Spring 2001

Methods of Detecting Differential Item Functioning: A Comparison of Item Response Theory and Confirmatory Factor Analysis

Ratchaneewan Wanichtanom
Old Dominion University

Recommended Citation

METHODS OF DETECTING DIFFERENTIAL ITEM FUNCTIONING:

A COMPARISON OF ITEM RESPONSE THEORY AND CONFIRMATORY

FACTOR ANALYSIS

by

Ratchaneewan Wanichtanom
B.A. February 1982, Thammasat, University of Thailand
M.S. July 1985, Mahidol University of Thailand
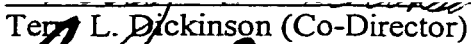M.S. December 1997, Old Dominion University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
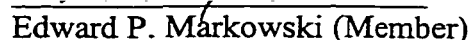Requirements for the Degree of

DOCTOR OF PHILOSOPHY

PSYCHOLOGY

OLD DOMINION UNIVERSITY
May 2001

Approved by:

Terry L. Dickinson (Co-Director)

Glynn D. Coates (Co-Director)

Louis H. Janda (Member)

Edward P. Markowski (Member)

ABSTRACT

METHODS TO DETECT DIFFERENTIAL ITEM FUNCTONING: A COMPARISON OF ITEM RESPONSE THEORY AND CONFIRMATORY FACTOR ANALYSIS

Ratchaneewan Wanichtanom
Old Dominion University, 2001
Co-Directors of Advisory Committee: Dr. Terry L. Dickinson
Dr. Glynn D. Coates

Differential item functioning (DIF) occurs when an item performs statistically differently for a reference group than for a focal group. DIF is a threat to the validity of a test, and it can lead to illegal usage of a test in situations such as employee selection. Thus, DIF has important implications for test construction and practice. This research is a Monte Carlo study that compares Item Response Theory (IRT) and three Confirmatory Factor Analysis (CFA) methods for detecting DIF. The three CFA methods were Model Comparison (MC), Modification Indexes (MI), and Modification Indexes-Divided sample (MI-Divided).

The research compared the detection rates of DIF by the methods for reference and focal groups. Each group consisted of 1000 examinees who responded to 50 items. Nine of the 50 items were designed to show DIF for the two groups. Responses were simulated using a two-parameter logistic model. Three types of DIF were manipulated for the nine items using the logistic model's a and b parameters. DIF was manipulated 1) only on the a (or discrimination) parameter, 2) only on the b (difficulty) parameter, and 3) on both the a and b parameters. In addition, the nine items were designed to have a crossing of the magnitude of the a and b parameters (i.e., low, medium, and high levels of the parameters). The amount of DIF was held constant to a value of .5 through use of

Raju's (1988) formula for the area separation between item characteristic curves of the reference and focal groups.

Results indicated that the all of the methods were very good at detecting DIF due to the b parameter (i.e., item difficulty). For DIF on the a and both the a and b parameters, the IRT and MI-Divided methods yielded significantly higher detection rates than the MC and MI methods. Further, the IRT and MI-Divided methods did not differ in their detection rates for these two types of DIF, and similarly, the MC and MI methods also did not differ in their detection rates. DIF due to both the a and b parameters was the hardest for all methods to detect. Although the MI-Divided method had a high detection rate, it also had false positive rates two to three times greater than expected. Future research on the methods was suggested for variables such as amount of DIF, sample size, and ability differences between focal and reference groups.

This dissertation is dedicated to my country 'Thailand'.

ACKNOWLEDGEMENTS

There are many people who I would like to thank for their help on my dissertation. I wish to thank my dissertation committee members: Dr. Terry Dickinson, Dr. Glynn Coates, Dr. Louis Janda, and Dr. Edward Markowski. Dr. Terry Dickinson has been my advisor since I entered graduate school in 1994 at Old Dominion University. He has helped to tailor my graduate study to suit my interests. He also provided me advice for my social adjustment in the United States. Because English is my second language, this required of him long hours of hard work for research guidance and editing of manuscripts. Without him, my dissertation would not be able to be completed. He taught me how to do good research, and I am so grateful that I have worked with him. It was an honor that Dr. Glynn Coates accepted to be co-director of my dissertation committee. He also provided guidance to me throughout my graduate study. He welcomed me to attend his classes several times so that I could gain knowledge as I improved in my English comprehension. While I was on my internship in Denver, he was my statistical backup. Whenever I had questions on statistical methods related to my internship, I could count on him to provide guidance. He is a teacher to be admired for his kindness and helpfulness. His example made me promise to myself that someday I would help my own students as he did for me. Dr. Louis Janda was also very helpful and patient. He agreed to become a member of my dissertation committee before I finished my qualifying examination. I did not have a proposal meeting until almost two years later, but he still agreed to remain a member of my dissertation committee.

Dr. Edward Markowski is the outside member of my dissertation committee. I contacted him less than a month before the proposal meeting. Although he was quite

busy with other students, he agreed to serve on my committee. He has provided me with good advice when I was a student in his class and also for my dissertation.

I would like to acknowledge the sponsorship of the Thai government. My government provided me the opportunity and financial support for graduate study at Old Dominion University from July 1994 to July 2000.

I would like to acknowledge Mary Boswell, Jacqueline Winston, and Peggy Kinard who work in the main Psychology office. They were supportive and helpful over my years at the university. They did anything they could to make my graduate experience easier. My special thanks to Peggy, who is also my friend. She supported me when I was away on my internship and had a hard time adjusting to life in Denver, Colorado.

Many students in the program were nice and helpful to me. Dr. Andrea Berndt, who was my peer advisor during my first year at the university, provided tremendous psychological support and helped with my language development. Dr. Johana Merritt and Dr. Jeff Sinn were also very kind and friendly to me. Arlene Pace Greene and Simon Bartle were also good friends and provided strong support.

Dr. Raymond Kirby is another person to whom I wish to express my gratitude. He retired from academic life recently. His kindness and help in his and other classes during the most difficult times are highly appreciated. My hope is also to become an excellent teacher.

Another good friend of mine, Mark Lesley from Mathematics Department, helped me with my language skills without any hesitation.

There are people from other universities who helped with my dissertation, and

I would like to express my gratitude toward them: Dr. Frans Jeroen Oort of the University of Amsterdam, and Dr. Frank Baker of the University of Wisconsin. Both of them freely provided materials that helped me accomplish this dissertation.

My thanks go to my parents, sister, and brothers who supported me throughout my many years of study in the United States.

Lastly, I thank Cenk Ozcan who unconditionally supported and helped me with all problems that I had during my time at Old Dominion University. He always cared for my comfort and happiness. I consider myself very lucky to know him.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

An item demonstrates differential item functioning (DIF) when it performs statistically differently for one group of examinees when compared to another group of examinees. The term "item bias" has been used synonymously for DIF. However, item bias is now understood to be a broader term that involves two procedures: (1) a statistical analysis that identifies the item as functioning differently for the groups, and (2) a follow-up analysis that generates substantive explanations for the group differences (Holland & Wainer, 1993). Using the term item bias implies that the item has been assessed with both procedures, whereas most of the time the term is intended to mean only that the item has been assessed statistically as functioning differently for the groups. DIF is a more neutral term to use for the statistical identification of biased items.

In a society with many ethnic groups like the United States, it is important to identify and eliminate DIF items from a test. DIF items are a threat to the validity of the test, and they can lead to unfair decision making in organizations for group members (e.g., job selection, promotion), creating legal problems that are both costly and time consuming to group members and organizations. Clearly, research into DIF has important implications for test construction and practice.

Types of Differential Item Functioning Detection Methods

There are two major types of DIF detection methods: external and internal methods. For external methods, all group members must have both a test score and a criterion score. Then, regression analysis is performed on the test and criterion scores.

The Journal of Applied Psychology was used as a style guide for formatting this dissertation

If the predicted criterion scores from the regression line are too high or too low for a group, the test shows evidence of DIF. The major concern in using an external method is the difficulty of finding criteria that are reliable and free of measurement error.

Internal methods use the test score as the criterion, which is more convenient than external methods because no other criterion is needed. The shortcoming of the internal method is that when a particular test contains many DIF items, some DIF items may not be significantly different from the test score. Thus, internal methods are prone to Type II errors in the presence of many DIF items.

This present study focuses on two internal methods of DIF detection. The ability of these two methods to detect several types of DIF are compared in a computer simulation. One method is based on Item Response Theory (IRT) and employs a likelihood ratio test. The second method is based on Confirmatory Factor Analysis (CFA), and it also employs a likelihood ratio test to detect DIF items. Because CFA is a relatively new approach to DIF detection, several variations of the CFA method are included in the study. Both IRT and CFA are explained in the following pages.

Item Response Theory and Classical Test Theory Methods

Before IRT became popular, DIF detection methods were based on Classical Test Theory (CTT). The CTT methods of DIF detection include Transformed Item Difficulty index, Adjustments to the Transformed Item Difficulty index (Angoff, 1972), The Golden Rule procedure (Golden Rule Insurance Company et al. v. Washburn et al., 1984), Analysis of Variance (Cleary & Hilton, 1968), and differences in item point biserial correlations. Camilli and Shepard (1994) reviewed all of these methods and recommended that test developers should not use them, due to their relative

ineffectiveness in comparison to IRT methods for detecting DIF.

