, regardless of the ability distribution. Our goal was to ensure consistency of estimation errors due to estimation method between different models.

4.5.2 Estimation of Parameters from Sampled Data Sets

Data sets of approximately 30% sample sizes were sampled from full datasets based on uniform, leverage, SLEV, and Adj-SLEV methods. Estimation of item parameters for each sampled data set was done by using the Markov Chain Monte Carlo (MCMC) method as implemented in the computer software OpenBUGS (Lunn et al., 2009). The following priors were used for MCMC estimation of item parameters:

$$\theta_j \sim Normal(0,1), \qquad j = 1, \dots, N,$$

$$b_i \sim Normal(0,1), \qquad i = 1, \dots, n,$$

$$a_i \sim Normal(0,1)\ and\ a_i > 0\ , \qquad i = 1, \dots, n,$$

$$c_i \sim Beta(5,17), \qquad i = 1, \dots, n,$$

where θ_j is ability of examinee j, and b_i is the item difficulty parameter, a_i is the item discrimination parameter and c_i is the pseudo-guessing parameter for item i, respectively.

CHAPTER 5

RESULTS

5.1 Accuracy Analyses

Analyses were conducted to compare accuracy of the parameter estimates from different data sampling methods. The full data sets with approximately normal and non-normal distributions given in Section 5 were sampled based on different data sampling methods given in Section 3. The estimated parameters from sampled data sets were analyzed for their accuracy compared to the estimates from full data sets. The indices used as a measure of accuracy were bias, root-mean-square error (RMSE), mean absolute error (MAE) and Pearson correlation. Before the accuracy analysis, the scales of estimates from data subsamples were all placed on the scale of the estimates from analysis of the full dataset by using mean/sigma equating method (Marco, 1977).

The bias, RMSE, MAE, and Pearson correlation were computed across 32 items and 50 replications for each sampling method. The following equations were used for calculating the accuracy indices for estimated item difficulty parameter from full data sets for item i ($\hat{b}_i$), and estimated item difficulty parameter from sampled data sets for item i from *r*th replication ($\hat{b}_{ir}$):

$$Bias(\hat{b}) = \frac{\sum_{r=1}^{50}\sum_{i=1}^{32}(\hat{b}_i - \hat{b}_{ir})}{50x32} \tag{26}$$

$$MAE(\hat{b}) = \frac{\sum_{r=1}^{50}\sum_{i=1}^{32}|\hat{b}_i - \hat{b}_{ir}|}{50x32} \tag{27}$$

$$RMSE(\hat{b}) = \sqrt{\frac{\sum_{r=1}^{50}\sum_{i=1}^{32}(\hat{b}_i - \hat{b}_{ir})^2}{50x32}} \tag{28}$$

$$Cor(\hat{b}, b) = \frac{1}{50}\sum_{r=1}^{50} Cor(\hat{b}_i, \hat{b}_{ir}) \tag{29}$$

5.2 Data Set with Normal Raw Score Distribution

Accuracy indices were calculated for comparing parameter estimates from the empirical data set with normal total score distribution and parameter estimates from samples of these datasets over 50 replications. The latent ability distribution was normal for Rasch and 2-pl models, and non-normal for 3-pl model (see Table 1). The results for different models from different sampling methods were compared in Tables 3-6. A factorial ANOVA test was conducted for each type of parameter using RMSE as the dependent variable, and sampling method as the independent variable (see Table 7). Pairwise comparisons with Bonferroni correction was administered for examining the significant differences in estimation accuracy between the sampling methods (see Table 7).

Results showed that the error in estimates increase as the number of parameters in the model increases, namely from Rasch model to 3-pl model (see Tables 3-6). For the Rasch model, there was a significant method effect for estimation of item difficulty (*b*) parameter (see Table 7). The smallest RMSE was achieved by random sampling method (see Tables 3-6), and pairwise comparisons indicated the RMSE from random sampling was significantly different than other sampling methods (see Table 7). For the 2-pl model, item discrimination (a) was best estimated by random sampling method (see Tables 3-7). The difference between random sampling and Adj-SLEV, however, was not significant at 0.01 level. Although the smallest RMSE for item difficulty (b) in 2-pl was produced by SLEV sampling method (see Tables 3-6), the RMSE estimates from leverage-based, SLEV, and Adj-SLEV methods were not significantly different than each other (see Table 7). Considering the trade-off between item discrimination and item difficulty parameters, Adj-SLEV sampling method resulted in the best recovery for 2-pl model. There was not a significant sampling method effect for estimation of item discrimination (a) in the 3-pl model (see Table 7). However, the smallest RMSEs were produced by random and Adj-SLEV methods (see Tables 3-6), and they were not significantly different than SLEV method (see Table 7). The smallest RMSE was achieved with SLEV sampling method for item difficulty (b) (see Tables 3-6), and it was not significantly different than the RMSE from leverage-based sampling method (see Table 7). RMSE from Adj-SLEV was significantly different than SLEV method, and it was not significantly different than leverage-based method at .05 level (see Table 7). The smallest RMSE was achieved by leverage-based sampling method for pseudo-guessing (c) parameters (see Tables 3-6). However, it was not significantly different than SLEV method (see Table 7). Adj-SLEV method was significantly different than leverage-based method, however was similar to SLEV method at .05 significance level. Determining a method that gives the best parameter estimates for all parameters in a 3-pl model was challenging. The trade-off for the parameter estimates apparently can be best achieved by using the SLEV method. Adj-SLEV sampling method may also produce a good trade-off which can be compared to results from the leverage-based sampling method.

Table 3: Accuracy Indices for Different Models from Random Sampling Method

	Random					
	Rasch	2-pl		3-pl		
	b	a	b	a	b	c
Bias	0.000	0.000	0.000	0.000	0.000	0.000
MAE	7.228	13.850	31.316	44.607	54.010	7.188
RMSE	8.950	17.645	44.935	57.795	78.842	12.206
Correlation	.998	.945	.957	.890	.755	.724

Note. Values for bias, MAE and RMSE are multiplied by 100.

Table 4: Accuracy Indices for Different Models from Leverage-based Sampling Method

	Leverage-based					
	Rasch	2-pl		3-pl		
	b	a	b	a	b	c
Bias	0.000	0.000	0.000	0.000	0.000	0.000
MAE	8.293	15.324	29.309	47.980	51.573	6.026
RMSE	10.492	20.675	41.768	61.518	75.475	11.098
Correlation	.997	.925	.963	.871	.776	.772

Note. Values for bias, MAE and RMSE are multiplied by 100.

Table 5: Accuracy Indices for Different Models from SLEV Sampling Method

	SLEV					
	Rasch	2-pl		3-pl		
	b	a	b	a	b	c
Bias	0.000	0.000	0.000	0.000	0.000	0.000
MAE	8.611	15.301	28.996	48.045	51.042	6.136
RMSE	10.812	20.550	41.221	61.246	74.802	11.182
Correlation	.997	9.257	.964	.872	.780	.769

Note. Values for bias, MAE and RMSE are multiplied by 100.

Table 6: Accuracy Indices for Different Models from Adj-SLEV

Sampling Method

	Adj-SLEV					
	Rasch	2-pl		3-pl		
	b	a	b	a	b	c
Bias	0.000	0.000	0.000	0.000	0.000	0.000
MAE	7.853	14.496	30.122	45.978	52.475	6.470
RMSE	9.889	19.319	42.720	58.742	76.487	11.582
Correlation	.997	.935	.961	.882	.770	.752

Note. Values for bias, MAE and RMSE are multiplied by 100.

Table 7: ANOVA and Pairwise Comparisons with Bonferroni Correction for RMSE

		ANOVA		Pairwise Comparisons			
Model	Parameter	*F*	*p*		Random	Leverage	SLEV
Rasch	*b*	9.923	.002	Leverage	<.001		
				SLEV	<.001	1.000	
				Adj-SLEV	.007	.360	.043
2-pl	*a*	6.828	.010	Leverage	<.001		
				SLEV	<.001	1.000	
				Adj-SLEV	.018	.204	.391
2-pl	*b*	7.243	.008	Leverage	.001		
				SLEV	<.001	1.000	
				Adj-SLEV	.047	1.000	.295
3-pl	*a*	0.380	.538	Leverage	.034		
				SLEV	.065	1.000	
				Adj-SLEV	1.000	.244	.409
3-pl	*b*	15.483	<.001	Leverage	<.001		
				SLEV	<.001	1.000	
				Adj-SLEV	<.001	.434	.018
3-pl	c	10.431	.001	Leverage	<.001		
				SLEV	<.001	1.000	
				Adj-SLEV	.001	.010	.056

Note. 1) Log transformation was applied to parameter estimates.

5.3 Data Set with Non-normal Raw Score Distribution

In this section, accuracy indices were compared for parameters from the empirical data set with the non-normal total score distribution and for parameters from subsamples of the non-normal dataset over 50 replications. The latent ability distribution was found to be non-normal for Rasch, 2-pl and 3-pl models (see Table 2). The accuracy indices for different models from different sampling methods were shown in Tables 8-11. The differences in RMSE estimates from different sampling methods were examined by factorial ANOVA test for each type of parameter (see Table 12). Pairwise comparisons with Bonferroni correction was administered for further investigation of differences in RMSE between the sampling methods (see Table 12).

Results showed that the errors from Empirical Study 2 were larger compared to the errors from Empirical Study 1, due to non-normality of the ability distribution. The errors also increased as the number of parameters in the model increased, similar to previous results in section 5.2. For the Rasch model, best recovery was achieved by the random sampling method (see Tables 8-12). The sampling method was not significant for estimation of 2-pl model parameters at .05 significance level (see Table 12), however there was a method effect for estimation of item discrimination (a) at .10 significance level. The smallest error was achieved by leverage-based and Adj-SLEV sampling methods for estimation of item discrimination(a) (see Tables 8-12), and these two were not significantly different than SLEV method (see Table 12). Similarly, random and Adj-SLEV sampling methods provided smallest errors for estimation of item difficulty (b) (see Tables 8-11), and they were not significantly different than SLEV method, although they were marginally different than leverage-based method (see Table 12). Considering the trade-off between item discrimination and item difficulty parameters, the Adj-SLEV sampling method can be used to result in the smallest RMSEs. The sampling methods which resulted in the smallest recovery indices varied for the 3-pl model depending on the type of the accuracy index. The best MAE for 3-pl was achieved with shrinkage based sampling method for item discrimination (a), with Adj-SLEV based sampling for item difficulty(b), and with leverage-based sampling method for pseudo-guessing (c) parameters (see Tables 8-12). The smallest RMSE for 3-pl model was achieved with the Adj-SLEV for item discrimination (a) (see Tables 8-12), although the results from different sampling methods were not significantly different (see Table 12). The sampling method effect was significant at .10 significance level for estimation of item difficulty (b), however it was not significant at .05 significance level. Adj-SLEV gave the smallest RMSE estimate for item difficulty (b) (see Tables 8-12), and it was not significantly different than random sampling method (see Table 12). The best RMSE estimates for estimation of pseudo-guessing (c) was from leverage-based and Adj-SLEV sampling methods (see Tables 8-12), and it was not significantly different than the estimate from SLEV method (see Table 12). The trade-off for all parameter estimates suggested using the Adj-SLEV method for estimation of this parameter. Overall, Adj-SLEV can be used for considering the trade-off between item discrimination, item difficulty and item pseudo-guessing parameters in 3-pl model.

Table 8: Accuracy Indices for Different Models from Random Sampling Method

	Random					
	Rasch	2-pl		3-pl		
	b	a	b	a	b	c
Bias	0.000	0.000	0.000	0.000	0.000	0.000
MAE	9.034	11.507	14.187	120.953	31.804	4.807
RMSE	11.645	14.872	19.515	154.669	41.402	8.419
Correlation	.997	.957	.988	.805	.849	.706

Note. Values for bias, MAE and RMSE are multiplied by 100.

Table 9: Accuracy Indices for Different Models from Leverage-based Sampling Method

	Leverage-based					
	Rasch	2-pl		3-pl		
	b	a	b	a	b	c
Bias	0.000	0.000	0.000	0.000	0.000	0.000
MAE	11.051	10.624	15.485	115.162	33.129	4.437
RMSE	13.855	13.701	21.085	155.653	42.754	8.086
Correlation	.996	.963	.987	.803	.839	.729

Note. Values for bias, MAE and RMSE are multiplied by 100.

Table 10: Accuracy Indices for Different Models from SLEV Sampling Method

	SLEV					
	Rasch	2-pl		3-pl		
	b	a	b	a	b	c
Bias	0.000	0.000	0.000	0.000	0.000	0.000
MAE	11.098	11.230	15.088	113.476	32.445	4.502
RMSE	13.738	14.759	20.555	151.424	42.003	8.150
Correlation	.996	.958	.987	.813	.844	.724

Note. Values for bias, MAE and RMSE are multiplied by 100.

Table 11: Accuracy Indices for Different Models from Adj-SLEV Sampling Method

	Adj-SLEV					
	Rasch	2-pl		3-pl		
	b	a	b	a	b	c
Bias	0.000	0.000	0.000	0.000	0.000	0.000
MAE	11.128	10.665	14.464	114.000	31.678	4.526
RMSE	13.958	13.705	19.578	149.892	40.857	8.093
Correlation	.996	.963	.988	.816	.853	.728

Note. Values for bias, MAE and RMSE are multiplied by 100.

Table 12: ANOVA and Pairwise Comparisons with Bonferroni Correction for RMSE

		ANOVA		Pairwise Comparisons			
Model	Parameter	F	p		Random	Leverage	SLEV
Rasch	b	40.762	<.001	Leverage	<.001		
				SLEV	<.001	1.000	
				Adj-SLEV	<.001	1.000	1.000
2pl	a	3.210	.075	Leverage	.044		
				SLEV	1.000	.078	
				Adj-SLEV	.037	1.000	.067
2pl	b	0.001	.982	Leverage	.024		
				SLEV	.242	1.000	
				Adj-SLEV	1.000	.052	.434
3pl	a	4.003	.047	Leverage	1.000		
				SLEV	1.000	.840	
				Adj-SLEV	.650	.290	1.000
3pl	b	3.582	.060	Leverage	.006		
				SLEV	.836	.370	
				Adj-SLEV	.867	<.001	.021
3pl	c*	3.038	.030	Leverage	.001		
				SLEV	.021	.520	
				Adj-SLEV	.004	.942	.605

Note. 1) Log transformation was applied to parameter estimates.
2) Welch's correction for unequal variances was shown with "*" if applied. Correction was also applied to pairwise comparisons.

CHAPTER 6

DISCUSSION

In this study, we compared different data sampling methods for Bayesian estimation of IRT model parameters. These methods were random, leverage-based, shrinkage leverage-based (SLEV), and adjusted shrinkage leverage-based (Adj-SLEV) sampling methods. Estimation of item parameters in IRT models were our interest in this study. Two empirical data sets consisting of binary scored responses to mathematics achievement items were provided. These data sets had normally and non-normally distributed total score distributions. Semi-parametric estimation of data sets with log-linear smoothing indicated normal ability distribution for the Rasch, and 2-pl models, and non-normal ability distribution for the 3-pl model for the data set with normal total score distribution. Similarly, semi-parametric estimation of data sets with log-linear smoothing indicated non-normal ability distribution for each of the Rasch, 2-pl, and 3-pl model for the data set with non-normal total score distribution. The MCMC method was administered for Bayesian estimation of the sampled data sets. Bayesian estimation requires determining a prior distribution for parameters to be estimated. The convention is assuming a normal prior distribution for the ability distribution. The errors in item parameter estimates may increase when the normality assumption for ability is violated.