Although the most promising DIF detection methods are based on IRT, these methods have limitations. One major concern in using IRT methods is that they need longer tests and larger sample sizes, which are not likely to be attained in some situations (e.g., classroom examinations). Consequently, whenever the assumptions of IRT seem unlikely to be attained, CTT methods should still be considered for DIF detection.

In order to understand the IRT method of DIF detection, IRT is summarized in the next section.

Item Response Theory

As mentioned previously, IRT has advantages over CTT. CTT has the disadvantages of being test and sample dependent and not being easily replicated in measurement. These disadvantages make test results incomparable from sample to sample, and from one test form to another test form. IRT makes it possible to compare test results from different groups of examinees and from different forms of tests because of its property of invariance of item and ability parameters. The property of invariance of item and ability parameters indicates that the item parameters do not depend on the sample of the examinees, and the ability parameters do not depend on the particular set of test items. Thus, IRT item parameters will theoretically retain the same values no matter the sample of examinees that is used to estimate the parameters (Suen, 1990).

IRT has two basic postulates: (1) The performance of an examinee on a test item can be predicted by a set of latent variables called traits, constructs, or abilities; and (2) the relationship between examinees' item performance and the set of traits underlying

item performance can be described by a monotonically increasing function called an item characteristic curve (ICC).

The ICC specifies that, as the level of the ability (trait or construct) increases, the probability of a correct response to an item increases. An example of an ICC is presented below:
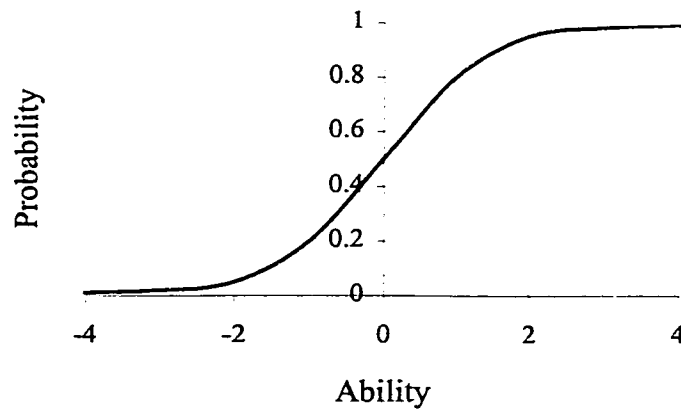
Figure 1. Item Characteristic Curve

Plots of ICCs are quite useful in DIF detection studies. Typically, each plot contains the ICCs for two groups. One group is labeled the reference group (e.g., majority group, males, native speaking), and the remaining group is the focal group (e.g., minority group, females, non-native speaking). The area between the two ICCs reflects the extent of DIF. If the ICCs overlap completely and there is no area between the ICCs, then the item shows no DIF. In other words, the item performs the same for the reference and focal groups, and two examinees with the same ability but from different groups will have the same probability of answering the item correctly. However, if the ICCs do not overlap, the item does not perform the same for the two groups, and any two examinees

with the same ability will likely have different probabilities of answering the item correctly.

Basic Assumptions of Item Response Theory

IRT has two assumptions that must be met. The unidimensionality assumption requires that only one ability is measured by a set of items in a test. The local independence assumption requires that the responses of examinees to any pair of items are statistically independent when the ability influencing test performance is held constant.

Popular Models in Item Response Theory

The most popular IRT models are the one-, two- and three-parameter logistic models (Hambleton, Swaminathan, & Roger, 1991). These models are appropriate for dichotomous item response data.

One-Parameter Logistic Model. The one-parameter logistic model is one of the most widely used IRT models. Item characteristic curves for the one-parameter logistic model are given by the equation

$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}} \quad i = 1,2,\ldots,n$$

In this equation, $P_i(\theta)$ is the probability that a randomly chosen examinee with ability θ answers item i correctly. Typically, the ability values of a group of examinees are standardized to have mean of zero and standard deviation of one. Although the values of ability can range from $-\infty$ to $+\infty$, most ICC figures display ability for the one-parameter model as ranging from -4 to $+4$.

The b_i parameter is called the difficulty parameter for item i. Values for the parameter can range from -2.0 to 2.0. The value of the b_i parameter is determined in

reference to the inflection point on the ICC of the logistic model. A vertical projection

from the inflection point onto the ability scale determines the value of the b_i parameter.

The b_i parameter can also be defined as the point on the ability scale where the

probability of a correct response is .5. Thus, the difficulty parameter indicates the

positioning of the ICC on the ability scale. The greater the value of the b_i parameter, the

greater the ability that is required for an examinee to have a 50% chance of getting the

item correct. Thus, items with greater values of the b_i parameter are more difficult for

examinees to answer correctly, and the items are positioned more towards the higher end

of the ability scale (i.e., towards +4).

Two-Parameter Logistic Model. Item characteristic curves for the two-parameter

logistic model are given by the equation (Birnbaum, 1968)

$$P(\theta) = \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}} \qquad i = 1, 2, \ldots, n$$

The two-parameter model has two additional elements compared to the one-parameter

model. D is a scaling constant usually set equal to 1.7 or 1.702 (Hulin, Drasgow &

Parsons, 1983). This scaling constant is added for historical reasons to make the shape of

the logistic function match as close as possible to that of the normal ogive function. The

constant does not change the basic nature of the logistic model.

The a_i parameter is called the discrimination parameter for item i. The a_i

parameter reflects the slope or steepness of the ICC at its inflection point. The steeper

the curve, the more useful the item for separating examinees into different ability levels.

The a_i parameter is defined on a scale from $-\infty$ to $+\infty$. However, items with negative

parameter values are discarded from ability tests because something is wrong with an

item if the probability of answering it correctly decreases as ability increases. It is also

unusual to have a_i values larger than 2. Thus, the typical range for item discrimination parameters is between 0 and 2.

The one-and two-parameter models make no allowance for guessing behavior. The assumption of no guessing is most plausible with free-response items, but it often can be met approximately with multiple-choice items when a test is not too difficult for examinees.

Three-Parameter Logistic Model. The mathematical expression for the three-parameter logistic model is

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} \quad i = 1, 2, \ldots, n$$

The c_i parameter is added to the two-parameter model to define the three-parameter model. This parameter is called the pseudo-chance level parameter (or guessing parameter). It is the lower asymptote of the ICC (or the probability of answering an item correctly when θ is $-\infty$), and the parameter corresponds to the probability of a correct response among respondents with a very low level of ability.

<center>*Types of Differential Item Functioning*</center>

Mellenberg (1982) distinguished two types of DIF: Uniform and nonuniform. Uniform DIF occurs when there is no interaction between ability level and group membership. That is, the probability of answering the item correctly is uniformly greater for one group than the other group over all levels of ability. Thus, for uniform DIF items only the difficulty parameter (i.e., b_i) is different between groups but the discrimination parameter (i.e., a_i) is the same.

Nonuniform DIF occurs when there is interaction between ability level and group membership. That is, the difference in the probabilities of a correct answer for the two

groups is not the same at all ability levels. Nonuniform DIF is reflected by ICCs that are not parallel. Thus, the discrimination parameter is different between groups but the difficulty parameter can be the same or different. Nonuniform DIF can be separated into two types of interaction: disordinal and ordinal.

Disordinal interaction between ability and group membership is indicated by ICCs that cross in the middle of the ability range. In contrast, ordinal interaction is indicated by ICCs that cross at either the low end or the high end of the ability scale, resulting in ICCs that appear to be similar over most of the ability range.

Item Response Theory Methods for Differential Item Functioning Detection

There are three IRT methods for detecting DIF: (1) comparison of item parameters, (2) measurement of the area between ICCs, and (3) assessment of the fit of an item response model to the data. For larger sample sizes (i.e., 1000) and longer tests (i.e., 50), all three methods give similar results (Kim, Cohen, & Kim, 1994). In the present research, methods (2) and (3) are used: Measurement of the area between ICCs, and assessment of the fit of an item response model to the data.

For measuring the area between ICCs, Raju (1988) provided an exact expression for computing the area between the ICCs for the one-, two-, and three-parameter models. The area between the ICCs for the three-parameter model is

$$Area = (1 - c) \left| \frac{2(a_2 - a_1)}{Da_1 a_2} \right| \ln\left[1 + e^{Da_1 a_2 (b_2 - b_1)/(a_2 - a_1)} \right] - (b_2 - b_1)$$

The area between ICCs is used to determine the amount and type of DIF in the experimental items utilized in this simulation study. The utilization of the area formula is described in method section.

Item Response Theory and the Likelihood Ratio Test

The likelihood ratio test is computed in the present research using the Bock-Aitkin marginal maximum likelihood estimation algorithm (Bock & Aitkin, 1981). This algorithm is implemented in several IRT computer programs, including MULTILOG (Thissen, 1991) which is the program used in this simulation study. MULTILOG can estimate parameters for binary and multiple category response models (e.g., logistic, and graded response models).

The likelihood ratio test makes use of anchor items in the analysis of DIF. The anchor items are known or believed to be items that do not exhibit DIF and that contain information over a range of the ability continuum. The anchor items and additional items that may show DIF are analyzed in two separate models. Each model is fitted to all of the item responses.

The models differ in the equality constrains that are imposed on item parameters. Because the anchor items do not function differently in the reference and focal groups, the parameters for these items are constrained to be equal across the groups in both of the models. For example, suppose 50 items are available for DIF analysis, utilizing the two-parameter logistic model. Of these items, 41 constitute the anchor set, whereas the remaining 9 items are thought to exhibit DIF. In both of the two-parameter logistic models that are fit to item responses, the a_i and b_i parameters for the anchor items are constrained to be equal. However, the models differ in how they treat the suspected items.

Using terminology for the models offered by Judd and McClelland (1989), the compact model (C model) requires that all items be constrained to have equal parameter

values across the reference and focal groups. The augmented model (A model) contains the same items as the C model. However, only the anchor items are constrained to have equal parameter values across the groups. The parameters for the items suspected of DIF are estimated separately for the reference and focal groups. That is, the suspected items are hypothesized to take on distinct parameter values in the reference and focal groups. (As an aside, only one suspected item at a time is usually added to the anchor items to define and estimate a C model. However, all of the items suspected of DIF could be examined simultaneously.)

Associated with the A and C models are maximum likelihood values. The natural log transformation of each value approximately follows a chi-square distribution. Each chi-square value can be used to evaluate the fit of the corresponding model to the item responses. More importantly, the fit of the two models can be compared by forming the ratio of their maximum likelihood values. Clearly, the likelihood values of the two models will differ due to the unconstrained suspected item(s) in the A model. The likelihood ratio of the two models is

$$LR = \frac{Likelihood(ModelC)}{Likelihood(ModelA)}$$

The natural log transformation of the LR also approximately follows a chi-square distribution. This chi-square distribution has degrees of freedom equal to the number of additional parameters in the augmented model. This chi-square distribution is

$$\chi^2_{(\#ofadditionalparameter)} \approx -2\ln(LR) = [-2\ln Likelihood(ModelC)] - [-2\ln Likelihood(ModelA)]$$

If the augmented model does not provide better fit to item responses than the compact model, then the additional parameters specified in the augmented model are not

necessary. Thus, the compact model (which contains fewer parameters) is a more parsimonious model and is to be preferred due to its greater simplicity.

If the augmented model does provide a better fit to item responses than the compact model, at least one of the suspected items is functioning differently in the reference and focal groups. As mentioned previously, a likelihood ratio test (i.e., chi-square statistic) is typically computed for each item separately to identify whether it is functioning differently in the reference and focal groups.

The likelihood ratio test (LRT) has been found to be one of the best methods in DIF detection. For simulated data with no DIF, the LRT provides a Type I error rate very close to the expected alpha levels considered for the two-parameter model. For the three-parameter model, Type I error rates are greater than expected alpha levels from .0005 to .005. However, for the three-parameter model and alpha levels from .01 to .10, the Type I error rates are close to expected values (Cohen, Kim, & Wollack, 1996). When compared to other IRT methods of DIF detection, Cohen et al. concluded that the LRT provides Type I error rates that are closest to those expected theoretically.

Confirmatory Factor Analysis

Confirmatory Factor Analysis (CFA) is another promising technique for exploring DIF (Camilli & Shepard, 1994). CFA is a general model for testing hypotheses about relations between observed and latent variables (i.e., factors). Latent variables are abstract concepts that are not directly measured. The term latent variable is used interchangeably with the IRT concepts of trait, construct, or ability. Observed variables are directly measured and serve as indicators for the latent variables.

The CFA model is represented as:

$$X = \Lambda \xi + \delta$$

where X is a vector of observed variables (e.g., items in a test or questionnaire), Λ is a matrix of structural coefficients (i.e., factor loadings) for the latent variables, ξ is a vector of latent variables, and δ is a vector of measurement errors in the model.

In using CFA, a model is specified (to indicate hypothesized relationships among observed and latent variables) and evaluated for its fit to item responses. Associated with a CFA model is a likelihood ratio value whose natural logarithm follows approximately the chi-square distribution. CFA also gives modification indexes, which indicate how to adjust the model (by estimating new parameters that formerly were fixed to zero) to improve the fit to the item responses. Of course, estimating new parameters is equivalent to re-specifying (defining anew) the hypothesized model.

For DIF analysis in CFA, group membership is specified in the hypothesized model as an additional latent variable. After the estimation of the model, modification indexes are assessed to identify items showing DIF. Items with large and significant modification indexes on the group membership variable are identified as demonstrating DIF.

Each modification index is a statistic that theoretically approximates a chi-square distribution with one degree of freedom. However, theoretical assumptions are often violated due to model misspecification and the nature of actual data. These violations can result in identification of anchor items as showing DIF (i.e., false positives). Thus, Oort (1992, 1996, 1998) proposed an adjusted critical value (AC) for evaluating the significance of a modification index. The adjusted critical value is:

$$AC = [\chi^2/(C + df - 1)] \, C$$

where C is a chosen critical value (e.g., 05 or .01). If an item's modification index is greater than adjusted critical value, then the particular item is considered to show DIF. The adjusted critical value is especially useful when all of the items are under suspicion for DIF (i.e., there are no anchor items).

Another statistic from CFA that is useful for detecting DIF is the expected parameter change (EPC). EPC is an estimate of what the value of a fixed parameter (i.e., one fixed to zero in the current model) will be when it is estimated as a free parameter in a revised model. EPC is useful to indicate the direction of DIF. For example, if the EPC for item i on group membership variable is positive, it indicates that the item is more attractive (or easier) for respondents with high scores on the group membership variable.

Oort (1996, 1998) used CFA in a computer simulation to detect DIF and concluded that CFA is a promising method. Oort found that CFA detects DIF for dichotomous items (i.e., correct/incorrect) as well as IRT methods. Furthermore, CFA was better than IRT methods for detecting DIF in polychotomous responses (e.g., Likert-type items). In particular, he found that CFA was better for seven-point items, large sample sizes, small trait differences between groups, equal group sizes, and large amounts of DIF.

Another approach to CFA for detecting DIF is a model comparison method, which is logically similar to the use of compact and augmented models by IRT. In CFA, a first model considers anchor items and items suspected of DIF. All of the items have structural coefficients estimated for the latent variable. However, none of the items have coefficients estimated for the group membership variable. This set of estimated coefficients is for the hypothesized model of no DIF for the suspected items. The second

model also considers the anchor and suspected items. However, all of the suspected items have structural coefficients estimated both for the latent variable and the group membership variable, whereas the anchor items have coefficients estimated only for the latent variable. This second set of estimated coefficients is for the hypothesized model of DIF for the suspected items. The difference between the chi-square statistics for the two models is computed to evaluate the hypothesis of DIF for the suspected items.

In the case of a single item which is suspected of DIF, the model comparison and modification index methods produce chi-square statistics that are identical in value. Both methods evaluate whether the structural coefficient of the suspected item is non-zero on the group membership variable.

In this study, the model comparison method is included because of its similarity to the IRT method. Both employ a likelihood ratio test and associated chi-square statistic to compare two models of item functioning. The modification index method is also included as it was used in previous research (Oort, 1998). Further, the modification index method is more practical in that it provides detection statistics for each suspected item in a single CFA analysis.

A third CFA method for DIF detection is used in the present research. This method is based on the Mantel-Haenszel method for DIF detection. Unlike IRT and CFA methods, the Mantel-Haenszel method is not a latent variable model. It deals only with observed variables. However, the Mantel-Haenszel method is very popular because it is appropriate for small samples and easily implemented in statistical program packages. A major problem with the Mantel-Haenszel method has been that it is not sensitive to

nonuniform DIF (Swaminathan & Rogers, 1990; Rogers & Swaminathan, 1993; Narayanan & Swaminathan, 1996).

Mazor, Clauser, and Hambleton (1994) proposed a variation on the Mantel-Haenszel method that is quite effective in detecting nonuniform DIF. The method requires three analyses of item responses. First, the total sample is analyzed for the presence of DIF. Next, the total sample is divided into high and low scoring subgroups. Then, both subgroups are analyzed separately for the presence of DIF. If one of the three analyses indicates an item to show DIF, then the item is flagged as detected. Mazor et al. showed that this divided sample method can increase DIF detection from 68% to 82% without increasing the Type I error rate.

Previous research indicates that IRT methods can detect nonuniform DIF, but there is limited research on the ability of CFA methods to detect nonuniform DIF. One of the major differences between IRT and CFA is that the relationships between observed variables (i.e., test items) and the latent variable (i.e., ability) in IRT are non-linear, whereas the relationships between observed variables and the latent variable in CFA are linear. Thus, there is the possibility that CFA methods may not be sensitive to nonuniform DIF. The divided sample approach may improve the detection rate of CFA for nonuniform DIF items.