Results showed that the errors in parameter estimates were higher when the ability distribution was non-normal. Errors also increased as the number of parameters in the model increased for both normally and non-normally distributed ability. The sampling method that provides the best item parameter estimates varied based on the model, based on the specific parameter in a model, and based on the ability distribution. The random sampling method appeared to provide best item parameter estimates for the Rasch model, both for the data sets with normal and non-normal ability distributions. For 2-pl model, the sampling methods exhibited a differential effect on item parameter estimation for the normally distributed ability, however not for the non-normally distributed ability. For the normal ability, the sampling method that provides the best estimate varied for item difficulty and item discrimination parameters, when they were evaluated individually. Adj-SLEV method, on the other hand, provided best estimates for this model when the results for both item parameters were considered together. For non-normal ability, Adj-SLEV and SLEV methods provided the best estimates for both type of item parameters based on the pairwise comparison tests with Bonferroni correction.

The effect of sampling method on estimation of 3-pl model varied depending on the parameter type and ability distribution. For normal ability, the sampling method effect was insignificant for estimation of item discrimination, although it was significant for estimation of item difficulty and item guessing parameters. Results from factorial ANOVA and pairwise comparisons yielded leverage-based, SLEV and Adj-SLEV methods to perform comparable when all parameters in a 3-pl model were considered together, although SLEV method may outperform the other two. For non-normal ability, the sampling method effect was marginally significant for estimation of item discrimination, insignificant for estimation of item difficulty, and significant for estimation of item guessing parameter. The trade-off for all parameter estimates suggested using the Adj-SLEV method for estimation of this parameter.

Overall, the most accurate estimates of item parameters were from random sampling method for Rasch model, and from either SLEV or Adj-SLEV for 2-pl and 3-pl models considering all parameters in the model. For these models, Adj-SLEV either provided the best estimates, or was a good alternative of the best model. Random sampling method, on the other hand, did not provide as accurate results as other sampling methods for 2-pl and 3-pl models when all parameters in the models were considered together.

BIBLIOGRAPHY

[1] Angoff, W. H. (1971). Scales, norms and equivalent scores. In R. L. Thorndike (Ed.), Educational measurement (2nd ed.). Washington, DC: American Council on Education.

[2] Anscombe, F. J., & Glynn, W. J. (1983). Distribution of the kurtosis statistic b2 for normal samples. Biometrika, 70(1), 227-234.

[3] Baker, F. (2001). The Basics of item response theory (Second edition). College Park: MD: ERIC Clearinghouse on Assessment and Evaluation. Retrieved from http://files.eric.ed.gov/fulltext/ED458219.pdf

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

[5] Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord, & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397-479). Reading, MA: Addison-Wesley.

[6] Cohen, M. B., Lee, Y. T., Musco, C., Musco, C., Peng, R., & Sidford, A. (2015, January). Uniform sampling for matrix approximation. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science* (pp. 181-190). ACM.

[7] Cox, D. R. (1958). *Planning of experiments*. New York: John Wiley.

[8] D'Agostino, R. B. (1970). Transformation to normality of the null distribution of g1. *Biometrika*, *57*(3), 679-681.

[9] Dai, Y. (2013). A mixture Rasch model with a covariate: A simulation study via Bayesian

Markov Chain Monte Carlo estimation. *Applied Psychological Measurement*, *37*, 375-396.

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

[11] Embretson, S. E. (1996). The new rules of measurement. *Psychological Assessment*, *8,* 341.

[12] Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, N.J.: Psychology Press.

[13] Fox, J. (1991). *Regression Diagnostics*. Sage: Newbury Park, CA.

[14] Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Sage: Newbury Park, CA.

[15] Ho, A. D., & Yu, C. C. (2015). Descriptive Statistics for Modern Test Score Distributions Skewness, Kurtosis, Discreteness, and Ceiling Effects. *Educational and Psychological Measurement*, *75*(3), 365-388.

[16] Hoaglin, D. C., & Welsch, R. E. (1978). The hat matrix in regression and ANOVA. *American Statistician*, *32*(1), 17-22.

[17] Holland, P. W. (1990). On the sampling theory foundations of item response theory models. *Psychometrika*, *55*, 577-601.

[18] Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York: Springer-Verlag.

[19] Komsta, L., & and Novomestky, F. (2015). moments: Moments, cumulants, skewness, kurto sis and related tests. R package version 0.14. Retrieved from http://CRAN.R-project.org/package=moments

[20] Lindsay, B., Clogg, C. C., & Grego, J. (1991). Semiparametric estimation in the Rasch mod

el and related exponential response models, including a simple latent class model for item analysis. *Journal of the American Statistical Association*, *86*(413), 96-107.

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

[22] Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores* (with contributions by A. Birnbaum). Reading, MA: Addison-Wesley.

[23] Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: Evolution, critique and future directions. *Statistics in medicine*, *28*(25), 3049.

[24] Ma, P., Mahoney, M. W., & Yu, B. (2015, January). A statistical perspective on algorithmic leveraging. *Journal of Machine Learning Research*, *16*(1), 861-911.

[25] Marco, G. L. (1977). Item characteristic curve solutions to three intractable testing problems. *Journal of Educational Measurement*, *14*, 139–160.

[26] Mislevy, R. L. (1986). Bayes modal estimation in item response models. *Psychometrika*, *51*, 177-195.

[27] OECD (2013), *PISA 2012 assessment and analytical framework: mathematics, reading, science, problem solving and financial literacy,* OECD Publishing. Retrieved from http://dx.doi.org/10.1787/9789264190511-en

[28] *OECD (2014), PISA 2012 results in focus.* Retrieved November 02, 2015, from http://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf

[29] R Core Team (2014). R: A language and environment for statistical computing. *R Foundation for Statistical Computing.* Vienna, Austria. URL http://www.R-project.org/.

[30] Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.*

Copenhagen: Nielson and Lydiche (for Danmarks Paedagogiske Institut).

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

[32] Robitzsch, A. (2014). *sirt: Supplementary Item Response Theory Models R package version 0.43-70* [Computer Software]. Retrieved from https://cran.r-project.org/package=sirt

[33] Robitzsch, A. (2015). *Package 'sirt'*. Retrieved from https://cran.r-project.org/web/packages/sirt/sirt.pdf

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

[35] Stocking, M. L. (1990). Specifying optimum examinees for item parameter estimation in item response theory. *Psychometrika, 55*, 461-475.

[36] Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied psychological measurement, 7*(2), 201-210.

[37] *Tay, L., Vermunt, J. K., & Wang, C. (2013).* Assessing the item response theory with covariate (IRT-C) procedure for ascertaining differential item functioning. *International Journal of Testing, 13*(3), *201–222.*

[38] Van der Linden, W. J., & Hambleton, R. K. (Eds.). (1997). *Handbook of modern item response theory*. New York: Springer.

[39] Wellner, J. A. (1986). Semiparametric models: progress and problems. *Centrum voor Wiskunde en Informatica, Report.*

[40] Xu, X., & von Davier, M. (2008). *Fitting the structured general diagnostic model to NAEP data*. ETS Research Report ETS RR-08-27. Princeton, NJ: Educational Testing Service.

APPENDIX A

ITEM PARAMETER ESTIMATES FROM APPROXIMATELY NORMAL DATA

(EMPIRICAL STUDY 1)

Table A1: Item Difficulty Estimates from Different Sampling Methods for Rasch Model

		Random				Leverage			
Item	bFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item1	1.962	1.942	1.736	2.238	0.095	1.964	1.781	2.210	0.086
Item2	-1.478	-1.466	-1.632	-1.257	0.090	-1.505	-1.728	-1.312	0.102
Item3	0.531	0.529	0.254	0.650	0.090	0.573	0.345	0.815	0.099
Item4	-0.897	-0.899	-1.068	-0.722	0.071	-0.812	-1.013	-0.609	0.089
Item5	-0.296	-0.316	-0.485	-0.152	0.082	-0.238	-0.424	-0.047	0.092
Item6	-0.067	-0.054	-0.242	0.134	0.081	-0.131	-0.401	0.058	0.099
Item7	0.261	0.253	0.025	0.509	0.113	0.318	0.134	0.475	0.078
Item8	-0.117	-0.142	-0.331	-0.007	0.079	-0.014	-0.190	0.158	0.073
Item9	0.099	0.092	-0.102	0.228	0.081	0.075	-0.091	0.357	0.085
Item10	-0.473	-0.483	-0.645	-0.319	0.071	-0.468	-0.699	-0.260	0.089
Item11	0.792	0.801	0.671	0.933	0.059	0.773	0.646	0.906	0.063
Item12	-1.249	-1.229	-1.409	-0.971	0.082	-1.226	-1.431	-1.004	0.094
Item13	3.116	3.095	2.857	3.410	0.120	3.073	2.875	3.271	0.080
Item14	-1.125	-1.120	-1.386	-0.923	0.095	-1.167	-1.377	-1.007	0.078
Item15	1.251	1.253	1.084	1.479	0.093	1.265	1.105	1.424	0.080
Item16	-0.908	-0.883	-1.099	-0.662	0.096	-0.875	-1.071	-0.685	0.084

Table A1 Continued: Item Difficulty Estimates from Different Sampling Methods

for Rasch Model

		SLEV				Adj-SLEV			
Item	bFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item1	1.962	1.958	1.741	2.166	0.092	1.967	1.796	2.181	0.099
Item2	-1.478	-1.497	-1.673	-1.224	0.103	-1.487	-1.718	-1.301	0.099
Item3	0.531	0.597	0.392	0.829	0.093	0.569	0.396	0.783	0.096
Item4	-0.897	-0.804	-1.015	-0.552	0.105	-0.856	-1.043	-0.640	0.080
Item5	-0.296	-0.224	-0.512	0.013	0.109	-0.237	-0.471	-0.051	0.101
Item6	-0.067	-0.101	-0.299	0.098	0.089	-0.101	-0.309	0.081	0.094
Item7	0.261	0.342	0.141	0.581	0.093	0.318	0.059	0.536	0.099
Item8	-0.117	-0.055	-0.220	0.141	0.083	-0.062	-0.218	0.150	0.084
Item9	0.099	0.071	-0.085	0.245	0.083	0.120	-0.083	0.287	0.082
Item10	-0.473	-0.494	-0.651	-0.303	0.079	-0.461	-0.702	-0.287	0.089
Item11	0.792	0.756	0.580	0.987	0.087	0.788	0.490	0.964	0.088
Item12	-1.249	-1.233	-1.418	-0.987	0.096	-1.237	-1.428	-1.051	0.085
Item13	3.116	3.073	2.852	3.331	0.097	3.111	2.920	3.363	0.111
Item14	-1.125	-1.137	-1.306	-0.989	0.073	-1.137	-1.285	-0.956	0.076
Item15	1.251	1.248	1.046	1.427	0.084	1.246	1.081	1.418	0.079
Item16	-0.908	-0.880	-1.151	-0.591	0.111	-0.885	-1.077	-0.717	0.084

Table A1 Continued: Item Difficulty Estimates from Different Sampling Methods for Rasch Model

		Random				Leverage			
Item	bFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item17	-0.918	-0.934	-1.113	-0.776	0.090	-0.922	-1.125	-0.758	0.085
Item18	-2.198	-2.216	-2.457	-1.985	0.100	-2.333	-2.716	-2.032	0.133
Item19	1.070	1.078	0.945	1.237	0.074	1.049	0.887	1.276	0.085
Item20	0.648	0.667	0.489	0.851	0.086	0.707	0.476	0.897	0.077
Item21	-0.656	-0.633	-0.83	-0.430	0.099	-0.619	-0.829	-0.427	0.091
Item22	0.936	0.936	0.695	1.212	0.100	0.968	0.802	1.086	0.071
Item23	-0.799	-0.794	-0.989	-0.646	0.083	-0.802	-0.96	-0.551	0.09
Item24	0.299	0.299	0.090	0.465	0.084	0.295	0.110	0.466	0.095
Item25	-1.323	-1.332	-1.532	-1.101	0.079	-1.270	-1.490	-1.042	0.095
Item26	-2.869	-2.873	-3.163	-2.610	0.128	-2.964	-3.209	-2.709	0.116
Item27	0.549	0.546	0.373	0.760	0.096	0.626	0.449	0.785	0.082
Item28	-0.883	-0.868	-1.067	-0.675	0.087	-0.841	-1.043	-0.690	0.077
Item29	-0.251	-0.245	-0.468	-0.093	0.087	-0.254	-0.431	-0.085	0.073
Item30	-1.693	-1.684	-1.874	-1.489	0.080	-1.720	-2.002	-1.418	0.101
Item31	1.999	2.005	1.808	2.178	0.090	1.946	1.712	2.078	0.075
Item32	3.044	3.032	2.841	3.358	0.104	2.886	2.665	3.117	0.099

Table A1 Continued: Item Difficulty Estimates from Different Sampling Methods for Rasch Model

		SLEV				Adj-SLEV			
Item	bFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item17	-0.918	-0.917	-1.157	-0.732	0.090	-0.905	-1.194	-0.647	0.089
Item18	-2.198	-2.345	-2.582	-2.146	0.097	-2.284	-2.639	-2.060	0.123
Item19	1.070	1.055	0.872	1.205	0.081	1.065	0.853	1.222	0.075
Item20	0.648	0.710	0.462	0.907	0.099	0.681	0.471	0.825	0.084
Item21	-0.656	-0.630	-0.812	-0.423	0.100	-0.642	-0.853	-0.487	0.075
Item22	0.936	0.994	0.806	1.168	0.079	0.953	0.713	1.079	0.083
Item23	-0.799	-0.800	-1.038	-0.62	0.092	-0.815	-0.996	-0.609	0.082
Item24	0.299	0.281	0.022	0.570	0.096	0.276	0.103	0.450	0.085
Item25	-1.323	-1.290	-1.487	-1.099	0.097	-1.324	-1.573	-1.127	0.085
Item26	-2.869	-2.965	-3.183	-2.736	0.127	-2.905	-3.138	-2.646	0.119
Item27	0.549	0.606	0.400	1.008	0.096	0.609	0.375	0.805	0.100
Item28	-0.883	-0.832	-1.036	-0.633	0.098	-0.865	-1.072	-0.617	0.096
Item29	-0.251	-0.234	-0.413	-0.044	0.083	-0.262	-0.476	-0.101	0.080
Item30	-1.693	-1.715	-1.916	-1.526	0.091	-1.739	-2.019	-1.503	0.107
Item31	1.999	1.920	1.769	2.129	0.078	1.958	1.771	2.152	0.077
Item32	3.044	2.899	2.659	3.103	0.093	2.901	2.676	3.102	0.101

Table A2: Item Difficulty Estimates from Different Sampling Methods for 2-pl Model