Variables Related to Differential Item Functioning Detection

Sample Size and Test Length

As mentioned previously, one of the disadvantages of IRT is that large sample sizes are usually required. Hulin, Lissak, and Drasgow (1982) indicated in their research that large numbers of items are not as necessary as large number of examinees. They

suggested that test lengths of 30 items with sample sizes of 500 appear sufficient for the two-parameter model and that test lengths of 50 items with sample sizes of 1,000 appear sufficient for the three-parameter model. Drasgow and Parsons (1983) also suggested that a sample size of 1,000 or more and a test length of 50 items or more appears to be adequate for the three-parameter logistic model when other assumptions about the data are satisfied (i.e. unidimensionality and local independence).

Swaminathan and Gifford (1983) studied the estimation of parameters in the three-parameter model using sample sizes of 50, 200, and 1,000. They found that increasing the sample size had only a slight effect in improving the accuracy of estimation of the b and c parameters and the ability of examinees, but increasing the sample size and the length of the test can improve the accuracy of the estimating the a parameter. For example, for a 20-item test using 1,000 examinees, the b and c parameter estimations are excellent, and a and ability parameter estimation are fair. For a 80 item test with 1,000 examinees, all parameters estimates are good.

Ree and Jensen (1983) studied the effect of sample size on the linear equating of item parameters. Equating is an important method for linking the metrics of several test forms. Simulation data of sample sizes 250, 500, 1,000, and 2,000 were used. The results indicated that the larger the sample size, the less the error of item parameter estimation. They concluded that a and b parameter estimates are stable and accurate if there are large numbers of examinees over a broad range of ability. However, the c parameter (guessing parameter) estimation requires a large number of examinees at very low ability levels. The study showed that the largest samples available are needed when item calibration (item parameter estimation) or equating (linking metrics) must be done.

In detail, they suggested that at least a sample size of 1,000 is recommended to minimize error in item parameter estimation.

In the present research, the sample size and the number of items are fixed at 1,000 and 50, respectively. The values were chosen based on the results of previous research on sample size and test length to ensure the stability of parameter estimation.

Research on Nonuniform DIF

Theoretically, the IRT method takes the discrimination parameter into account for DIF detection, and thus, it should be able to detect nonuniform DIF (i.e., interaction between ability level and group membership). This detection of DIF should occur when there are differences between the reference and focal groups in the discrimination parameter or in the discrimination and difficulty parameters.

Although it has been maintained that nonuniform DIF is rare, several researchers have found nonuniform DIF in real data using the IRT method (Ellis, 1989, Huang, Church, & Katigbak, 1997, Linn, Levine, Hastings, & Wardrop, 1981). Ellis found 10 of 251 items of a translated test showing DIF. Seven of 10 items showed nonuniform DIF. Huang et al. studied the translation of NEO-PI items and found that a large proportion (40%) of the NEO-PI items functioned differently across American and Philippine samples. Both discrimination and difficulty parameters functioned differently between the two cultural samples, which indicated that nonuniform DIF can occur in real data. Clearly, researchers should consider the effectiveness of detection methods for uniform and nonuniform DIF.

Previous research also shows that the IRT method can detect nonuniform DIF better than the popular Mantel-Haenszel method (Maranon, Garcia, & Costas, 1997).

Maranon et al. found that IRT detected 93% of the nonuniform items, Mantel-Haenszel detected 68% of the nonuniform items, and the Mantel-Haenzel divided sample approach detected 81% of the nonuniform items.

For CFA, there is only one article that addressed the detection of nonuniform DIF (Oort, 1998). Oort concluded that there was no difference in detection rates for uniform and nonuniform DIF. Unfortunately, Oort did not control the area between the ICCs for uniform and nonuniform items. Thus, type of DIF and the magnitude of DIF were confounded such that no firm conclusions may be drawn about the ability of CFA to detect nonuniform DIF.

In this study, both uniform and nonuniform DIF are investigated. There are three types of DIF: DIF on the a parameter only (nonuniform DIF), DIF on the a and b parameters (nonuniform DIF), and DIF on the b parameter only (uniform DIF).

Designing Differential Item Functioning

Research on DIF is usually conducted with a Monte Carlo study that allows the researcher to create data with 'known' item parameters. For example, item parameters for the reference group can be defined to represent a range of item difficulty and discrimination. These same item parameters would also be used for the focal group. However, a subset of the items would be specified as showing DIF for the focal group. For these specified items, constant values (e.g., .2, .5 or.8) are subtracted or added to their parameter values to create DIF. Using the item parameters for each group, item responses are then generated. Next, the responses for the two groups are analyzed for the effects of the "experimental" variables (e.g., sample size, detection method) that are hypothesized to influence the identification of DIF for the items designed to have DIF.

The parameters values for DIF can be designed in two different ways. In most previous studies, the researcher simply adds or subtracts constant values to item parameters to define an amount of DIF. This procedure works quite well for studies that investigate only uniform DIF (i.e., difficulty parameters are different between groups but the discrimination parameters are the same). However, if the researcher is interested in nonuniform DIF, this procedure does not work well, because adding or subtracting constant values to item parameters is not linearly related to the amount of DIF as reflected by the area separation of the ICCs for the reference and focal groups. This means that constant differences in item parameters do not yield constant differences in the area separation of item ICCs.

A second procedure actually uses the formula for the area between ICCs to find values that will hold constant the area between ICCs (Rogers & Swaminathan, 1993; Narayanan & Swaminathan, 1996). For example, suppose a researcher wants to define an item with nonuniform DIF as having an area equal to .5 between the ICCs for the reference and focal groups. Further, suppose the researcher wants to define the item to have low discrimination and low difficulty. For the reference group, values of $a = .5$ and $b = -1.5$ could define this item. For the focal group, the nonuniform item could be defined as less discriminating (e.g., $a = .39$). In this case, the item for the focal group can only take on a limited number of difficulty values (e.g., $b = -1.7$, -.2, 1.8) in order for the area between ICCs to be equal to .5. This fact is obtained by algebraic substitutions into Raju's (1988) area formula of the area value (i.e., .5), discrimination values for the reference and focal groups (i.e., .5 and .39, respectively), and the reference group

difficulty value (i.e., -1.5). As shown by this example, the difficulty values do not show a constant difference in magnitude for an equal area separation between the ICCs.

Level of Differential Item Functioning

Previous research indicates that most DIF can be successfully detected when the area between ICCs is at least .5. For example, Cohen, Kim, and Baker (1993) set the levels of DIF in their study to .5 and 1.0. Most of the items with DIF were detected. Oort (1996, 1998) studied three levels of DIF: Weak (.2), moderate (.5), and strong (.8). Most of the items with moderate and strong DIF were detected, but not those with weak DIF. However, Oort noted that the percentage of false positives was high (15-20%) in item detection. In this study, the level of DIF is fixed at .5 by calculating the area between ICCs using Raju's (1988) formula.

Proportion of Items with Differential Item Functioning

As mentioned before, one of disadvantages of internal methods of DIF detection is that they use the test score as the criterion. In particular, when there are many items in the test that exhibit DIF, these items may not be detected because the test score is itself highly determined by these items. For example, Oshima and Miller (1992) concluded that when the number of items with DIF increases, the detection rate decreases. Kim and Cohen (1992) also found that the number of false negatives (failures to detect the suspected items) increased as the percentage of items with DIF increased on the test. Finally, Cohen and Kim (1993) reported that when sample sizes are smaller (e.g., 100), tests are shorter (e.g., 20 items), or the proportion of items with DIF is large (e.g., greater than 20%), the rate of false negatives tends to increase. In the present research, the proportion of items with DIF is fixed at 18%.

Level of Item Parameters

The level of item parameters (i.e., magnitude of item difficulty and item discrimination parameters) may be related to the detection rate. For example, Rogers and Swaminathan (1993) used Mantel-Haenszel and Logistic Regression methods for DIF detection and found that the lowest detection rate occurred with items of medium difficulty and low discrimination, whereas the highest detection rate occurred with items of medium difficulty and high discrimination. In a similar study, Narayanan and Swaminathan (1996) used Simultaneous Item Bias and Logistic Regression methods. They found that the detection rates were highest for items with low difficulty and high discrimination, followed by medium difficulty and high discrimination items. The lowest detection rate was found for items with medium difficulty and low discrimination. No previous research has reported the influence of level of item parameters on detection using IRT or CFA methods.

In the present research, the difficulty and discrimination parameter values are set to three levels for items designed to show DIF: low, medium, and high. The levels of parameters of items designed not to show DIF are sample randomly.

Purpose of the Study

The purpose of this study is to compare the IRT method and three CFA methods for detecting DIF. The research design variables include type of DIF (uniform and nonuniform) and level of item parameters (low, medium, and high).

CHAPTER II

METHOD

Design

Uniform and nonuniform DIF were studied through comparisons of reference and focal groups. The three types of DIF were defined by (1) group differences on *a* parameters (nonuniform DIF), (2) group differences only on *a* and *b* parameters (nonuniform DIF), and (3) group differences only on *b* parameters (uniform DIF). DIF was also studied through a combination of low, medium and high discrimination parameter values with low, medium and high difficulty parameter values. Test size was fixed at 50 items. Thus, for the reference and focal groups, 41 of the 50 items were non-DIF, whereas the remaining 9 items were DIF.