		Random				Leverage			
Item	bFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item1	1.445	1.833	1.613	2.168	0.135	1.818	1.573	2.138	0.113
Item2	-1.519	-2.016	-2.333	-1.537	0.185	-2.001	-2.415	-1.669	0.167
Item3	0.474	0.520	0.182	0.666	0.108	0.499	0.232	0.771	0.113
Item4	-0.688	-0.984	-1.176	-0.752	0.092	-0.907	-1.085	-0.726	0.079
Item5	-0.276	-0.469	-0.747	-0.254	0.114	-0.390	-0.598	-0.149	0.107
Item6	-0.055	-0.140	-0.362	0.101	0.116	-0.272	-0.591	-0.018	0.130
Item7	0.298	0.286	-0.024	0.653	0.151	0.285	0.054	0.483	0.096
Item8	-0.060	-0.203	-0.388	-0.084	0.077	-0.116	-0.268	0.061	0.068
Item9	0.112	0.043	-0.207	0.222	0.094	-0.022	-0.202	0.261	0.092
Item10	-0.409	-0.633	-0.891	-0.432	0.103	-0.622	-0.875	-0.363	0.108
Item11	0.549	0.616	0.518	0.770	0.058	0.578	0.448	0.683	0.062
Item12	-1.206	-1.601	-1.917	-1.280	0.145	-1.530	-1.813	-1.303	0.138
Item13	1.733	2.409	2.165	2.898	0.126	2.526	2.279	2.758	0.108
Item14	-0.881	-1.214	-1.455	-0.922	0.117	-1.287	-1.510	-1.057	0.097
Item15	0.856	1.021	0.848	1.228	0.096	1.034	0.851	1.225	0.084
Item16	-1.118	-1.474	-2.006	-1.066	0.192	-1.371	-1.716	-1.004	0.150

Table A2 Continued: Item Difficulty Estimates from Different Sampling Methods

for 2-pl Model

		SLEV				Adj-SLEV			
Item	bFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item1	1.445	1.819	1.523	2.188	0.143	1.831	1.597	2.096	0.117
Item2	-1.519	-1.963	-2.336	-1.545	0.159	-1.973	-2.450	-1.750	0.173
Item3	0.474	0.527	0.313	0.794	0.104	0.520	0.363	0.827	0.114
Item4	-0.688	-0.897	-1.100	-0.646	0.111	-0.938	-1.145	-0.761	0.084
Item5	-0.276	-0.377	-0.706	-0.075	0.129	-0.384	-0.624	-0.140	0.118
Item6	-0.055	-0.234	-0.495	0.008	0.118	-0.232	-0.470	0.100	0.130
Item7	0.298	0.314	0.029	0.594	0.116	0.307	-0.017	0.639	0.125
Item8	-0.060	-0.155	-0.312	0.048	0.080	-0.151	-0.315	0.022	0.080
Item9	0.112	-0.026	-0.206	0.166	0.089	0.039	-0.171	0.205	0.093
Item10	-0.409	-0.650	-0.824	-0.461	0.094	-0.610	-0.947	-0.453	0.100
Item11	0.549	0.564	0.408	0.777	0.075	0.587	0.355	0.744	0.074
Item12	-1.206	-1.535	-1.845	-1.225	0.133	-1.542	-1.759	-1.169	0.103
Item13	1.733	2.524	2.297	2.856	0.122	2.535	2.225	2.941	0.151
Item14	-0.881	-1.255	-1.491	-1.035	0.100	-1.250	-1.446	-1.027	0.102
Item15	0.856	1.000	0.805	1.148	0.076	1.022	0.863	1.218	0.077
Item16	-1.118	-1.398	-1.773	-0.992	0.168	-1.452	-1.930	-1.076	0.185

Table A2 Continued: Item Difficulty Estimates from Different Sampling Methods

for 2-pl Model

		Random				Leverage			
Item	bFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item17	-0.934	-1.308	-1.723	-0.99	0.166	-1.230	-1.553	-1.027	0.112
Item18	-5.941	-4.208	-4.545	-3.796	0.195	-4.400	-4.892	-3.898	0.219
Item19	0.733	0.857	0.703	1.028	0.073	0.827	0.647	1.013	0.077
Item20	0.545	0.647	0.463	0.935	0.097	0.609	0.411	0.825	0.088
Item21	-0.569	-0.802	-1.096	-0.548	0.132	-0.783	-1.017	-0.578	0.098
Item22	0.791	0.949	0.764	1.258	0.13	0.880	0.730	1.088	0.082
Item23	-0.676	-0.950	-1.182	-0.764	0.107	-0.957	-1.305	-0.689	0.109
Item24	0.299	0.299	0.034	0.552	0.111	0.228	-0.045	0.421	0.107
Item25	-1.218	-1.654	-1.957	-1.370	0.13	-1.518	-1.827	-1.286	0.133
Item26	-4.405	-4.143	-4.684	-3.552	0.256	-4.096	-4.773	-3.564	0.232
Item27	0.589	0.676	0.426	0.989	0.13	0.663	0.334	0.873	0.103
Item28	-0.841	-1.149	-1.550	-0.896	0.136	-1.058	-1.371	-0.888	0.104
Item29	-0.156	-0.294	-0.462	-0.148	0.082	-0.346	-0.497	-0.171	0.076
Item30	-1.411	-1.863	-2.202	-1.520	0.153	-1.932	-2.174	-1.610	0.123
Item31	1.302	1.669	1.394	1.954	0.13	1.621	1.413	1.777	0.075
Item32	1.681	2.321	2.111	2.774	0.135	2.314	2.051	2.624	0.116

Table A2 Continued: Item Difficulty Estimates from Different Sampling Methods

for 2-pl Model

		SLEV				Adj-SLEV			
Item	bFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item17	-0.934	-1.231	-1.553	-1.025	0.110	-1.249	-1.549	-0.959	0.137
Item18	-5.941	-4.421	-4.892	-4.046	0.204	-4.347	-4.770	-3.812	0.216
Item19	0.733	0.831	0.647	0.975	0.072	0.836	0.672	0.962	0.064
Item20	0.545	0.614	0.411	0.854	0.104	0.599	0.386	0.774	0.087
Item21	-0.569	-0.792	-1.017	-0.576	0.111	-0.813	-1.078	-0.599	0.092
Item22	0.791	0.912	0.730	1.213	0.092	0.901	0.622	1.073	0.103
Item23	-0.676	-0.941	-1.305	-0.720	0.123	-0.962	-1.168	-0.711	0.102
Item24	0.299	0.215	-0.045	0.522	0.111	0.221	0.034	0.481	0.104
Item25	-1.218	-1.531	-1.827	-1.213	0.130	-1.583	-1.839	-1.303	0.129
Item26	-4.405	-4.112	-4.773	-3.656	0.240	-4.094	-4.551	-3.640	0.212
Item27	0.589	0.630	0.334	1.089	0.116	0.679	0.378	1.000	0.133
Item28	-0.841	-1.050	-1.371	-0.781	0.135	-1.109	-1.362	-0.794	0.123
Item29	-0.156	-0.325	-0.497	-0.153	0.084	-0.342	-0.617	-0.179	0.084
Item30	-1.411	-1.913	-2.174	-1.645	0.132	-1.908	-2.203	-1.607	0.140
Item31	1.302	1.593	1.413	1.888	0.093	1.633	1.418	1.844	0.087
Item32	1.681	2.307	2.051	2.531	0.104	2.274	1.999	2.586	0.117

Table A3: Item Discrimination Estimates from Different Sampling Methods

for 2-pl Model

		Random				Leverage			
Item	aFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item1	1.514	1.507	1.185	1.925	0.185	1.474	1.193	1.951	0.194
Item2	0.897	0.850	0.634	1.150	0.122	0.791	0.593	1.110	0.106
Item3	1.194	1.198	0.890	1.567	0.139	1.235	0.982	1.532	0.128
Item4	1.350	1.411	1.095	1.889	0.172	1.519	1.135	1.997	0.177
Item5	0.933	0.889	0.685	1.165	0.112	0.856	0.452	1.162	0.123
Item6	0.805	0.730	0.448	0.973	0.108	0.739	0.546	0.981	0.096
Item7	0.857	0.780	0.580	1.031	0.108	0.800	0.531	1.039	0.103
Item8	1.455	1.529	1.247	1.988	0.170	1.622	1.353	2.059	0.165
Item9	1.114	1.125	0.870	1.474	0.136	1.126	0.792	1.442	0.127
Item10	1.070	1.059	0.763	1.338	0.121	1.100	0.885	1.596	0.124
Item11	2.035	2.172	1.834	2.517	0.179	2.323	1.980	2.681	0.183
Item12	0.970	0.950	0.679	1.201	0.128	0.964	0.734	1.310	0.132
Item13	2.650	2.378	1.928	2.811	0.202	2.143	1.726	2.466	0.186
Item14	1.327	1.390	1.053	1.834	0.166	1.398	1.119	1.739	0.145
Item15	1.909	2.005	1.635	2.659	0.205	1.986	1.549	2.323	0.174
Item16	0.702	0.612	0.385	0.880	0.110	0.546	0.333	0.748	0.106

Table A3 Continued: Item Discrimination Estimates from Different Sampling Methods for 2-pl Model

		SLEV				Adj-SLEV			
Item	aFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item1	1.514	1.451	1.164	1.970	0.175	1.477	1.006	1.777	0.164
Item2	0.897	0.823	0.563	1.116	0.119	0.839	0.602	1.262	0.117
Item3	1.194	1.206	0.902	1.519	0.125	1.184	0.953	1.682	0.147
Item4	1.350	1.505	1.205	2.078	0.169	1.512	1.121	1.981	0.162
Item5	0.933	0.834	0.500	1.182	0.123	0.865	0.591	1.125	0.116
Item6	0.805	0.739	0.510	0.966	0.109	0.730	0.525	0.943	0.095
Item7	0.857	0.807	0.621	1.098	0.117	0.801	0.547	1.030	0.128
Item8	1.455	1.633	1.339	1.903	0.135	1.605	1.348	2.050	0.150
Item9	1.114	1.088	0.769	1.322	0.118	1.120	0.864	1.309	0.111
Item10	1.070	1.094	0.738	1.315	0.123	1.093	0.832	1.468	0.125
Item11	2.035	2.259	1.804	2.770	0.204	2.280	1.875	2.624	0.189
Item12	0.970	0.966	0.595	1.308	0.123	0.988	0.742	1.338	0.116
Item13	2.650	2.127	1.614	2.511	0.190	2.187	1.764	2.586	0.219
Item14	1.327	1.396	1.018	1.782	0.173	1.395	1.071	1.769	0.153
Item15	1.909	2.084	1.765	2.478	0.169	1.925	1.579	2.387	0.174
Item16	0.702	0.531	0.338	0.730	0.088	0.543	0.291	0.839	0.126

Table A3 Continued: Item Discrimination Estimates from Different Sampling Methods for 2-pl Model

		Random				Leverage			
Item	aFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item17	0.893	0.850	0.610	1.140	0.121	0.844	0.634	1.066	0.113
Item18	0.305	0.447	0.273	0.594	0.077	0.274	0.061	0.558	0.107
Item19	1.964	2.103	1.715	2.574	0.198	2.157	1.726	2.550	0.188
Item20	1.307	1.271	0.926	1.621	0.148	1.379	1.051	1.705	0.144
Item21	1.093	1.081	0.840	1.397	0.134	1.109	0.833	1.447	0.132
Item22	1.253	1.234	0.925	1.571	0.162	1.336	0.970	1.649	0.135
Item23	1.149	1.172	0.959	1.436	0.120	1.227	0.884	1.712	0.154
Item24	1.043	1.009	0.732	1.298	0.142	1.012	0.817	1.347	0.129
Item25	1.040	1.037	0.772	1.369	0.145	1.080	0.718	1.473	0.149
Item26	0.568	0.765	0.600	1.126	0.115	0.699	0.506	1.010	0.121
Item27	0.885	0.820	0.468	1.040	0.115	0.808	0.534	1.073	0.121
Item28	0.974	0.955	0.589	1.331	0.160	1.028	0.726	1.321	0.139
Item29	1.546	1.646	1.252	1.911	0.133	1.614	1.358	1.886	0.132
Item30	1.222	1.298	0.845	1.678	0.167	1.234	0.890	1.609	0.135
Item31	1.978	1.990	1.532	2.564	0.199	2.012	1.645	2.382	0.177
Item32	2.750	2.488	2.118	3.006	0.205	2.319	2.017	2.879	0.186

Table A3 Continued: Item Discrimination Estimates from Different Sampling Methods for 2-pl Model

		SLEV				Adj-SLEV			
Item	aFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item17	0.893	0.833	0.630	1.094	0.111	0.816	0.592	1.136	0.125
Item18	0.305	0.283	0.117	0.443	0.074	0.320	0.156	0.442	0.067
Item19	1.964	2.122	1.767	2.486	0.171	2.133	1.784	2.446	0.167
Item20	1.307	1.362	1.105	1.679	0.127	1.358	0.985	1.605	0.128
Item21	1.093	1.121	0.915	1.432	0.122	1.096	0.784	1.377	0.138
Item22	1.253	1.310	1.057	1.634	0.134	1.269	1.038	1.572	0.137
Item23	1.149	1.277	1.021	1.771	0.186	1.239	0.976	1.586	0.150
Item24	1.043	0.990	0.686	1.367	0.135	1.026	0.749	1.389	0.142
Item25	1.040	1.102	0.839	1.439	0.150	1.090	0.812	1.389	0.143
Item26	0.568	0.701	0.479	1.107	0.131	0.708	0.537	0.930	0.094
Item27	0.885	0.841	0.637	1.203	0.110	0.802	0.534	1.067	0.116
Item28	0.974	1.021	0.758	1.337	0.144	0.987	0.749	1.284	0.129
Item29	1.546	1.627	1.256	2.027	0.161	1.622	1.308	1.966	0.150
Item30	1.222	1.247	1.005	1.531	0.123	1.313	0.958	1.638	0.142
Item31	1.978	2.022	1.651	2.613	0.206	1.988	1.536	2.515	0.206
Item32	2.750	2.350	1.985	3.009	0.194	2.443	1.969	2.902	0.203

Table A4: Item Difficulty Estimates from Different Sampling Methods

for 3-pl Model

		Random				Leverage			
Item	bFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item1	0.920	1.601	1.405	1.869	0.103	1.609	1.431	1.831	0.087
Item2	-0.044	-1.098	-1.374	-0.768	0.135	-1.045	-1.479	-0.771	0.147
Item3	0.467	0.665	0.357	0.811	0.103	0.636	0.411	0.872	0.099
Item4	-0.178	-0.559	-0.687	-0.398	0.069	-0.503	-0.645	-0.324	0.072
Item5	0.058	0.020	-0.227	0.235	0.081	0.048	-0.169	0.386	0.103
Item6	0.578	0.609	0.114	0.974	0.124	0.482	0.162	0.791	0.139
Item7	0.458	0.658	0.327	0.943	0.135	0.602	0.448	0.867	0.095
Item8	0.122	0.028	-0.125	0.149	0.064	0.078	-0.081	0.223	0.064
Item9	0.453	0.441	0.212	0.593	0.103	0.389	0.190	0.630	0.089
Item10	0.219	-0.063	-0.303	0.119	0.082	-0.057	-0.260	0.208	0.101
Item11	0.482	0.638	0.545	0.761	0.055	0.618	0.502	0.711	0.050
Item12	-0.299	-0.897	-1.082	-0.656	0.094	-0.899	-1.153	-0.698	0.114
Item13	1.070	2.069	1.847	2.496	0.107	2.190	1.961	2.416	0.096
Item14	0.124	-0.473	-0.788	-0.261	0.112	-0.437	-0.648	-0.231	0.099
Item15	0.634	0.961	0.802	1.124	0.079	0.984	0.830	1.113	0.069
Item16	-0.099	-0.540	-0.920	-0.148	0.152	-0.541	-0.750	-0.263	0.112

Table A4 Continues: Item Difficulty Estimates from Different Sampling Methods

for 3-pl Model

		SLEV				Adj-SLEV			
Item	bFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item1	0.920	1.606	1.384	1.891	0.107	1.616	1.416	1.807	0.091
Item2	-0.044	-1.010	-1.272	-0.482	0.156	-1.049	-1.441	-0.782	0.148
Item3	0.467	0.659	0.486	0.903	0.089	0.666	0.484	0.937	0.102
Item4	-0.178	-0.483	-0.663	-0.254	0.094	-0.530	-0.693	-0.360	0.067
Item5	0.058	0.068	-0.257	0.297	0.112	0.059	-0.186	0.258	0.101
Item6	0.578	0.504	0.253	0.756	0.117	0.498	0.251	0.720	0.107
Item7	0.458	0.628	0.460	0.837	0.100	0.627	0.402	1.000	0.118
Item8	0.122	0.050	-0.114	0.229	0.068	0.063	-0.073	0.212	0.068
Item9	0.453	0.404	0.247	0.652	0.081	0.448	0.239	0.659	0.079
Item10	0.219	-0.068	-0.278	0.136	0.102	-0.062	-0.285	0.158	0.097
Item11	0.482	0.614	0.483	0.776	0.063	0.629	0.472	0.758	0.062
Item12	-0.299	-0.891	-1.115	-0.641	0.113	-0.897	-1.114	-0.643	0.099
Item13	1.070	2.180	1.996	2.411	0.096	2.188	1.935	2.562	0.126
Item14	0.124	-0.427	-0.680	-0.224	0.112	-0.436	-0.719	-0.179	0.110
Item15	0.634	0.953	0.774	1.066	0.062	0.979	0.829	1.129	0.066
Item16	-0.099	-0.536	-0.871	-0.222	0.141	-0.546	-0.764	-0.359	0.100

Table A4 Continued: Item Difficulty Estimates from Different Sampling Methods

for 3-pl Model

		Random				Leverage			
Item	bFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item17	0.164	-0.417	-0.780	-0.109	0.134	-0.367	-0.711	-0.160	0.118
Item18	-5.600	-2.571	-2.856	-2.255	0.135	-2.757	-3.196	-2.418	0.163
Item19	0.573	0.829	0.699	0.972	0.064	0.817	0.680	0.958	0.062
Item20	0.467	0.735	0.573	1.011	0.090	0.683	0.539	0.870	0.078
Item21	-0.023	-0.266	-0.539	-0.064	0.110	-0.276	-0.548	-0.094	0.092
Item22	0.618	0.990	0.806	1.274	0.114	0.917	0.776	1.089	0.068
Item23	0.123	-0.335	-0.638	-0.141	0.096	-0.325	-0.508	-0.083	0.099
Item24	0.521	0.621	0.398	0.882	0.106	0.601	0.379	0.784	0.095
Item25	-0.376	-0.955	-1.108	-0.751	0.093	-0.850	-1.136	-0.618	0.120
Item26	-1.454	-3.042	-3.463	-2.585	0.188	-2.987	-3.373	-2.551	0.173
Item27	0.553	0.920	0.696	1.273	0.124	0.864	0.700	1.071	0.100
Item28	-0.204	-0.537	-0.791	-0.307	0.114	-0.525	-0.716	-0.300	0.084
Item29	0.236	0.031	-0.147	0.186	0.070	0.061	-0.153	0.225	0.076
Item30	0.017	-1.041	-1.256	-0.744	0.126	-0.989	-1.370	-0.648	0.149
Item31	0.853	1.461	1.246	1.649	0.101	1.446	1.250	1.563	0.061
Item32	1.037	1.989	1.797	2.287	0.107	2.003	1.801	2.215	0.093

Table A4 Continued: Item Difficulty Estimates from Different Sampling Methods

for 3-pl Model

		SLEV				Adj-SLEV			
Item	bFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item17	0.164	-0.390	-0.618	-0.042	0.122	-0.391	-0.680	-0.093	0.134
Item18	-5.600	-2.790	-3.078	-2.422	0.146	-2.702	-3.032	-2.336	0.159
Item19	0.573	0.822	0.660	0.950	0.060	0.826	0.695	0.926	0.054
Item20	0.467	0.687	0.531	0.880	0.089	0.683	0.514	0.848	0.075
Item21	-0.023	-0.296	-0.488	-0.039	0.102	-0.303	-0.560	-0.113	0.077
Item22	0.618	0.943	0.801	1.206	0.079	0.946	0.706	1.130	0.095
Item23	0.123	-0.315	-0.510	0.018	0.112	-0.322	-0.501	-0.110	0.080
Item24	0.521	0.576	0.394	0.936	0.105	0.581	0.399	0.854	0.098
Item25	-0.376	-0.851	-1.052	-0.626	0.114	-0.918	-1.128	-0.644	0.098
Item26	-1.454	-3.005	-3.524	-2.586	0.190	-2.969	-3.313	-2.534	0.164
Item27	0.553	0.832	0.539	1.149	0.101	0.892	0.647	1.187	0.122
Item28	-0.204	-0.519	-0.720	-0.309	0.100	-0.538	-0.722	-0.307	0.109
Item29	0.236	0.083	-0.065	0.253	0.069	0.056	-0.116	0.223	0.071
Item30	0.017	-0.966	-1.213	-0.607	0.137	-1.036	-1.379	-0.627	0.155
Item31	0.853	1.418	1.266	1.667	0.075	1.453	1.284	1.635	0.070
Item32	1.037	1.990	1.779	2.195	0.087	1.962	1.728	2.189	0.097

Table A5: Item Discrimination Estimates from Different Sampling Methods

for 3-pl Model

		Random				Leverage			
Item	aFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item1	3.251	3.529	2.755	4.527	0.390	3.310	2.671	4.348	0.391
Item2	2.220	1.617	1.054	2.359	0.310	1.555	1.130	2.608	0.305
Item3	2.271	2.537	1.921	3.451	0.337	2.552	1.909	3.195	0.325
Item4	2.295	2.751	2.211	3.579	0.341	2.984	2.142	4.144	0.405
Item5	1.566	1.812	1.319	2.623	0.275	1.683	0.748	2.389	0.298
Item6	3.597	2.682	1.737	4.376	0.494	2.651	1.974	3.651	0.466
Item7	1.680	1.769	1.204	2.550	0.290	1.727	1.185	2.507	0.260
Item8	2.455	2.983	2.443	4.072	0.372	3.077	2.514	3.945	0.335
Item9	3.318	3.105	2.457	3.889	0.373	3.037	2.434	4.153	0.390
Item10	2.596	2.495	1.637	3.823	0.452	2.660	2.070	4.345	0.403
Item11	4.074	4.701	3.886	5.769	0.404	4.929	4.068	5.939	0.438
Item12	1.703	1.757	1.106	2.353	0.287	1.720	1.256	2.577	0.277
Item13	5.235	4.820	3.824	5.621	0.411	4.184	3.303	4.960	0.423
Item14	4.427	3.600	2.437	4.776	0.521	4.077	2.695	5.271	0.501
Item15	3.834	4.313	3.556	5.685	0.448	4.123	3.215	5.165	0.404
Item16	1.299	1.105	0.511	1.630	0.252	0.937	0.504	1.419	0.244

Table A5 Continued: Item Discrimination Estimates from Different Sampling Methods for 3-pl Model

		SLEV				Adj-SLEV			
Item	aFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item1	3.251	3.276	2.782	4.159	0.379	3.421	2.670	4.661	0.406
Item2	2.220	1.662	1.038	2.658	0.334	1.655	1.073	2.944	0.323
Item3	2.271	2.478	1.839	3.345	0.318	2.440	1.889	3.319	0.337
Item4	2.295	2.970	2.404	4.014	0.362	2.943	2.182	4.345	0.354
Item5	1.566	1.661	0.928	2.529	0.287	1.739	1.108	2.269	0.258
Item6	3.597	2.698	1.710	3.981	0.574	2.536	1.501	3.869	0.530
Item7	1.680	1.768	1.320	2.380	0.267	1.726	1.208	2.282	0.286
Item8	2.455	3.095	2.548	3.782	0.298	3.073	2.516	3.960	0.289
Item9	3.318	2.994	2.281	4.124	0.429	3.188	2.341	4.159	0.467
Item10	2.596	2.658	1.633	3.351	0.364	2.564	1.836	3.626	0.407
Item11	4.074	4.862	3.893	6.037	0.455	4.873	4.075	5.638	0.424
Item12	1.703	1.747	0.857	2.430	0.267	1.808	1.308	2.412	0.237
Item13	5.235	4.164	3.173	4.898	0.375	4.338	3.432	5.207	0.437
Item14	4.427	3.987	2.843	5.270	0.572	3.947	2.546	5.078	0.599
Item15	3.834	4.343	3.607	5.113	0.363	4.032	3.355	4.824	0.379
Item16	1.299	0.927	0.458	1.376	0.220	0.980	0.279	2.234	0.321

Table A5 Continued: Item Discrimination Estimates from Different Sampling Methods for 3-pl Model

		Random				Leverage			
Item	aFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item17	2.439	1.974	1.338	2.932	0.359	2.059	1.422	2.794	0.330
Item18	0.325	0.411	-0.024	0.795	0.191	0.062	-0.340	0.698	0.241
Item19	3.887	4.509	3.553	5.537	0.454	4.505	3.774	5.372	0.383
Item20	2.312	2.598	1.901	3.728	0.372	2.695	2.120	3.268	0.284
Item21	2.057	2.250	1.647	3.106	0.330	2.319	1.683	3.122	0.327
Item22	2.428	2.704	1.999	3.506	0.374	2.833	2.033	3.747	0.348
Item23	2.879	2.661	2.022	3.304	0.304	2.930	2.111	4.106	0.420
Item24	2.914	2.578	1.744	3.523	0.409	2.973	2.116	4.195	0.472
Item25	1.740	1.954	1.372	2.708	0.331	2.100	1.233	2.950	0.356
Item26	0.821	1.094	0.672	1.742	0.260	0.949	0.485	1.556	0.254
Item27	1.614	1.895	1.108	3.069	0.320	1.720	1.041	2.323	0.299
Item28	1.650	1.836	1.170	2.502	0.322	1.963	1.458	2.805	0.318
Item29	3.828	3.853	3.179	4.505	0.325	4.202	3.099	5.267	0.490
Item30	4.026	2.848	1.649	3.718	0.397	2.992	2.105	4.031	0.436
Item31	3.996	4.470	3.634	5.820	0.517	4.188	3.315	5.024	0.415
Item32	5.575	5.100	4.418	6.047	0.415	4.615	3.935	5.661	0.368

Table A5 Continued: Item Discrimination Estimates from Different Sampling Methods for 3-pl Model

		SLEV				Adj-SLEV			
Item	aFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item17	2.439	1.967	1.303	3.456	0.349	1.916	1.270	2.879	0.380
Item18	0.325	0.096	-0.317	0.449	0.182	0.185	-0.267	0.489	0.168
Item19	3.887	4.459	3.658	5.461	0.449	4.476	3.606	5.241	0.399
Item20	2.312	2.660	1.975	3.215	0.267	2.691	1.968	3.451	0.325
Item21	2.057	2.299	1.719	3.152	0.316	2.236	1.600	3.021	0.317
Item22	2.428	2.762	2.151	3.917	0.322	2.711	2.095	3.326	0.301
Item23	2.879	3.089	2.229	5.009	0.521	2.967	2.099	4.038	0.419
Item24	2.914	2.696	1.817	4.449	0.448	2.837	2.114	3.937	0.424
Item25	1.740	2.146	1.582	3.002	0.327	2.072	1.470	2.733	0.316
Item26	0.821	0.944	0.526	1.683	0.257	0.992	0.584	1.460	0.205
Item27	1.614	1.806	1.239	3.050	0.290	1.750	1.225	2.247	0.248
Item28	1.650	1.925	1.347	2.594	0.308	1.897	1.316	2.653	0.284
Item29	3.828	4.237	3.280	5.232	0.411	4.14	3.433	5.322	0.418
Item30	4.026	3.042	2.362	3.927	0.407	3.030	2.240	4.338	0.419
Item31	3.996	4.209	3.348	5.236	0.464	4.228	3.222	5.053	0.429
Item32	5.575	4.685	3.873	5.861	0.404	4.922	3.963	6.013	0.404

Table A6: Item Guessing Estimates from Different Sampling Methods

for 3-pl Model

		Random				Leverage			
Item	cFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item1	0.011	-0.044	-0.072	-0.017	0.014	-0.035	-0.065	-0.010	0.011
Item2	0.434	0.323	0.267	0.394	0.030	0.351	0.270	0.481	0.042
Item3	0.033	0.038	0.001	0.083	0.021	0.031	0.004	0.064	0.014
Item4	0.089	0.140	0.060	0.204	0.030	0.120	0.081	0.197	0.026
Item5	0.066	0.150	0.087	0.222	0.029	0.132	0.087	0.189	0.026
Item6	0.305	0.313	0.157	0.432	0.056	0.311	0.203	0.407	0.046
Item7	0.088	0.128	0.067	0.198	0.030	0.108	0.062	0.163	0.022
Item8	0.000	0.048	0.006	0.096	0.019	0.019	-0.016	0.050	0.014
Item9	0.180	0.172	0.103	0.267	0.040	0.162	0.076	0.263	0.045
Item10	0.212	0.212	0.141	0.357	0.043	0.211	0.135	0.282	0.035
Item11	0.010	-0.027	-0.054	0.006	0.015	-0.027	-0.050	0.000	0.013
Item12	0.240	0.242	0.158	0.311	0.030	0.214	0.155	0.303	0.030
Item13	0.000	-0.094	-0.116	-0.067	0.009	-0.074	-0.099	-0.054	0.009
Item14	0.359	0.333	0.217	0.481	0.058	0.386	0.277	0.488	0.047
Item15	0.006	-0.043	-0.071	-0.004	0.014	-0.038	-0.066	-0.012	0.010
Item16	0.252	0.270	0.207	0.364	0.030	0.251	0.177	0.306	0.027