Data were simulated for 25 replications to compare reference and focal groups. The number of replications was chosen based on the advice of Harwell, Stone, Hsu, and Kirisci (1996), who suggested this number for a Monte Carlo study in IRT. Thus, the total number of data sets is 100: 25 replications for each of the three types of DIF (or focal groups), and 25 data sets for the reference group.

Data Generation

The data generation followed several phases using SAS program (SAS Institute, 1998). The core programs are presented in Appendix A. First, ability parameters were generated. The ability parameters of the reference and focal groups were drawn from a normal distribution with a mean of zero and standard deviation of one (See Appendix A; p. 56).

Second, item parameter values were generated (see Appendix A; p. 72). The first

9 of the 50 items in each data set were the items designed to show DIF. As mentioned

previously, the parameter values for these items reflected a crossing of three levels of

discrimination (i.e., low, medium, and high) with three levels of difficulty (i.e., low,

medium, and high). The parameter values were set for the reference group following

recommendations found in the literatures that were discussed previously. The parameter

values were set for the focal group using Raju's (1988) area formula. These values for

the groups are shown in Table 1. The ICCs for the items showing DIF are shown in

Appendix B.

Table 1

Discrimination (a) and Difficulty (b) Parameters for Items showing DIF

	Reference Group		Focal Group1 (a DIF)		Focal Group2 (ab DIF)		Focal Group3 (b DIF)	
Item	a	b	A	b	a	b	a	b
1	0.5001	-1.5000	0.3830	-1.5000	0.3880	-1.7000	0.5001	-2.0000
2	0.5001	0	0.3830	0	0.3880	-0.2000	0.5001	-0.5000
3	0.5001	2.0000	0.3830	2.0000	0.3880	1.8000	0.5001	1.5000
4	1.0000	-1.5000	0.6200	-1.5000	0.6340	-1.7000	1.0000	-2.0000
5	1.0000	0	0.6200	0	0.6340	-0.2000	1.0000	-0.5000
6	1.0000	2.0000	0.6200	2.0000	0.6340	1.8000	1.0000	1.5000
7	2.0000	-1.5000	0.8990	-1.5000	0.9280	-1.7000	2.0000	-2.0000
8	2.0000	0	0.8990	0	0.9280	-0.2000	2.0000	-0.5000
9	2.0000	2.0000	0.8990	2.0000	0.9280	1.8000	2.0000	1.5000

As shown in Table 1, there are three types of DIF: DIF from *a* parameters

(nonuniform DIF), DIF from both *a* and *b* parameters (nonuniform DIF), and DIF from *b*

parameters only (uniform DIF). Each type of DIF is represented as a focal group.

For focal group 1, the *b* parameter was equal to that of the reference group, and

the *a* parameter was chosen to make the area between the ICCs equal to .5. For the focal

group 2, the *b* parameter was chosen to be .2 less than that for the reference group. Then,

the a parameter was chosen to make the area between the ICCs equal to .5. For focal group 3, the b was chosen to be .5 less that for the reference group, which also gives an area between the ICCs (for the same a) equal to .5.

For the remaining 41 items, both discrimination and difficulty parameters were sampled randomly. Using a SAS program, the discrimination parameter was sampled from a uniform distribution with a range from 0 to 2.0, whereas the difficulty parameter was sampled from a normal distribution with a range from -2.0 to 2.0 (See Appendix A; p.73). Of course, the discrimination and difficulty parameter values for the 41 items were identical for the reference and focal groups.

Finally, a SAS program was used to generate item responses for 1000 examinees separately for the reference and focal groups (See Appendix A; p.82). The probability of a correct answer to each item was calculated using the formula for the two-parameter logistic model. In order to introduce measurement error, a number was sampled from a uniform distribution (within an interval from 0 to 1). A simulated item response was scored as correct if the sampled number was less than the probability of a correct response as calculated with the logistic model. If the sampled number was more than the probability of a correct response, it was scored as an incorrect response.

Differential Item Functioning Detection Methods

The simulated data of 50 item responses of 1,000 examinees for the reference group and 1,000 examinees for the focal group were investigated for DIF using the IRT and CFA methods.

IRT Method

This method compares a compact model and augmented model for 42 items.

These items consist of the 41 items designed to be non-DIF, and one item designed to exhibit DIF. The compact model specifies all of the 42 items to be non-DIF for reference and focal groups. That is, the compact model estimates common parameter values (i.e., difficulty and discrimination) for the two groups. The augmented model specifies 41 items to be non-DIF and the remaining item to be the item suspected of DIF. That is, the augmented model estimates common parameter values for the 41 items for the two groups, and separate parameter values for the suspicious item for the two groups. Using MULTILOG, the steps in the analysis are as follows:

1. Estimate item parameter to obtain the chi-square goodness-of-fit statistic $G(1)$ for 42 items (items numbered 10 to 50 are anchor items, and one of the items numbered 1 to 9 is chosen as a suspected item.) For this analysis, impose the constraints that the parameter values (a and b) for every item are equal for the reference and focal groups. This analysis provides the statistic $G(1)$ for the compact model.

2. Estimate item parameter values as in step 1 except for one change to obtain the chi-square goodness-of-fit statistic $G(2)$. For this analysis, impose the constraints that the parameter values (a and b) for items 10 to 50 are equal for the reference and focal groups, but allow the suspected item to have its parameter values estimated separately for the reference and focal groups. This analysis provides the statistic $G(2)$ for the augmented model.

3. Compute the difference between $G(2)$ and $G(1)$. This difference approximately follows a chi-square distribution with 2 degrees of freedom.

4. If the difference $G(2)-G(1)$ exceeds the critical value, then statistically, the

suspected item shows DIF (See Appendix C).

5. Repeat steps 1-4 for the remaining suspected items.

CFA Methods

Three CFA methods were investigated: Model Comparison (MC), Modification Index (MI), and Modification Index-Divided Sample (MI-Divided). Each method specified two latent variables: ability and group membership. For all CFA methods, Proc Calis (SAS Institute, 1998) was used to perform analyses.

The Model Comparison method is logically similar to the IRT method. One series of steps for the Model Comparison method is as follows:

1. Estimate a "compact" model that includes all anchor items (item 10-50) plus one of the suspected DIF items. All items load on the ability factor. Only the group membership indicator variable loads on the group membership factor. The CFA provides $C(1)$, a chi-square goodness-of-fit value.

2. Estimate an "augmented" model that includes all anchor items plus the suspected item. Again, all items load on the ability factor, however, the group membership variable and the suspected item both load on the group membership factor. The CFA analysis provides $C(2)$, a chi-square goodness-of-fit value.

3. Compute the difference $C(2) - C(1)$. This difference approximately follows a chi-square distribution with 1 degree of freedom (See Appendix D).

4. If the difference exceeds the critical value (the second model is a significantly better fit than the first model), then the suspected item shows DIF.

5. Repeat steps 1-4 for the remaining suspected items.

As mentioned previously, the MC method and the MI method provide identical chi-square values when both evaluate a single suspected item. Therefore, the MI method for a single suspected item was actually used to calculate the chi-square values for the MC method. This strategy was more practical, because it required estimating only the compact model and noting the modification index for the single suspected item, rather than estimating the compact and augmented models and calculating the difference in their chi-square values.

The steps for the Modification Index method are as follows:

1. All items (Item 1-50) are included in the model. There are two latent variables: ability and group membership. All of the items load on the ability factor. Only the group membership indicator variable loads on the group membership factor. The analysis provides a modification index for each item on the group membership variable. Each modification index is a statistic from a distribution that approximately follows a chi-square distribution with 1 degree of freedom.

2. If a suspected item (i.e., items 1-9) has a modification index that exceeds the critical value of a chi-square distribution with 1 degree of freedom at .05 (i.e., 3.841), the item is detected as showing DIF. Because the present study has a small percentage of suspicious items (i.e., 18%), an adjusted critical value will not be used in this study (See Appendix E).

The steps for the MI-Divided method are as follows:

1. The steps of the MI method are repeated for three groups of respondents: (1) total group, (2) high performing group (i.e., total score of 25 or higher),

and (3) low performing group (i.e., total score of less than 25).

2. If the modification index for a suspected item in one or more of the three

    groups exceeds the critical value, the item is identified as showing DIF (See

    Appendix E).

Comparison of DIF Detection Methods

The McNemar Test (McNemar, 1962) is used to compare the detection rates of

the methods (See Appendix F, and G). The test indicates whether each method has better

detection rates than others.

CHAPTER III

RESULTS

Overview

The results of the analyses are presented in this order. First, the frequencies and detection rates of each method are described. This is done for each type of DIF (i.e., DIF on a, b, and ab parameters). The detection rates of the methods are compared between all pairings of types of DIF using the McNemar test for correlated proportions. Next, the frequencies and detection rates of each method for each type of DIF are described for the combinations of low, medium, high values of the a and b parameters. Finally, the false positive rates of detection for the MI and MI-Divided are presented.