Table A6 Continued: Item Guessing Estimates from Different Sampling Methods

for 3-pl Model

		SLEV				Adj-SLEV			
Item	cFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item1	0.011	-0.035	-0.055	-0.017	0.010	-0.037	-0.057	-0.004	0.012
Item2	0.434	0.355	0.276	0.437	0.041	0.338	0.253	0.424	0.039
Item3	0.033	0.029	-0.005	0.077	0.017	0.034	-0.002	0.075	0.017
Item4	0.089	0.125	0.073	0.182	0.024	0.123	0.064	0.222	0.031
Item5	0.066	0.134	0.096	0.200	0.023	0.133	0.087	0.196	0.028
Item6	0.305	0.309	0.199	0.402	0.050	0.303	0.208	0.408	0.051
Item7	0.088	0.109	0.050	0.174	0.025	0.106	0.045	0.224	0.030
Item8	0.000	0.022	0.001	0.046	0.010	0.029	-0.004	0.066	0.017
Item9	0.180	0.173	0.102	0.291	0.044	0.175	0.087	0.282	0.044
Item10	0.212	0.220	0.121	0.415	0.047	0.204	0.131	0.306	0.044
Item11	0.010	-0.020	-0.052	0.010	0.011	-0.025	-0.049	0.008	0.015
Item12	0.240	0.217	0.136	0.292	0.031	0.222	0.180	0.282	0.025
Item13	0.000	-0.075	-0.090	-0.058	0.008	-0.082	-0.099	-0.061	0.009
Item14	0.359	0.372	0.245	0.501	0.060	0.370	0.259	0.550	0.060
Item15	0.006	-0.039	-0.059	-0.016	0.010	-0.037	-0.069	-0.011	0.013
Item16	0.252	0.259	0.205	0.314	0.026	0.270	0.185	0.374	0.038

Table A6 Continued: Item Guessing Estimates from Different Sampling Methods

for 3-pl Model

		Random				Leverage			
Item	cFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item17	0.341	0.317	0.237	0.481	0.045	0.323	0.237	0.413	0.041
Item18	0.000	0.558	0.471	0.681	0.052	0.521	0.432	0.616	0.042
Item19	0.007	-0.039	-0.061	-0.014	0.011	-0.034	-0.053	-0.006	0.01
Item20	0.000	0.011	-0.016	0.042	0.014	-0.002	-0.035	0.038	0.014
Item21	0.129	0.185	0.129	0.269	0.030	0.171	0.101	0.247	0.033
Item22	0.020	0.015	-0.019	0.063	0.018	0.005	-0.026	0.034	0.014
Item23	0.261	0.237	0.116	0.38	0.052	0.247	0.116	0.348	0.047
Item24	0.154	0.133	0.062	0.240	0.041	0.162	0.044	0.241	0.036
Item25	0.218	0.255	0.201	0.329	0.027	0.247	0.167	0.331	0.036
Item26	0.656	0.443	0.365	0.540	0.036	0.434	0.376	0.517	0.029
Item27	0.043	0.096	0.056	0.173	0.026	0.073	0.030	0.111	0.020
Item28	0.142	0.197	0.129	0.278	0.034	0.172	0.126	0.254	0.026
Item29	0.133	0.116	0.035	0.190	0.029	0.150	0.079	0.227	0.034
Item30	0.490	0.356	0.252	0.458	0.051	0.416	0.319	0.538	0.049
Item31	0.002	-0.064	-0.088	-0.035	0.011	-0.054	-0.079	-0.035	0.009
Item32	0.000	-0.094	-0.110	-0.067	0.010	-0.073	-0.099	-0.053	0.009

Table A6 Continued: Item Guessing Estimates from Different Sampling Methods for 3-pl Model

		SLEV				Adj-SLEV			
Item	cFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item17	0.341	0.308	0.210	0.484	0.050	0.311	0.210	0.441	0.052
Item18	0.000	0.523	0.412	0.630	0.040	0.540	0.449	0.691	0.051
Item19	0.007	-0.033	-0.058	-0.007	0.011	-0.036	-0.055	-0.011	0.011
Item20	0.000	-0.003	-0.034	0.025	0.011	0.001	-0.017	0.022	0.010
Item21	0.129	0.163	0.111	0.217	0.027	0.170	0.116	0.241	0.030
Item22	0.020	0.005	-0.024	0.043	0.014	0.011	-0.030	0.042	0.016
Item23	0.261	0.246	0.162	0.362	0.047	0.255	0.144	0.401	0.050
Item24	0.154	0.149	0.081	0.286	0.036	0.154	0.081	0.259	0.041
Item25	0.218	0.253	0.179	0.390	0.039	0.244	0.193	0.342	0.033
Item26	0.656	0.434	0.357	0.495	0.030	0.441	0.371	0.495	0.026
Item27	0.043	0.073	0.033	0.109	0.019	0.079	0.019	0.126	0.024
Item28	0.142	0.167	0.079	0.223	0.032	0.186	0.111	0.258	0.034
Item29	0.133	0.151	0.065	0.238	0.041	0.150	0.082	0.236	0.034
Item30	0.490	0.420	0.304	0.525	0.052	0.388	0.282	0.531	0.056
Item31	0.002	-0.056	-0.077	-0.038	0.009	-0.057	-0.079	-0.037	0.010
Item32	0.000	-0.075	-0.093	-0.059	0.008	-0.081	-0.098	-0.059	0.008

APPENDIX B

ITEM PARAMETER ESTIMATES FROM NON-NORMAL DATA

(EMPIRICAL STUDY 2)

Table B1 : Item Difficulty Estimates from Different Sampling Methods

for Rasch Model

		Random				Leverage			
Item	bFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item1	3.458	3.442	3.144	3.740	0.126	3.555	3.257	3.938	0.140
Item2	-0.030	-0.037	-0.257	0.153	0.088	-0.024	-0.180	0.148	0.075
Item3	2.379	2.403	2.124	2.752	0.111	2.464	2.338	2.644	0.074
Item4	1.135	1.170	0.891	1.365	0.097	1.185	1.012	1.353	0.075
Item5	1.382	1.398	1.168	1.709	0.113	1.464	1.228	1.758	0.117
Item6	1.009	1.064	0.874	1.337	0.104	1.202	1.024	1.390	0.094
Item7	2.569	2.610	2.392	2.878	0.118	2.587	2.346	2.772	0.093
Item8	1.930	1.980	1.770	2.227	0.103	1.824	1.680	1.997	0.079
Item9	0.873	0.916	0.722	1.114	0.087	1.010	0.745	1.227	0.113
Item10	1.138	1.153	0.950	1.324	0.075	1.192	0.982	1.379	0.097
Item11	2.549	2.576	2.274	2.902	0.135	2.506	2.296	2.717	0.082
Item12	0.723	0.738	0.571	0.909	0.069	0.646	0.499	0.845	0.078
Item13	4.452	4.258	3.940	4.615	0.163	4.226	4.025	4.516	0.112
Item14	0.492	0.516	0.330	0.690	0.084	0.505	0.281	0.693	0.098
Item15	3.409	3.394	3.020	3.765	0.148	3.295	3.138	3.522	0.091
Item16	0.565	0.563	0.400	0.719	0.073	0.636	0.440	0.850	0.087

Table B1 Continued: Item Difficulty Estimates from Different Sampling Methods for Rasch Model

		SLEV				Adj-SLEV			
Item	bFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item1	3.458	3.534	3.300	3.758	0.106	3.551	3.356	3.808	0.112
Item2	-0.030	-0.032	-0.237	0.171	0.104	-0.029	-0.207	0.174	0.093
Item3	2.379	2.491	2.258	2.757	0.115	2.439	2.269	2.683	0.109
Item4	1.135	1.178	0.966	1.396	0.102	1.214	0.992	1.404	0.091
Item5	1.382	1.465	1.256	1.639	0.085	1.466	1.286	1.776	0.092
Item6	1.009	1.185	0.905	1.399	0.095	1.148	0.941	1.374	0.103
Item7	2.569	2.582	2.425	2.736	0.079	2.575	2.328	2.830	0.117
Item8	1.930	1.841	1.660	2.060	0.089	1.867	1.709	2.056	0.081
Item9	0.873	0.993	0.782	1.181	0.087	0.957	0.689	1.176	0.092
Item10	1.138	1.172	0.951	1.381	0.101	1.212	1.058	1.427	0.086
Item11	2.549	2.520	2.241	2.829	0.108	2.536	2.274	2.793	0.125
Item12	0.723	0.689	0.485	0.872	0.090	0.701	0.478	0.993	0.102
Item13	4.452	4.216	3.971	4.387	0.100	4.194	3.855	4.484	0.123
Item14	0.492	0.500	0.252	0.686	0.096	0.470	0.260	0.697	0.099
Item15	3.409	3.299	3.113	3.561	0.081	3.312	3.108	3.630	0.126
Item16	0.565	0.634	0.452	0.954	0.101	0.614	0.448	0.898	0.090

Table B1 Continued: Item Difficulty Estimates from Different Sampling Methods for Rasch Model

		Random				Leverage			
Item	bFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item17	-0.122	-0.135	-0.299	0.019	0.078	-0.153	-0.351	0.124	0.099
Item18	-1.886	-1.983	-2.176	-1.825	0.079	-2.030	-2.234	-1.758	0.097
Item19	3.317	3.302	3.109	3.558	0.109	3.163	2.978	3.355	0.090
Item20	2.236	2.288	2.046	2.504	0.110	2.382	2.145	2.656	0.121
Item21	0.543	0.563	0.390	0.784	0.085	0.592	0.410	0.733	0.078
Item22	1.925	1.978	1.757	2.278	0.110	1.969	1.711	2.229	0.096
Item23	0.554	0.613	0.422	0.754	0.076	0.533	0.352	0.671	0.068
Item24	1.447	1.472	1.179	1.720	0.097	1.508	1.332	1.904	0.107
Item25	0.466	0.481	0.347	0.688	0.072	0.500	0.244	0.712	0.102
Item26	-1.541	-1.647	-1.837	-1.460	0.098	-1.685	-1.936	-1.422	0.112
Item27	1.753	1.763	1.531	1.944	0.092	1.878	1.660	2.061	0.093
Item28	0.124	0.117	-0.045	0.325	0.080	0.125	-0.124	0.437	0.109
Item29	1.220	1.242	1.020	1.400	0.085	1.245	1.060	1.490	0.085
Item30	-0.685	-0.720	-0.957	-0.518	0.091	-0.802	-1.048	-0.616	0.104
Item31	2.731	2.758	2.492	3.005	0.119	2.783	2.581	2.962	0.080
Item32	4.251	4.129	3.759	4.588	0.179	4.083	3.867	4.291	0.097

Table B1 Continued: Item Difficulty Estimates from Different Sampling Methods for Rasch Model

		SLEV				Adj-SLEV			
Item	bFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item17	-0.122	-0.152	-0.372	0.030	0.088	-0.128	-0.373	0.062	0.112
Item18	-1.886	-2.052	-2.228	-1.862	0.094	-2.054	-2.236	-1.769	0.100
Item19	3.317	3.175	2.952	3.392	0.097	3.198	3.027	3.452	0.092
Item20	2.236	2.381	2.107	2.588	0.112	2.367	2.165	2.687	0.114
Item21	0.543	0.598	0.419	0.815	0.085	0.600	0.387	0.814	0.083
Item22	1.925	1.974	1.755	2.171	0.084	2.008	1.842	2.293	0.094
Item23	0.554	0.556	0.344	0.726	0.090	0.557	0.305	0.712	0.083
Item24	1.447	1.506	1.343	1.710	0.095	1.516	1.345	1.759	0.097
Item25	0.466	0.522	0.280	0.670	0.091	0.494	0.342	0.668	0.073
Item26	-1.541	-1.668	-1.818	-1.530	0.069	-1.714	-1.857	-1.515	0.076
Item27	1.753	1.873	1.655	2.121	0.098	1.870	1.624	2.144	0.095
Item28	0.124	0.116	-0.115	0.383	0.110	0.134	-0.080	0.293	0.09
Item29	1.220	1.262	1.023	1.416	0.085	1.260	1.021	1.429	0.095
Item30	-0.685	-0.833	-0.988	-0.673	0.078	-0.787	-0.982	-0.618	0.092
Item31	2.731	2.773	2.586	3.053	0.096	2.785	2.580	3.070	0.112
Item32	4.251	4.068	3.862	4.321	0.113	4.031	3.840	4.265	0.098

Table B2 Continued: Item Difficulty Estimates from Different Sampling Methods for 2-pl Model

		Random				Leverage			
Item	bFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item1	2.814	2.792	2.307	3.352	0.203	2.911	2.698	3.357	0.127
Item2	0.013	-0.084	-0.363	0.132	0.122	-0.059	-0.236	0.086	0.079
Item3	2.091	2.088	1.757	2.764	0.192	2.129	1.928	2.379	0.095
Item4	0.883	0.866	0.630	1.166	0.110	0.932	0.804	1.097	0.065
Item5	1.379	1.363	1.010	1.736	0.146	1.357	1.089	1.654	0.118
Item6	1.445	1.492	1.164	2.086	0.211	1.388	1.118	1.763	0.130
Item7	2.110	2.167	1.832	2.689	0.192	2.208	2.002	2.586	0.118
Item8	1.360	1.363	1.172	1.561	0.084	1.359	1.215	1.528	0.065
Item9	1.484	1.527	0.865	1.981	0.263	1.280	0.971	1.656	0.143
Item10	1.146	1.126	0.900	1.478	0.141	1.069	0.895	1.277	0.089
Item11	1.811	1.820	1.528	2.161	0.120	1.916	1.702	2.098	0.081
Item12	0.523	0.480	0.333	0.597	0.065	0.466	0.358	0.606	0.060
Item13	2.717	2.852	2.572	3.138	0.149	2.987	2.807	3.220	0.103
Item14	0.597	0.578	0.356	0.985	0.125	0.479	0.254	0.696	0.100
Item15	2.245	2.274	1.987	2.505	0.116	2.350	2.203	2.492	0.062
Item16	0.616	0.557	0.346	0.830	0.099	0.596	0.355	0.782	0.096

Table B2 Continued: Item Difficulty Estimates from Different Sampling Methods

for 2-pl Model

		SLEV				Adj-SLEV			
Item	bFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item1	2.814	2.908	2.610	3.231	0.136	2.897	2.606	3.202	0.134
Item2	0.013	-0.055	-0.323	0.192	0.124	-0.058	-0.258	0.163	0.109
Item3	2.091	2.158	1.900	2.485	0.121	2.112	1.919	2.429	0.118
Item4	0.883	0.926	0.773	1.095	0.077	0.950	0.776	1.091	0.075
Item5	1.379	1.386	1.178	1.617	0.106	1.385	1.165	1.654	0.098
Item6	1.445	1.390	1.060	1.662	0.130	1.394	1.119	1.740	0.144
Item7	2.110	2.155	1.932	2.562	0.119	2.198	1.870	2.492	0.135
Item8	1.360	1.348	1.228	1.505	0.078	1.352	1.214	1.483	0.063
Item9	1.484	1.280	0.999	1.702	0.137	1.324	1.064	1.699	0.148
Item10	1.146	1.051	0.814	1.241	0.101	1.087	0.913	1.307	0.087
Item11	1.811	1.907	1.744	2.162	0.089	1.904	1.685	2.179	0.103
Item12	0.523	0.500	0.284	0.664	0.076	0.494	0.344	0.703	0.071
Item13	2.717	2.980	2.808	3.147	0.087	2.936	2.621	3.197	0.120
Item14	0.597	0.491	0.256	0.655	0.093	0.467	0.255	0.723	0.110
Item15	2.245	2.348	2.140	2.548	0.077	2.345	2.196	2.626	0.113
Item16	0.616	0.612	0.404	0.926	0.102	0.589	0.435	0.856	0.092

Table B2 Continued: Item Difficulty Estimates from Different Sampling Methods for 2-pl Model

		Random				Leverage			
Item	bFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item17	-0.102	-0.216	-0.508	0.020	0.118	-0.215	-0.468	0.081	0.117
Item18	-3.526	-2.979	-3.357	-2.674	0.184	-2.955	-3.249	-2.440	0.177
Item19	2.080	2.145	1.91	2.549	0.124	2.258	2.106	2.389	0.068
Item20	1.900	1.908	1.442	2.357	0.189	2.058	1.858	2.281	0.108
Item21	0.574	0.530	0.251	0.758	0.113	0.554	0.359	0.691	0.082
Item22	1.510	1.494	1.219	1.861	0.124	1.588	1.313	1.767	0.084
Item23	0.651	0.672	0.444	0.919	0.110	0.509	0.344	0.623	0.072
Item24	2.319	2.268	1.739	3.091	0.256	1.814	1.544	2.054	0.122
Item25	0.407	0.353	0.218	0.531	0.069	0.410	0.182	0.579	0.087
Item26	-1.419	-1.675	-2.194	-1.362	0.169	-1.616	-1.803	-1.217	0.125
Item27	1.600	1.571	1.251	1.871	0.126	1.691	1.450	1.916	0.102
Item28	0.200	0.120	-0.096	0.472	0.113	0.098	-0.248	0.447	0.128
Item29	1.721	1.734	1.287	2.149	0.196	1.376	1.172	1.571	0.108
Item30	-0.751	-0.925	-1.199	-0.656	0.136	-0.831	-1.095	-0.608	0.116
Item31	2.237	2.239	2.002	2.668	0.173	2.275	2.071	2.434	0.087
Item32	2.635	2.769	2.475	3.292	0.190	2.887	2.676	3.048	0.076

Table B2 Continued: Item Difficulty Estimates from Different Sampling Methods

for 2-pl Model

		SLEV				Adj-SLEV			
Item	bFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item17	-0.102	-0.200	-0.502	0.067	0.115	-0.185	-0.436	0.051	0.130
Item18	-3.526	-3.011	-3.376	-2.711	0.166	-3.042	-3.494	-2.660	0.181
Item19	2.080	2.263	2.069	2.443	0.085	2.231	2.045	2.379	0.082
Item20	1.900	2.051	1.732	2.308	0.112	2.044	1.821	2.319	0.120
Item21	0.574	0.570	0.400	0.819	0.095	0.584	0.350	0.782	0.082
Item22	1.510	1.563	1.411	1.719	0.075	1.591	1.429	1.845	0.086
Item23	0.651	0.545	0.349	0.749	0.088	0.553	0.273	0.754	0.091
Item24	2.319	1.781	1.525	2.055	0.140	1.828	1.616	2.178	0.121
Item25	0.407	0.435	0.221	0.587	0.086	0.400	0.245	0.562	0.069
Item26	-1.419	-1.608	-1.910	-1.385	0.102	-1.631	-1.900	-1.409	0.115
Item27	1.600	1.695	1.479	1.859	0.073	1.724	1.463	1.927	0.103
Item28	0.200	0.106	-0.186	0.381	0.117	0.129	-0.099	0.345	0.110
Item29	1.721	1.425	1.166	1.633	0.131	1.454	1.081	1.838	0.142
Item30	-0.751	-0.847	-1.073	-0.694	0.094	-0.832	-1.059	-0.546	0.117
Item31	2.237	2.247	2.057	2.555	0.101	2.240	2.001	2.496	0.117
Item32	2.635	2.866	2.703	3.105	0.092	2.804	2.657	3.063	0.099

Table B3: Item Discrimination Estimates from Different Sampling Methods for 2-pl Model

		Random				Leverage			
Item	aFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item1	1.315	1.409	1.067	1.833	0.166	1.822	1.100	1.822	0.174
Item2	0.773	0.715	0.493	0.959	0.097	1.040	0.673	1.040	0.092
Item3	1.198	1.245	1.001	1.551	0.132	1.440	1.023	1.440	0.116
Item4	1.538	1.526	1.224	2.004	0.170	1.919	1.32	1.919	0.117
Item5	1.028	1.031	0.749	1.246	0.101	1.251	0.772	1.251	0.097
Item6	0.660	0.658	0.503	0.822	0.076	0.837	0.455	0.837	0.086
Item7	1.328	1.349	1.016	1.782	0.18	1.496	0.958	1.496	0.128
Item8	1.821	1.839	1.524	2.240	0.155	2.211	1.537	2.211	0.143
Item9	0.547	0.548	0.405	0.776	0.085	0.744	0.384	0.744	0.080
Item10	1.027	1.022	0.772	1.230	0.108	1.329	0.879	1.329	0.091
Item11	1.748	1.788	1.456	2.213	0.182	2.080	1.450	2.080	0.142
Item12	1.787	1.730	1.451	2.103	0.132	2.471	1.805	2.471	0.129
Item13	2.282	2.146	1.639	2.623	0.222	2.375	1.752	2.375	0.151
Item14	0.856	0.831	0.566	1.053	0.113	1.231	0.701	1.231	0.113
Item15	2.005	2.038	1.578	2.519	0.212	2.470	1.787	2.470	0.149
Item16	0.968	0.950	0.764	1.196	0.097	1.197	0.737	1.197	0.112

Table B3 Continued: Item Discrimination Estimates from Different Sampling Methods for 2-pl Model

		SLEV				Adj-SLEV			
Item	Afull	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item1	1.315	1.366	0.946	1.806	0.166	1.396	1.105	1.690	0.151
Item2	0.773	0.830	0.620	1.074	0.105	0.818	0.645	0.990	0.076
Item3	1.198	1.212	0.873	1.488	0.141	1.220	0.928	1.500	0.120
Item4	1.538	1.560	1.328	1.873	0.130	1.510	1.173	1.784	0.130
Item5	1.028	0.996	0.714	1.250	0.100	1.016	0.855	1.204	0.084
Item6	0.660	0.635	0.471	0.849	0.088	0.645	0.458	0.827	0.092
Item7	1.328	1.323	0.938	1.732	0.142	1.257	1.027	1.512	0.121
Item8	1.821	1.970	1.616	2.261	0.135	1.903	1.596	2.233	0.147
Item9	0.547	0.529	0.316	0.768	0.082	0.510	0.320	0.683	0.087
Item10	1.027	1.115	0.899	1.402	0.105	1.117	0.935	1.319	0.096
Item11	1.748	1.738	1.388	2.163	0.150	1.730	1.371	2.106	0.166
Item12	1.787	1.966	1.617	2.279	0.140	1.904	1.595	2.265	0.154
Item13	2.282	2.078	1.708	2.398	0.165	2.125	1.696	2.502	0.172
Item14	0.856	0.893	0.688	1.128	0.099	0.883	0.740	1.014	0.072
Item15	2.005	2.078	1.794	2.473	0.174	2.046	1.750	2.528	0.189
Item16	0.968	0.930	0.792	1.127	0.073	0.955	0.789	1.168	0.084

Table B3 Continued: Item Discrimination Estimates from Different Sampling Methods for 2-pl Model

		Random				Leverage			
Item	aFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item17	0.754	0.733	0.553	0.956	0.096	0.857	0.558	0.857	0.066
Item18	0.477	0.632	0.474	0.791	0.070	0.647	0.345	0.647	0.067
Item19	2.308	2.198	1.800	2.54	0.174	2.367	1.817	2.367	0.112
Item20	1.270	1.330	1.075	1.779	0.164	1.502	0.98	1.502	0.103
Item21	1.013	1.012	0.673	1.314	0.127	1.171	0.798	1.171	0.078
Item22	1.471	1.543	1.245	1.917	0.150	1.696	1.222	1.696	0.106
Item23	0.881	0.851	0.697	1.007	0.084	1.136	0.707	1.136	0.095
Item24	0.568	0.590	0.470	0.854	0.081	0.800	0.442	0.800	0.082
Item25	1.337	1.319	1.082	1.631	0.145	1.581	1.133	1.581	0.113
Item26	1.134	1.126	0.789	1.506	0.175	1.407	0.844	1.407	0.124
Item27	1.153	1.185	0.961	1.529	0.135	1.259	0.899	1.259	0.094
Item28	0.778	0.746	0.499	1.027	0.119	0.921	0.552	0.921	0.092
Item29	0.664	0.664	0.508	0.865	0.082	0.921	0.542	0.921	0.095
Item30	0.839	0.815	0.604	1.089	0.101	1.243	0.792	1.243	0.105
Item31	1.328	1.394	1.163	1.729	0.148	1.769	1.151	1.769	0.143
Item32	2.222	2.114	1.696	2.572	0.204	2.491	1.807	2.491	0.161

Table B3 Continued: Item Discrimination Estimates from Different Sampling Methods for 2-pl Model

		SLEV				Adj- SLEV			
Item	aFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item17	0.754	0.713	0.469	0.994	0.103	0.700	0.535	0.927	0.094
Item18	0.477	0.476	0.290	0.683	0.069	0.518	0.375	0.677	0.065
Item19	2.308	2.083	1.769	2.330	0.146	2.168	1.846	2.540	0.166
Item20	1.270	1.225	0.993	1.572	0.115	1.223	1.000	1.509	0.107
Item21	1.013	0.956	0.727	1.177	0.109	0.927	0.765	1.145	0.087
Item22	1.471	1.529	1.193	1.799	0.120	1.492	1.274	1.803	0.137
Item23	0.881	0.887	0.727	1.121	0.092	0.892	0.617	1.106	0.104
Item24	0.568	0.631	0.463	0.873	0.097	0.654	0.445	0.899	0.091
Item25	1.337	1.316	1.086	1.540	0.117	1.329	1.009	1.568	0.115
Item26	1.134	1.039	0.645	1.399	0.141	1.097	0.855	1.376	0.119
Item27	1.153	1.095	0.889	1.347	0.093	1.071	0.891	1.243	0.097
Item28	0.778	0.744	0.597	0.925	0.089	0.726	0.522	0.929	0.081
Item29	0.664	0.690	0.507	0.961	0.098	0.704	0.545	0.834	0.081
Item30	0.839	0.937	0.706	1.355	0.115	0.927	0.750	1.237	0.114
Item31	1.328	1.432	1.098	1.962	0.170	1.442	1.165	1.765	0.142
Item32	2.222	2.104	1.706	2.473	0.161	2.169	1.650	2.542	0.197

Table B4: Item Difficulty Estimates from Different Sampling Methods

for 3-pl Model

		Random				Leverage			
Item	bFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item1	0.976	1.476	1.175	1.689	0.109	1.825	1.411	1.825	0.072
Item2	0.587	0.107	-0.117	0.413	0.110	0.251	-0.089	0.251	0.085
Item3	0.867	1.076	0.878	1.426	0.114	1.213	0.980	1.213	0.056
Item4	0.574	0.335	0.183	0.538	0.078	0.492	0.274	0.492	0.044
Item5	0.725	0.702	0.446	0.980	0.101	0.843	0.495	0.843	0.080
Item6	0.816	0.896	0.648	1.264	0.139	1.074	0.609	1.074	0.097
Item7	0.843	1.085	0.919	1.350	0.104	1.342	0.991	1.342	0.070
Item8	0.698	0.644	0.521	0.760	0.053	0.741	0.532	0.741	0.045
Item9	0.803	0.916	0.498	1.231	0.156	1.010	0.678	1.010	0.079
Item10	0.686	0.566	0.426	0.807	0.091	0.679	0.450	0.679	0.057
Item11	0.809	0.924	0.731	1.152	0.076	1.070	0.830	1.070	0.052
Item12	0.489	0.092	-0.004	0.214	0.051	0.145	-0.044	0.145	0.043
Item13	1.002	1.599	1.416	1.853	0.105	1.789	1.508	1.789	0.062
Item14	0.634	0.356	0.138	0.601	0.094	0.475	0.151	0.475	0.078
Item15	0.900	1.210	1.024	1.356	0.071	1.316	1.128	1.316	0.041
Item16	0.550	0.244	0.125	0.456	0.079	0.403	0.064	0.403	0.073

Table B4: Item Difficulty Estimates from Different Sampling Methods

for 3-pl Model

		SLEV				Adj- SLEV			
Item	bFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item1	0.976	1.541	1.385	1.713	0.072	1.514	1.382	1.665	0.066
Item2	0.587	0.063	-0.133	0.267	0.098	0.074	-0.189	0.313	0.103
Item3	0.867	1.121	0.988	1.335	0.071	1.092	0.953	1.251	0.071
Item4	0.574	0.358	0.276	0.462	0.049	0.380	0.279	0.482	0.049
Item5	0.725	0.707	0.552	0.910	0.076	0.707	0.552	0.887	0.068
Item6	0.816	0.824	0.584	1.068	0.099	0.842	0.659	1.094	0.107
Item7	0.843	1.096	0.967	1.340	0.071	1.113	0.928	1.287	0.073
Item8	0.698	0.613	0.532	0.744	0.052	0.620	0.520	0.697	0.042
Item9	0.803	0.826	0.614	1.115	0.093	0.866	0.716	1.069	0.089
Item10	0.686	0.524	0.380	0.690	0.070	0.548	0.409	0.697	0.065
Item11	0.809	0.959	0.854	1.124	0.057	0.955	0.820	1.116	0.064
Item12	0.489	0.096	-0.062	0.211	0.054	0.092	-0.036	0.192	0.053
Item13	1.002	1.625	1.488	1.732	0.054	1.592	1.402	1.758	0.072
Item14	0.634	0.312	0.144	0.477	0.081	0.303	0.122	0.512	0.083
Item15	0.900	1.225	1.097	1.335	0.046	1.214	1.121	1.377	0.065
Item16	0.550	0.254	0.105	0.445	0.070	0.237	0.095	0.430	0.067

Table B4 Continued: Item Difficulty Estimates from Different Sampling Methods for 3-pl Model