Detection by the Four Methods for the Types of DIF

Table 2 displays the frequencies and detection rates by the four methods for each type of DIF. The table shows that the IRT and MI-Divided methods (84% and 85%, respectively) yield higher total rates of DIF detection than the MC and MI methods (71% and 70%, respectively).

Table 2 also shows that the type of DIF has an effect on the detection rate. DIF on the b parameter is detected best by all of the methods. Recall that DIF on the b parameter is uniform in nature (i.e., reference and focal group ICCs are parallel). Although the highest detection rate for the b DIF is obtained by the MC method, all of the methods appear to be sensitive to this type of DIF. Pairwise comparisons of the detection rates using the McNemar test confirm this conclusion. As shown in Table 3, none of the comparisons reach statistical significance ($p > .01$).

In contrast, the methods differ in their detection rates for DIF on the a and ab

Table 2

Frequencies and Detection Rates by the Methods for the Types of DIF

	IRT	MC	MI	MI-Divided
a DIF	185/225 (82%)	127/225 (56%)	128/225 (57%)	187/225 (83%)
ab DIF	164/225 (73%)	128/225 (57%)	125/225 (56%)	170/225 (76%)
b DIF	219/225 (97%)	221/225 (98%)	218/225 (97%)	219/225 (97%)
Total	568/675 (84%)	476/675 (71%)	471/675 (70%)	576/675 (85%)

Note. a DIF = discrimination (i.e., *a* parameter) differs between reference and focal groups. *ab* DIF = discrimination and difficulty (i.e., *a* and *b* parameters) differ between reference and focal groups. *b* DIF = difficulty (i.e., *b* parameter) differs between reference and focal groups.

parameters. Recall that DIF on these parameters is nonuniform in nature (i.e., reference and focal group ICCs are nonparallel). As shown in Table 2, both the IRT and MI-Divided are superior to the MC and MI methods in their detection rates. Comparisons of the methods using the McNemar test are shown in Table 3. The comparisons indicate that the IRT and MI-Divided methods are not statistically significantly different from each other in detection rate, whereas both are significantly greater than the MC and MI methods. Further, the MC and MI methods are not statistically significantly different from each other in their detection rates for DIF on the *a* and *ab* parameters.

Table 3

Comparisons of the Detection Rates by the Methods for the Types of DIF Using the McNemar Test

	a DIF			
	IRT	MC	MI	MI-Divided
IRT	-	52.56*	52.57*	0.22
MC	-	-	1.00	60.00*
MI	-	-	-	59.00*
MI-Divided	-	-	-	-

	ab DIF			
	IRT	MC	MI	MI-Divided
IRT	-	24.00*	26.68*	1.50
MC	-	-	0.60	38.35*
MI	-	-	-	45.00*
MI-Divided	-	-	-	-

	b DIF			
	IRT	MC	MI	MI-Divided
IRT	-	2.00	1.00	0.00
MC	-	-	3.00	2.00
MI	-	-	-	1.00
MI-Divided	-	-	-	-

Note. The critical value is 6.635 for chi-square with df = 1.
* ($p < .01$).

Detection Rates for a DIF

The nine items for all focal groups (i.e., a, ab, and b DIF) reflect a crossing of

low, medium, and high values of the a and b parameters. For the a DIF focal group, the

a parameter values are lower in magnitude than those of the reference group. However,

the b parameter values for the focal group are identical to those of the reference group (see Table 1). The a and b parameter values for the nine items achieve the same amount of DIF in terms of the area (i.e., .5) between the ICCs of the reference and focal groups. Three items (i.e., items 2, 5, and 8) reflect DIF that is nonuniform and disordinal in nature, whereas the remaining six items reflect DIF that is nonuniform and ordinal (see Appendix B). Further, the ICCs for the three disordinal items intersect at the ability scale value of zero to reflect that the reference and focal groups do not differ in ability.

The detection rates for the nine items with a DIF are shown in Table 4. Inspection of the table indicates that the nature of a DIF has an influence on detection rates. The MC and MI methods are quite poor (i.e., 0% to 8%) in detecting the three items showing a DIF at medium values of the b parameter (i.e., nonuniform and disordinal items). In contrast, the IRT and MI-Divided methods show greater sensitivity to these nonuniform and disordinal items (i.e., 24% to 100%). Although the IRT and MI-Divided methods are somewhat poor in detection for the low a, medium b item (i.e., 24%, 32%, respectively), they are quite sensitive for the remaining two nonuniform and disordinal items (i.e., 88% to 100%).

All of the methods show smaller detection rates for low values of the a parameter (across the levels of the b parameter). This is expected because with a low value for the a parameter, an item does not provide much differentiation of ability. Nonetheless, the IRT and MI-Divided methods are somewhat better in detection at low a values than the MC and MI methods.

In summary, the methods of detection can be divided into two groups: 1) methods that cannot detect a DIF for the medium value of the b parameter (i.e., nonuniform and

Table 4

Frequencies and Detection Rates of a DIF by Each Method for the Levels of a and b Parameters

Item Response Theory

	b parameter			
a parameter	Low	Med.	High	Total
Low	12/25 (48%)	6/25 (24%)	20/25 (80%)	38/75 (51%)
Med.	23/25 (92%)	25/25 (100%)	25/25 (100%)	73/75 (97%)
High	24/25 (96%)	25/25 (100%)	25/25 (100%)	74/75 (99%)
Total	59/75 (79%)	56/75 (75%)	70/75 (93%)	185/225 (82%)

Model Comparison

	b parameter			
a parameter	Low	Med.	High	Total
Low	9/25 (36%)	2/25 (8%)	16/25 (64%)	27/75 (36%)
Med.	23/25 (92%)	0/25 (0%)	25/25 (100%)	48/75 (64%)
High	25/25 (100%)	2/25 (8%)	25/25 (100%)	52/75 (69%)
Total	57/75 (76%)	4/75 (5%)	66/75 (88%)	127/225 (56%)

Table 4 *(continued)*

Modification Indexes

	b parameter			
a parameter	Low	Med.	High	Total
Low	10/25 (40%)	2/25 (8%)	16/25 (64%)	28/75 (37%)
Med.	23/25 (92%)	0/25 (0%)	25/25 (100%)	48/75 (64%)
High	25/25 (100%)	2/25 (8%)	25/25 (100%)	52/75 (69%)
Total	58/75 (77%)	4/75 (5%)	66/75 (88%)	128/225 (57%)

Modification Indexes-Divided Sample

	b parameter			
a parameter	Low	Med.	High	Total
Low	14/25 (56%)	8/25 (32%)	20/25 (80%)	42/75 (56%)
Med.	23/25 (92%)	22/25 (88%)	25/25 (100%)	70/75 (93%)
High	25/25 (100%)	25/25 (100%)	25/25 (100%)	75/75 (100%)
Total	62/75 (83%)	55/75 (73%)	70/75 (93%)	187/225 (83%)

disordinal items), which are MC and MI, and 2) methods that can detect a DIF for the medium value of the b parameter, which are IRT and MI-Divided. All methods are poorer in their detection of DIF for low values of the a parameter. For the remaining four items, all of the methods are quite sensitive in detecting DIF.

Figure 2 presents the detection rates of each method for the nine items showing a DIF. As seen in the figure, the detection rates toward the right side of the graph are greater in magnitude. Namely, for medium and high values of the a parameter, the detection rates are greater. Exceptions occur for the MC and MI methods on items 5 and 8 (medium values for the b parameter), because these methods cannot detect DIF that is nonuniform and disordinal in nature.
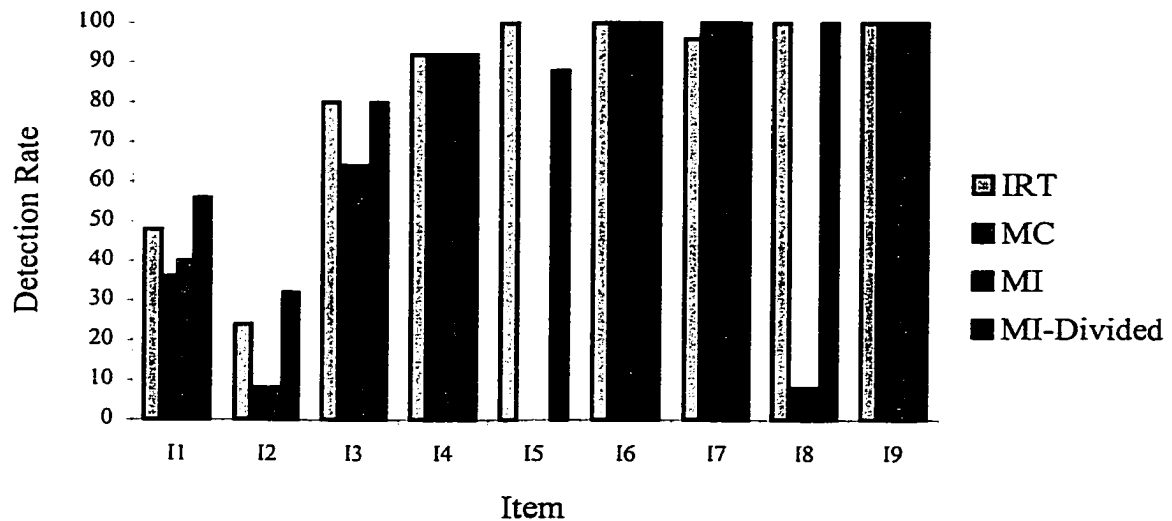
Figure 2. Detection of a DIF by the Four Methods

Note. Item 1 = low a low b; item 2 = low a medium b, item 3 = low a high b, item 4 = medium a low b, item 5 = medium a medium b, item 6 = medium a high b, item 7 = high a low b, item 8 = high a medium b, and item 9 = high a high b.

Figure 2 also shows that the detection rates toward the left side of the graph are smaller in magnitude. These items have small a values and are poor in discriminating ability levels. Nonetheless, the IRT and MI-Divided methods still show greater sensitivity in detection for the items compared to the MC and MI methods.

Detection Rates for ab DIF

The nine items that show DIF for this focal group have both a and b parameter values that are smaller in magnitude than those of the reference group. Nonetheless, the a and b parameters still achieve the same amount of DIF in terms of area (i.e., .5) between the ICCs of the reference and focal groups. Six of the nine items reflect DIF that is nonuniform and disordinal in nature (see Appendix B). Further, because the b parameter differs for the focal and reference groups, the ICCs intersect (i.e., are disordinal) above or below the ability scale value of zero to reflect the difference in group abilities. The remaining three items (i.e., items 3, 6, and 9) reflect DIF that is nonuniform and ordinal in nature.

The detection rates for the nine items with ab DIF are shown in Table 5. The methods are similar in their pattern of detection for ab DIF. That is, all of the methods are quite sensitive in detecting ab DIF for high values of the b parameter for all of the levels of the a parameter. As mentioned previously, these three items show DIF that is nonuniform and ordinal in nature.

For medium values of the b parameter, the IRT and MI-Divided methods are again superior in their detection rates to the MC and MI methods. However, the MC and MI methods are relatively better in detection than was the case for a DIF. This result occurs because ab DIF is due to reference and focal group differences on both of the

Table 5

Frequencies and Detection Rates of ab DIF by Each Method by the Levels of a and b Parameters

Item Response Theory

a parameter	b parameter			Total
	Low	Med.	High	
Low	9/25 (36%)	6/25 (24%)	20/25 (80%)	35/75 (47%)
Med.	15/25 (60%)	25/25 (100%)	25/25 (100%)	65/75 (87%)
High	14/25 (56%)	25/25 (100%)	25/25 (100%)	64/75 (85%)
Total	38/75 (51%)	56/75 (75%)	70/75 (93%)	164/225 (73%)

Model Comparison

a parameter	b parameter			Total
	Low	Med.	High	
Low	2/25 (8%)	4/25 (16%)	23/25 (92%)	29/75 (39%)
Med.	7/25 (28%)	12/25 (48%)	25/25 (100%)	44/75 (59%)
High	11/25 (44%)	19/25 (76%)	25/25 (100%)	55/75 (73%)
Total	20/75 (27%)	35/75 (47%)	73/75 (97%)	128/225 (57%)

Table 5 *(continued)*

Modification Indexes

a parameter	*b* parameter			Total
	Low	Med.	High	
Low	3/25 (12%)	4/25 (16%)	21/25 (84%)	28/75 (37%)
Med.	11/25 (44%)	10/25 (40%)	25/25 (100%)	46/75 (61%)
High	12/25 (48%)	14/25 (56%)	25/25 (100%)	51/75 (68%)
Total	26/75 (35%)	28/75 (37%)	71/75 (95%)	125/225 (56%)

Modification Indexes-Divided Sample

a parameter	*b* parameter			Total
	Low	Med.	High	
Low	10/25 (40%)	7/25 (28%)	22/25 (88%)	39/75 (52%)
Med.	16/25 (64%)	24/25 (96%)	25/25 (100%)	65/75 (87%)
High	16/25 (64%)	25/25 (100%)	25/25 (100%)	66/75 (88%)
Total	42/75 (56%)	56/75 (75%)	72/75 (96%)	170/225 (76%)

parameters. Thus, the MC and MI methods can detect *ab* DIF that is uniform and disordinal when this type of DIF includes difference on the *b* parameter (i.e., items 2, 5, and 8).

For the low values of the *b* parameter (i.e., item 1, 4, and 7), all of the methods are poorer in detection compared to their detection of *a* DIF. However, the MC and MI methods are much poorer in detection relative to the IRT and MI methods. Apparently, the result also occurs because *ab* DIF is due to reference and focal group differences on both of the parameters.

All of the methods again show smaller detection rates for low values of the *a* parameter. The three items (i.e., 1, 2, and 3) provide relatively less differentiation across ability levels.

In summary, the methods of detection still divide themselves into two groups. The MC and MI methods are again poorer in their detection of DIF that is nonuniform and disordinal in nature compared to the IRT and MI-Divided methods. However, the IRT and MI-Divided methods are even poor at detecting *ab* DIF for the low level of the *b* parameter. This is especially troublesome because items with low difficulty (i.e., *b* parameter) and medium and high discrimination (i.e., *a* parameter) may be most useful to measure focal group ability.

Figure 3 presents the detection rates of each method for the nine items that reflect *ab* DIF. For items 3, 6, and 9, which are nonuniform and ordinal in nature, there are small differences between the methods in detection rates. However, for the remaining items, which are nonuniform and disordinal, the IRT and MI-Divided methods are superior in their detection rates. Further, the detection rates for these items decrease as

you move to the left side of the graph. Thus, as the difficulty and discriminating power

of the items decrease (i.e., *a* and *b* parameter values), all of the methods are less capable

of detecting nonuniform and disordinal DIF.
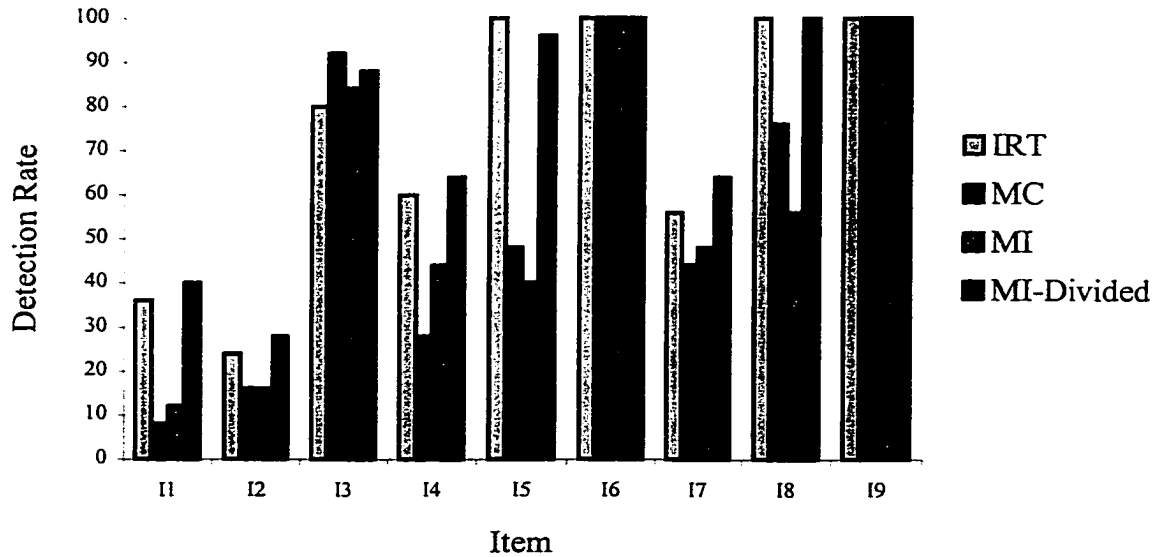
Figure 3. Detection of *ab* DIF by the Four Methods

Note. Item 1 = low *a* low *b*; item 2 = low *a* medium *b*, item 3 = low *a* high *b*, item 4 =
medium *a* low *b*, item 5 = medium *a* medium *b*, item 6 = medium *a* high *b*, item 7 = high
a low *b*, item 8 = high *a* medium *b*, and item 9 = high *a* high *b*.

Detection Rates for b DIF

The nine items showing DIF for this focal group have only *b* parameter values

that are smaller in magnitude than those of the reference group. All of these items reflect

DIF that is uniform in nature.

The frequencies and detection rates for the items with *b* DIF are shown in Table 6.