		Random				Leverage			
Item	bFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item17	0.507	-0.060	-0.233	0.109	0.071	0.118	-0.299	0.118	0.093
Item18	-3.137	-1.912	-2.067	-1.741	0.080	-1.583	-2.024	-1.583	0.098
Item19	0.860	1.130	0.974	1.398	0.078	1.258	1.094	1.258	0.041
Item20	0.832	0.986	0.694	1.219	0.116	1.203	0.940	1.203	0.068
Item21	0.536	0.222	0.025	0.425	0.093	0.333	0.083	0.333	0.057
Item22	0.737	0.737	0.546	0.951	0.082	0.885	0.593	0.885	0.055
Item23	0.633	0.388	0.213	0.546	0.072	0.465	0.164	0.465	0.069
Item24	0.804	0.996	0.832	1.383	0.097	1.091	0.846	1.091	0.055
Item25	0.424	0.024	-0.078	0.170	0.058	0.176	-0.120	0.176	0.062
Item26	-1.147	-1.316	-1.640	-1.142	0.093	-1.015	-1.405	-1.015	0.084
Item27	0.775	0.810	0.583	1.006	0.089	1.008	0.706	1.008	0.067
Item28	0.540	0.130	-0.113	0.470	0.119	0.313	-0.140	0.313	0.091
Item29	0.770	0.832	0.589	1.074	0.098	0.884	0.654	0.884	0.061
Item30	0.408	-0.526	-0.756	-0.279	0.094	-0.294	-0.727	-0.294	0.102
Item31	0.885	1.164	0.991	1.428	0.100	1.270	1.051	1.270	0.051
Item32	0.980	1.534	1.332	1.855	0.125	1.656	1.451	1.656	0.046

Table B4 Continued: Item Difficulty Estimates from Different Sampling Methods for 3-pl Model

		SLEV				Adj- SLEV			
Item	bFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item17	0.507	-0.095	-0.262	0.227	0.089	-0.078	-0.284	0.105	0.103
Item18	-3.137	-1.906	-2.066	-1.724	0.088	-1.941	-2.171	-1.732	0.100
Item19	0.860	1.175	1.063	1.284	0.053	1.152	1.040	1.246	0.052
Item20	0.832	1.064	0.873	1.226	0.070	1.059	0.889	1.217	0.075
Item21	0.536	0.228	0.099	0.405	0.072	0.247	0.108	0.386	0.059
Item22	0.737	0.756	0.651	0.871	0.052	0.780	0.671	0.939	0.059
Item23	0.633	0.337	0.225	0.479	0.059	0.352	0.213	0.571	0.083
Item24	0.804	0.933	0.798	1.056	0.059	0.936	0.819	1.053	0.052
Item25	0.424	0.069	-0.092	0.198	0.061	0.043	-0.053	0.179	0.051
Item26	-1.147	-1.261	-1.395	-1.169	0.059	-1.293	-1.459	-1.146	0.061
Item27	0.775	0.862	0.719	0.981	0.049	0.886	0.729	1.024	0.070
Item28	0.540	0.079	-0.199	0.302	0.097	0.133	-0.094	0.294	0.101
Item29	0.770	0.790	0.617	0.904	0.074	0.795	0.591	0.971	0.073
Item30	0.408	-0.514	-0.750	-0.221	0.101	-0.504	-0.681	-0.254	0.098
Item31	0.885	1.162	1.050	1.329	0.057	1.150	0.989	1.272	0.064
Item32	0.980	1.545	1.448	1.667	0.050	1.500	1.410	1.615	0.054

Table B5: Item Discrimination Estimates from Different Sampling Methods for 3-pl Model

		Random				Leverage			
Item	aFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item1	7.535	7.642	6.237	9.446	0.773	7.122	5.644	8.743	0.745
Item2	5.876	3.617	2.486	6.129	0.759	4.842	3.542	6.856	0.857
Item3	5.469	6.636	4.813	9.078	0.996	5.908	4.623	7.451	0.678
Item4	5.363	6.329	4.791	8.440	0.855	6.549	5.469	7.784	0.551
Item5	3.789	4.623	2.705	6.190	0.722	4.384	2.790	5.502	0.583
Item6	3.053	3.075	1.871	4.987	0.602	2.674	1.472	3.610	0.525
Item7	7.508	7.258	5.589	9.467	0.938	6.244	4.776	7.564	0.583
Item8	6.580	8.015	6.341	10.459	0.880	8.495	6.714	10.581	0.759
Item9	6.835	4.193	2.438	7.400	1.142	4.055	2.150	6.498	0.953
Item10	5.208	5.453	3.694	7.614	0.794	5.889	4.172	7.875	0.840
Item11	6.611	7.965	6.006	10.048	1.017	7.593	6.329	9.376	0.701
Item12	6.547	8.169	6.527	10.284	0.793	9.619	7.854	11.719	0.777
Item13	9.853	9.515	7.218	11.097	0.814	9.670	8.129	10.875	0.638
Item14	6.207	4.837	3.126	7.782	0.955	5.600	4.109	7.205	0.870
Item15	8.458	9.670	7.740	11.659	0.951	10.166	8.379	11.813	0.787
Item16	3.619	3.658	2.400	5.654	0.622	3.722	2.293	5.305	0.576

Table B5 Continued: Item Discrimination Estimates from Different Sampling Methods for 3-pl Model

		SLEV				Adj- SLEV			
Item	aFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item1	7.535	7.019	5.658	8.521	0.676	7.475	5.874	9.009	0.661
Item2	5.876	4.694	2.927	6.164	0.772	4.510	3.420	6.431	0.676
Item3	5.469	6.030	4.551	7.899	0.681	6.054	4.574	7.958	0.731
Item4	5.363	6.538	5.316	8.281	0.672	6.302	4.478	8.249	0.749
Item5	3.789	4.317	2.948	5.578	0.568	4.398	3.309	5.603	0.520
Item6	3.053	2.687	1.658	3.800	0.613	2.852	1.574	4.125	0.574
Item7	7.508	6.569	4.770	8.143	0.666	6.372	5.258	7.579	0.529
Item8	6.580	8.834	7.109	10.374	0.691	8.492	6.932	10.281	0.822
Item9	6.835	4.283	2.106	7.593	1.059	4.328	3.019	7.530	0.905
Item10	5.208	6.020	4.409	8.162	0.821	6.071	4.487	9.482	0.825
Item11	6.611	7.712	6.290	10.125	0.685	7.591	5.907	10.105	0.841
Item12	6.547	9.425	7.385	10.775	0.755	8.906	7.401	11.079	0.865
Item13	9.853	9.599	7.913	10.957	0.614	9.604	7.506	10.839	0.665
Item14	6.207	5.609	4.334	7.550	0.735	5.357	3.869	8.078	0.886
Item15	8.458	10.009	7.921	11.461	0.768	10.115	8.612	11.784	0.714
Item16	3.619	3.646	2.749	4.545	0.423	3.649	2.747	5.105	0.447

Table B5 Continued: Item Discrimination Estimates from Different Sampling Methods for 3-pl Model

		Random				Leverage			
Item	aFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item17	3.627	2.748	1.580	4.800	0.726	2.625	1.621	3.604	0.436
Item18	0.544	0.76	-0.466	1.789	0.492	0.275	-0.652	1.206	0.371
Item19	9.652	9.651	7.316	12.141	0.920	9.502	7.913	10.736	0.553
Item20	4.924	5.776	4.525	7.572	0.745	5.225	4.044	6.318	0.518
Item21	3.819	4.250	2.891	5.726	0.705	3.699	2.709	4.558	0.423
Item22	5.268	6.488	5.048	8.595	0.696	6.169	5.129	7.961	0.610
Item23	7.079	5.487	3.000	8.144	1.036	5.929	4.579	7.551	0.676
Item24	11.529	8.052	6.379	10.542	0.963	7.448	4.874	9.110	0.906
Item25	4.076	4.960	3.750	6.590	0.745	5.267	4.406	6.633	0.609
Item26	1.314	3.353	1.686	5.234	0.921	3.326	2.083	5.397	0.668
Item27	3.830	4.782	3.717	6.694	0.746	4.209	3.114	4.959	0.450
Item28	3.554	3.085	1.661	4.538	0.697	3.260	2.121	5.026	0.571
Item29	8.546	7.058	4.262	11.256	1.497	5.946	4.644	8.097	0.820
Item30	4.086	2.716	1.395	4.280	0.613	4.053	2.619	5.981	0.664
Item31	6.588	7.460	5.763	10.179	0.916	7.483	6.168	9.409	0.756
Item32	10.126	9.793	8.352	11.914	0.834	10.126	8.569	11.584	0.635

Table B5 Continued: Item Discrimination Estimates from Different Sampling Methods for 3-pl Model

		SLEV				Adj- SLEV			
Item	aFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item17	3.627	2.728	1.310	4.075	0.614	2.533	1.550	3.547	0.507
Item18	0.544	0.216	-0.673	1.321	0.384	0.404	-0.804	1.557	0.418
Item19	9.652	9.303	7.993	10.560	0.697	9.577	7.518	11.166	0.813
Item20	4.924	5.233	4.422	6.646	0.486	5.269	4.063	7.078	0.545
Item21	3.819	3.834	2.640	5.509	0.617	3.631	2.531	4.775	0.506
Item22	5.268	6.401	5.056	8.185	0.616	6.307	5.166	8.217	0.697
Item23	7.079	5.641	4.263	7.411	0.725	5.992	4.338	9.252	0.918
Item24	11.529	7.678	5.753	9.300	0.882	7.698	5.967	9.746	0.858
Item25	4.076	5.163	3.991	6.249	0.517	5.108	3.436	6.461	0.569
Item26	1.314	3.105	1.169	4.945	0.707	3.350	2.160	4.669	0.560
Item27	3.830	4.119	3.247	5.534	0.458	4.087	3.055	5.068	0.508
Item28	3.554	3.162	1.980	4.154	0.464	3.253	2.188	5.335	0.643
Item29	8.546	6.111	4.202	8.761	0.961	6.008	4.067	7.809	0.842
Item30	4.086	3.993	2.753	6.263	0.641	3.805	2.719	5.351	0.718
Item31	6.588	7.314	6.037	9.234	0.723	7.736	6.184	10.015	0.888
Item32	10.126	10.082	8.860	11.465	0.601	10.240	8.997	11.788	0.662

Table B6: Item Guessing Estimates from Different Sampling Methods for 3-pl Model

		Random				Leverage			
Item	cFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item1	0.013	-0.023	-0.044	-0.009	0.007	-0.020	-0.035	-0.009	0.006
Item2	0.328	0.259	0.188	0.372	0.039	0.259	0.169	0.352	0.038
Item3	0.022	0.008	-0.008	0.029	0.009	0.004	-0.012	0.021	0.008
Item4	0.010	0.019	-0.004	0.048	0.012	0.014	-0.005	0.038	0.01
Item5	0.031	0.055	0.022	0.097	0.016	0.044	0.015	0.080	0.015
Item6	0.120	0.131	0.084	0.185	0.021	0.114	0.076	0.177	0.023
Item7	0.026	-0.006	-0.029	0.012	0.010	-0.004	-0.017	0.012	0.008
Item8	0.000	-0.013	-0.027	0.000	0.006	-0.009	-0.021	0.011	0.006
Item9	0.239	0.203	0.142	0.267	0.032	0.206	0.119	0.277	0.035
Item10	0.097	0.093	0.048	0.158	0.021	0.087	0.047	0.155	0.022
Item11	0.000	-0.025	-0.041	-0.008	0.007	-0.021	-0.034	-0.012	0.005
Item12	0.036	0.051	0.026	0.080	0.014	0.043	0.013	0.073	0.014
Item13	0.000	-0.043	-0.053	-0.033	0.005	-0.037	-0.053	-0.025	0.005
Item14	0.226	0.188	0.133	0.253	0.031	0.191	0.132	0.275	0.030
Item15	0.001	-0.034	-0.046	-0.023	0.005	-0.029	-0.046	-0.014	0.005
Item16	0.089	0.101	0.066	0.151	0.021	0.087	0.053	0.135	0.018

Table B6 Continued: Item Guessing Estimates from Different Sampling Methods for 3-pl Model

		SLEV				Adj- SLEV			
Item	cFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item1	0.013	-0.020	-0.032	0.001	0.007	-0.022	-0.037	-0.008	0.007
Item2	0.328	0.260	0.183	0.353	0.037	0.261	0.189	0.344	0.035
Item3	0.022	0.006	-0.021	0.028	0.011	0.006	-0.014	0.026	0.009
Item4	0.01	0.016	-0.004	0.035	0.010	0.019	-0.002	0.040	0.009
Item5	0.031	0.050	0.019	0.083	0.014	0.049	0.022	0.082	0.015
Item6	0.12	0.115	0.074	0.151	0.021	0.123	0.075	0.187	0.027
Item7	0.026	-0.006	-0.022	0.010	0.007	-0.003	-0.017	0.019	0.009
Item8	0.000	-0.011	-0.022	0.005	0.006	-0.009	-0.021	0.013	0.007
Item9	0.239	0.208	0.150	0.265	0.032	0.219	0.160	0.282	0.030
Item10	0.097	0.087	0.045	0.121	0.019	0.087	0.028	0.151	0.024
Item11	0.000	-0.021	-0.032	-0.009	0.005	-0.023	-0.036	-0.007	0.006
Item12	0.036	0.045	0.019	0.083	0.014	0.046	0.020	0.089	0.015
Item13	0.000	-0.036	-0.046	-0.026	0.003	-0.040	-0.046	-0.029	0.004
Item14	0.226	0.196	0.132	0.281	0.031	0.197	0.133	0.277	0.032
Item15	0.001	-0.029	-0.039	-0.022	0.004	-0.030	-0.041	-0.016	0.006
Item16	0.089	0.094	0.060	0.134	0.017	0.089	0.056	0.133	0.017

Table B6 Continued: Item Guessing Estimates from Different Sampling Methods for 3-pl Model

		Random				Leverage			
Item	cFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item17	0.265	0.212	0.161	0.302	0.033	0.202	0.141	0.282	0.027
Item18	0.000	0.355	0.312	0.436	0.031	0.346	0.305	0.395	0.021
Item19	0.000	-0.039	-0.048	-0.030	0.004	-0.033	-0.047	-0.02	0.004
Item20	0.007	-0.007	-0.024	0.012	0.008	-0.003	-0.027	0.014	0.009
Item21	0.089	0.107	0.059	0.171	0.026	0.092	0.044	0.135	0.016
Item22	0.000	-0.007	-0.025	0.008	0.007	-0.005	-0.016	0.011	0.006
Item23	0.221	0.184	0.133	0.256	0.025	0.199	0.135	0.264	0.030
Item24	0.167	0.165	0.115	0.220	0.022	0.176	0.107	0.231	0.025
Item25	0.007	0.056	0.024	0.082	0.013	0.058	0.035	0.094	0.013
Item26	0.000	0.213	0.155	0.280	0.023	0.210	0.178	0.232	0.013
Item27	0.000	0.014	-0.008	0.034	0.011	0.005	-0.010	0.020	0.008
Item28	0.210	0.196	0.131	0.297	0.035	0.204	0.127	0.278	0.033
Item29	0.176	0.168	0.118	0.217	0.025	0.162	0.095	0.219	0.026
Item30	0.384	0.253	0.199	0.356	0.029	0.279	0.194	0.364	0.036
Item31	0.018	-0.008	-0.023	0.006	0.007	-0.004	-0.019	0.014	0.008
Item32	0.001	-0.041	-0.050	-0.032	0.004	-0.035	-0.046	-0.024	0.005

Table B6 Continued: Item Guessing Estimates from Different Sampling Methods for 3-pl Model

		SLEV				Adj-SLEV			
Item	cFULL	Mean	Min	Max	Stdv	Mean	Min	Max	Stdv
Item17	0.265	0.209	0.157	0.359	0.035	0.201	0.162	0.261	0.027
Item18	0.000	0.349	0.307	0.404	0.026	0.345	0.301	0.403	0.027
Item19	0.000	-0.033	-0.041	-0.024	0.004	-0.037	-0.044	-0.024	0.004
Item20	0.007	-0.004	-0.022	0.013	0.008	-0.004	-0.018	0.014	0.008
Item21	0.089	0.096	0.060	0.140	0.018	0.099	0.046	0.127	0.017
Item22	0.000	-0.008	-0.020	0.017	0.007	-0.006	-0.021	0.018	0.008
Item23	0.221	0.190	0.130	0.245	0.03	0.199	0.125	0.300	0.036
Item24	0.167	0.167	0.118	0.209	0.024	0.160	0.097	0.214	0.025
Item25	0.007	0.057	0.032	0.082	0.013	0.055	0.030	0.088	0.013
Item26	0.000	0.211	0.170	0.263	0.022	0.205	0.163	0.256	0.019
Item27	0.000	0.006	-0.010	0.036	0.009	0.009	-0.008	0.028	0.009
Item28	0.210	0.192	0.126	0.272	0.036	0.206	0.158	0.297	0.030
Item29	0.176	0.162	0.113	0.220	0.025	0.158	0.109	0.224	0.024
Item30	0.384	0.278	0.198	0.394	0.039	0.268	0.206	0.370	0.037
Item31	0.018	-0.007	-0.023	0.010	0.009	-0.007	-0.025	0.015	0.008
Item32	0.001	-0.035	-0.042	-0.025	0.004	-0.038	-0.048	-0.024	0.005

APPENDIX C

SAMPLE R CODES

C.1.Sample R Code for Plotting Item Characteristic Curves (ICCs) in Chapter 2 (Figures 1-3)

```
#ICC for Rasch
Model######################################################
#generate model parameters for ICC#1
a1<-1     #item discrimination
b1<--1    #item difficulty
cc1<-0    #item guessing

#generate theta
theta<-rnorm(1000,0,1)

#Calculate model probabilities
prob1<-cc1+(1-cc1)/(1+exp(-a1*(theta-b1)))

#create and order dataset to plot
data1<-cbind(theta,prob1)
k1<-data1[order(theta),]
k1<-as.data.frame(k1)

#plot ICC#1
plot(k1$theta,k1$prob1,type="l",xlab="Ability",ylab=expression(paste("P"[i],"
(",theta,")")),main="Item Characteristic Curve (ICC) for Rasch Model")
text(-1.2,0.6, substitute(b[1]==b1, list(b1 = b1)))
axis(2, at=seq(0, 1, by=0.1), labels = F)
axis(1, at=seq(-4, 5, by=1), labels = F)

#generate model parameters for ICC#2
a2<-1     #item discrimination
b2<-0    #item difficulty
cc2<-0    #item guessing

#Calculate model probabilities
prob2<-cc2+(1-cc2)/(1+exp(-a2*(theta-b2)))

#create and order dataset to plot
data2<-cbind(theta,prob2)
k2<-data2[order(theta),]
k2<-as.data.frame(k2)

#plot ICC#2
par(new=T)
plot(k2$theta,k2$prob2,type="l",xlab="", ylab="", xaxt="n", yaxt="n")
text(-0.91,0.38, substitute(b[2]==b2, list(b2 = b2)))

#generate model parameters for ICC#3
a3<-1     #item discrimination
b3<-1    #item difficulty
cc3<-0    #item guessing

#Calculate model probabilities
prob3<-cc3+(1-cc3)/(1+exp(-a3*(theta-b3)))
```

```
#create and order dataset to plot
data3<-cbind(theta,prob3)
k3<-data3[order(theta),]
k3<-as.data.frame(k3)

#plot ICC#3
par(new=T)
plot(k3$theta,k3$prob3,type="l",xlab="", ylab="", xaxt="n", yaxt="n")
text(0.7,0.30, substitute(b[3]==b3, list(b3 = b3)))

#ICC for 2-pl
Model###########################################################
#generate model parameters for ICC#1
a1<-0.80     #item discrimination
b1<-0   #item difficulty
cc1<-0    #item guessing

#generate theta
theta<-rnorm(1000,0,1)

#Calculate model probabilities
prob1<-cc1+(1-cc1)/(1+exp(-a1*(theta-b1)))

#create and order dataset to plot
data1<-cbind(theta,prob1)
k1<-data1[order(theta),]
k1<-as.data.frame(k1)

#plot ICC#1
plot(k1$theta,k1$prob1,type="l",xlab="Ability",ylab=expression(paste("P"[i],"
(",theta,")")),main=bquote(atop(paste("Item Characteristic Curve (ICC) for 2-
pl Model"),paste("b"[i],"=0"))))
text(-1.5,0.3, substitute(a[1]==a1, list(a1 = a1)))
axis(2, at=seq(0, 1, by=0.1), labels = F)
axis(1, at=seq(-4, 5, by=1), labels = F)

#generate model parameters for ICC#2
a2<-1.5  #item discrimination
b2<-0    #item difficulty
cc2<-0   #item guessing

#Calculate model probabilities
prob2<-cc2+(1-cc2)/(1+exp(-a2*(theta-b2)))

#create and order dataset to plot
data2<-cbind(theta,prob2)
k2<-data2[order(theta),]
k2<-as.data.frame(k2)

#plot ICC#2
par(new=T)
plot(k2$theta,k2$prob2,type="l",xlab="", ylab="", xaxt="n", yaxt="n")
text(-1.15,0.15, substitute(a[2]==a2, list(a2 = a2)),cex=1)

#generate model parameters for ICC#3
a3<-3     #item discrimination
b3<-0    #item difficulty
cc3<-0    #item guessing

#Calculate model probabilities
prob3<-cc3+(1-cc3)/(1+exp(-a3*(theta-b3)))
```

```
#create and order dataset to plot
data3<-cbind(theta,prob3)
k3<-data3[order(theta),]
k3<-as.data.frame(k3)

#plot ICC#3
par(new=T)
plot(k3$theta,k3$prob3,type="l",xlab="", ylab="", xaxt="n", yaxt="n")
text(0.2,0.30, substitute(a[3]==a3, list(a3 = a3)))

#ICC for 3-pl Model########################################################
#generate model parameters for ICC#1
a1<-1     #item discrimination
b1<-0  #item difficulty
c1<-0.0  #item guessing

#generate theta
theta<-rnorm(1000,0,1)

#Calculate model probabilities
prob1<-c1+(1-c1)/(1+exp(-a1*(theta-b1)))

#create and order dataset to plot
data1<-cbind(theta,prob1)
k1<-data1[order(theta),]
k1<-as.data.frame(k1)

#plot ICC#1
plot(k1$theta,k1$prob1,type="l",xlab="Ability",ylab=expression(paste("P"[i],"
(",theta,")")),main=bquote(atop(paste("Item Characteristic Curve (ICC) for 3-
pl Model"),paste("b"[i],"=0",", ","a"[i],"=1"))))

text(-2.80,0.1, substitute(c[1]==c1, list(c1 = c1)))
axis(2, at=seq(0, 1, by=0.1), labels = F)
axis(1, at=seq(-4, 4, by=1), labels = F)

#generate model parameters for ICC#2
a2<-1     #item discrimination
b2<-0    #item difficulty
c2<-0.10    #item guessing

#Calculate model probabilities
prob2<-c2+(1-c2)/(1+exp(-a2*(theta-b2)))

#create and order dataset to plot
data2<-cbind(theta,prob2)
k2<-data2[order(theta),]
k2<-as.data.frame(k2)

#plot ICC#2
par(new=T)
plot(k2$theta,k2$prob2,type="l",xlab="", ylab="", xaxt="n", yaxt="n",ylim=c(0
,1))
text(-2.70,0.2, substitute(c[2]==c2, list(c2 = c2)),cex=1)

#generate model parameters for ICC#3
a3<-1     #item discrimination
b3<-0    #item difficulty
c3<-0.20    #item guessing

#Calculate model probabilities
prob3<-c3+(1-c3)/(1+exp(-a3*(theta-b3)))
```

```
#create and order dataset to plot
data3<-cbind(theta,prob3)
k3<-data3[order(theta),]
k3<-as.data.frame(k3)

#plot ICC#3
par(new=T)
plot(k3$theta,k3$prob3,type="l",xlab="", ylab="", xaxt="n", yaxt="n",ylim=c(0
,1))
text(-2.70,0.30, substitute(c[3]==c3, list(c3 = c3)))
```

C.2. Sample R Code for Semi-parametric IRT Analyses in Chapter 4

```
install.packages("sirt")

library(sirt)

#read data
math_data<-read.csv("C:/.csv",sep=",")

#Rasch model#################################################################

#Rasch model assuming normal trait distribution
irtRaschnormal<- rasch.mml2(math_data)
summary(irtRaschnormal)

#Rasch model with log-linear smoothing up to two moments

irtRaschSkew1<- rasch.mml2( math_data,distribution.trait="smooth2")
summary(irtRaschSkew1)

#Rasch model with log-linear smoothing up to three moments
irtRaschSkew2<- rasch.mml2( math_data,distribution.trait="smooth3")
summary(irtRaschSkew2)

#Rasch model with log-linear smoothing up to four moments
irtRaschSkew3<- rasch.mml2( math_data,distribution.trait="smooth4")
summary(irtRaschSkew3)

#comparison of models for model fit
IRT.compareModels( irtRaschnormal,irtRaschSkew1,irtRaschSkew2,irtRaschSkew3)

#2-PL model##################################################################
I <- ncol(math_data) # number of items

#2pl model assuming normal trait distribution;
irt2plnorm<- rasch.mml2(math_data,est.a = 1:I )
summary(irt2plnorm)

#2pl model with log-linear smoothing up to two moments
irt2plskew1<- rasch.mml2(math_data,est.a = 1:I,distribution.trait="smooth2" )
summary(irt2plskew1)

#2pl model with log-linear smoothing up to three moments
irt2plskew2<- rasch.mml2(math_data,est.a = 1:I,distribution.trait="smooth3" )
summary(irt2plskew2)

#2pl model with log-linear smoothing up to four moments;
irt2plskew3<- rasch.mml2(math_data,est.a = 1:I,distribution.trait="smooth4" )
summary(irt2plskew3)

#compare models
IRT.compareModels( irt2plnorm , irt2plskew1, irt2plskew2, irt2plskew3 )
```

```
#3-PL model###################################################################

#3pl model assuming normal trait distribution;
irt3plnorm<- rasch.mml2( math_data, est.a = 1:I , est.c = 1:I,mmliter = 10000
) # maximal 10,000 iterations
summary(irt3plnorm)

#3pl with log-linear smoothing up to two moments
irt3plskew1<- rasch.mml2( math_data, est.a = 1:I , est.c = 1:I,mmliter = 1000
0,distribution.trait="smooth2")
summary(irt3plskew1)

#3pl with log-linear smoothing up to three moments
irt3plskew2<- rasch.mml2( math_data, est.a = 1:I , est.c = 1:I,mmliter = 1000
0, distribution.trait="smooth3")
summary(irt3plskew2)

#3pl with log-linear smoothing up to three moments
irt3plskew3<- rasch.mml2( math_data, mmliter =10000,est.a = 1:I , est.c = 1:I
,distribution.trait="smooth4")
summary(irt3plskew3)

#compare models
IRT.compareModels(irt3plnorm,irt3plskew1,irt3plskew2,irt3plskew3)
```

C.3. Sampling Data Sets based on Sampling Methods in Chapter 4

```
#read full data to be sampled

data_sample <-read.csv("C:/.csv",sep=",")

##Random Sampling
##############################################################
# probability of selecting a data point based on random sampling method
m=1/2058

#vector of probabilities
prandom=rep(m,2058)

#replication 1 (sample full dataset based on random sampling method)
datarandom1 <- data_sample[sample(1:nrow(data_sample),548,replace=F,prob=pran
dom),]

#make an excel datafile

datatowrite1<- datarandom1
write.csv(datatowrite1"C:/.csv",row.names=FALSE)

##Leverage-based Sampling
#######################################################

#generate a covariate which has 0.90 correlation with total mathematics score
n     <- 2058                   # sample size
rho   <- 0.90                  # desired correlation
theta <- acos(rho)             # corresponding angle
x1    <- data_sample$sum        # total mathematics scores
x2    <- rnorm(n,5,20)      # new random variable
X     <- cbind(x1, x2)         # create new data matrix
Xcen  <- scale(X, center=TRUE, scale=FALSE)   # center x1 and x2 (mean 0)

Identity   <- diag(n)                         # identity matrix of size n
```

```
QR    <- qr.Q(qr(Xcen[ , 1, drop=FALSE]))     # QR-decomposition
Project    <- tcrossprod(QR)                  # projection onto space defined
by x1
x2ort  <- (Identity - Project) %*% Xcen[,2]   # find x2 orthogonal to x1
Xnew  <- cbind(Xcen[ , 1], x2ort)             # bind to matrix
Y    <- Xnew %*% diag(1/sqrt(colSums(Xnew^2))) # scale columns to have length
of 1

x <- Y[,2] + (1/tan(theta)) * Y[,1]           # generated covariate
cor(x1, x)                                    # check correlation

#rescale covariate
d<-round(x,3)
cov<-d*20
cor(x1,cov)

#create new dataset
newd90<-cbind(newdata$sum,cov)
newd90<-as.data.frame(newd90)

#regression for calculating leverage scores
summary(m90 <- glm( V1~ cov, data=newd90))

h90<-hatvalues(m90)    #gives leverage values, calculated only based on X's
h_total90<-sum(h90)
h_normal90<-h90/h_total90   #normalized leverage scores, a.k.a. leverage-base
d probabilities

#replication 1 (sample full dataset based on leverage-based sampling method)
dataleverage901 <- data_sample[sample(1:nrow(data_sample),548,replace=F,prob=
h_normal90),]  #sample data rows

##Shrinkage-based Sampling
#######################################################
alpha=0.9
pp=1/2058
m=rep(pp,2058)

## probability of selecting a data point based on shrinkage-based sampling me
thod
probshrinkage90=alpha*h_normal90+(1-alpha)*m

#generate leverage90 data

#replication 1 (sample full dataset based on shrinkage-based sampling method)
datashrink901 <- data_sample[sample(1:nrow(data_sample),548,replace=F,prob=pr
obshrinkage90),]  #sample data rows


##Adjusted Shrinkage-based Sampling
#################################################
#normalized leverage scores
h_normal90<-as.matrix(h_normal90)

# random probability of selecting a data point
p=1/2058
m=rep(p,2058)
n<-as.matrix(m)

#function for probability of selecting a data point based on adjusted shrinka
ge-based sampling method
piadjshrink<-matrix(,2058,)
for (i in 1:2058){
  if (h_normal90[i,]>n[i,]){
    piadjshrink[i,]<-h_normal90[i,]
```

```
  }
  else{
    piadjshrink[i,]<-n[i,]
  }
}

#replication 1
dataadjshrink901 <- data_sample[sample(1:nrow(data_sample),548,replace=F,prob
=piadjshrink),]  #sample data rows
```