All of the methods are effective in detection this type of DIF. The smaller detection rates

occur for items with low discriminating power (i.e., items 1, 2, and 3). Nonetheless,

Table 6

Frequencies and Detection Rates of b DIF by Each Method by the Levels of a and b Parameters

Item Response Theory

	b parameter			
a parameter	Low	Med.	High	Total
Low	23/25 (92%)	24/25 (96%)	23/25 (92%)	70/75 (93%)
Med.	24/25 (96%)	25/25 (100%)	25/25 (100%)	74/75 (99%)
High	25/25 (100%)	25/25 (100%)	25/25 (100%)	75/75 (100%)
Total	72/75 (96%)	74/75 (99%)	73/75 (97%)	219/225 (97%)

Model Comparison

	b parameter			
a parameter	Low	Med.	High	Total
Low	25/25 (100%)	24/25 (96%)	23/25 (92%)	72/75 (96%)
Med.	24/25 (96%)	25/25 (100%)	25/25 (100%)	74/75 (99%)
High	25/25 (100%)	25/25 (100%)	25/25 (100%)	75/75 (100%)
Total	74/75 (99%)	74/75 (99%)	73/75 (97%)	221/225 (98%)

Table 6 *(continued)*

Modification Indexes

a parameter	*b* parameter			Total
	Low	Med.	High	
Low	22/25 (88%)	24/25 (96%)	23/25 (92%)	69/75 (92%)
Med.	24/25 (96%)	25/25 (100%)	25/25 (100%)	74/75 (99%)
High	25/25 (100%)	25/25 (100%)	25/25 (100%)	75/75 (100%)
Total	71/75 (95%)	74/75 (99%)	73/75 (97%)	218/225 (97%)

Modification Indexes-Divided Sample

a parameter	*b* parameter			Total
	Low	Med.	High	
Low	23/25 (92%)	24/25 (96%)	23/25 (92%)	70/75 (93%)
Med.	24/25 (96%)	25/25 (100%)	25/25 (100%)	74/75 (99%)
High	25/25 (100%)	25/25 (100%)	25/25 (100%)	75/75 (100%)
Total	72/75 (96%)	74/75 (99%)	73/75 (97%)	219/225 (97%)

these detection rates are still quite high (i.e., 88% to 100%).

Figure 4 shows the detection rates for the nine items showing *b* DIF. All of

the methods were completely successful in detection of DIF for items five through nine.

For the remaining items, the methods were less sensitive but they still show strong

capability to detect *b* DIF. One anomaly is the detection rate of the MC method for item

1.
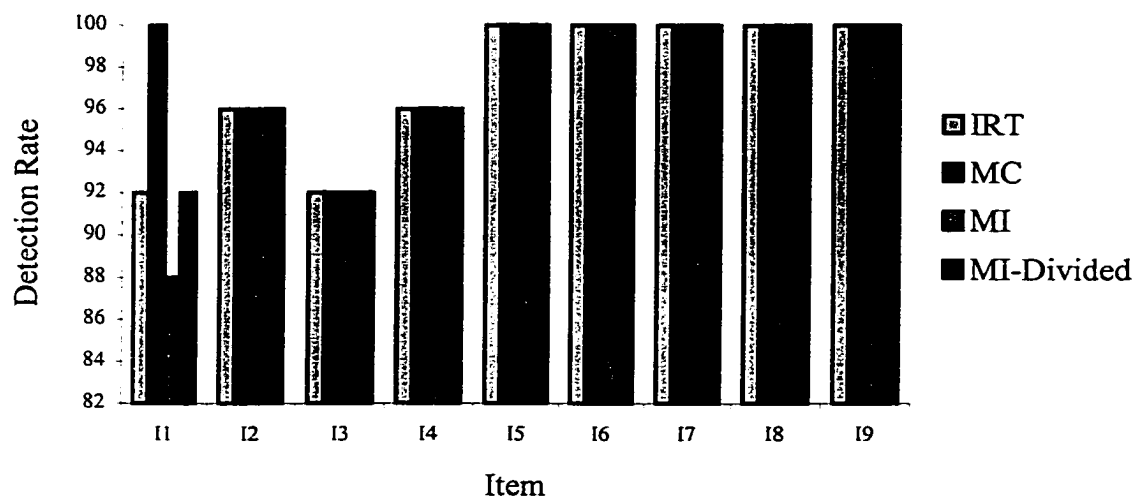
Figure 4. Detection of *b* DIF by the Four Methods

Note. Item 1 = low *a* low *b*; item 2 = low *a* medium *b*, item 3 = low *a* high *b*, item 4 = medium *a* low *b*, item 5 = medium *a* medium *b*, item 6 = medium *a* high *b*, item 7 = high *a* low *b*, item 8 = high *a* medium *b*, and item 9 = high *a* high *b*.

False Positive Rates of the MI and MI-Divided Methods

The MI and MI-Divided methods have a decided advantage over the IRT and MC

methods. The IRT and MC methods require a priori identification of items that are

known not to exhibit DIF. In contrast, the MI and MI-Divided methods do not have this

requirement. Consequently, the MI and MI-Divided methods could be used to identify

items showing DIF within the entire set of items. This advantage would be especially useful in the early stages of test development. Because of this advantage, the MI and MI-Divided methods were evaluated for error rates in identifying items with no DIF to show DIF (i.e., false positive rates). High rates of misidentification by a method are undesirable, because valuable items would be discarded.

As shown in Table 7, the MI method holds the false positive rate to an acceptable amount (i.e., 5%) for a and ab DIF. However, this method identifies twice as many items showing b as would be expected by chance. The MI-Divided method is less effective, showing error rates two to three times greater than would be expected by chance.

Table 7

False Positive Rates of MI and MI-Divided Methods

Types of DIF	MI	MI-Divided
a DIF	3%	9%
ab DIF	4%	11%
b DIF	11%	18%

CHAPTER IV

CONCLUSIONS

The general findings of the study, limitations of the research, and future research are discussed in the following sections.

General Findings

Uniform DIF was easier to detect (i.e., 97%) than both types of nonuniform (a and ab) DIF (i.e., 83% and 76%, respectively). This result replicates previous research findings. For example, Swaminathan and Rogers (1990) reported that the detection rate of uniform DIF was 94% for logistic regression and 96% for MH, whereas the detection rate of nonuniform DIF was 70% for logistic regression and 1% for MH.

All of the methods in the present study were able to detect uniform (b) DIF successfully. The detection rates of the four methods did not differ significantly from one another. For this type of DIF, the MI method is the practical choice for detection. A single analysis provides detection of items that show uniform DIF. The MI method could be used initially to remove these items before moving to methods for identifying items that may show nonuniform DIF. The MI method also has the advantage that it does not require a priori separation of items into a set known not to exhibit DIF and a second set suspected of DIF. Clearly, the MI method is quite useful in the early stages of measure development when little research is available about the functioning of items in focal groups.

For both types of nonuniform (a and ab) DIF, the detection rates of the IRT and MI-Divided methods are not statistically different from one another, and they show statistically significant superior detection rates compared to the MC and MI methods.

Furthermore, the detection rates of the MC and MI methods are not statistically different from one another. The differences in the detection rates between the two groups of methods are due to an ability to detect nonuniform DIF that is disordinal in nature.

The poor detection of nonuniform DIF by the MI method does not agree with the results reported by Oort (1998). He concluded there was no difference in the detection of uniform and nonuniform DIF. As suggested previously, Oort's results are probably an artifact of the confounding of type of DIF with the magnitude of DIF. In the present research, the amount of DIF was held constant to .5 using Raju's (1988) area formula.

Interestingly, the detection rates of the IRT and MI-Divided were poorer for nonuniform ab DIF (73% and 76%, respectively) than their detection rates for nonuniform a DIF (i.e., 82% and 83%, respectively). The greater number of disordinal items showing ab DIF (i.e., 6 of 9) compared to the number for a DIF (i.e., 3 of 9) may account for this finding. There is only one previous study that compared detection rates for types of nonuniform DIF (i.e., a vs. ab DIF). In contrast to the present study, Rogers and Swaminathan (1993) reported that logistic regression and MH methods had higher detection rates for ab DIF (78% and 80%, respectively) compared to a DIF (71% and 5%, respectively). Unfortunately, the amount of DIF and the sizes of the item parameter values used by Rogers and Swaminathan are not directly comparable to those used in the present research study.

The greater detection rate of the MI-Divided method compared to that of the MI method indicates that the technique of splitting a sample into low and high performance groups can increase detection rates. This result is similar to that found for the MH method when samples were split into performance groups. The improvement in detection

rates is due to the detection of items that show nonuniform DIF that is disordinal in nature. Such items have ICCs that intersect in the middle of the ability scale (cf. items 2, 5, and 8 for a and ab DIF in Appendix B). Apparently, when low and high performance groups are identified, this produces at least one pair of ICCs that are separated and do not intersect. Thus, the low and high groups used by the MI-Divided method actually show DIF due to differences in item difficulty (i.e., b DIF). It is precisely this type of DIF that the MI method can identify.

Limitations

This research has not studied many of the variables that occur in the real situations. For example, the level of DIF used was .5, which is a moderate amount of DIF. The purpose of using .5 was to provide a reasonable basis for comparing the DIF detection methods. For actual test items, there would be a mixture of items that also show weak (e.g., .2) and strong (e.g., .8) amounts of DIF.

Another important limitation of the present research is that there was no difference between reference and focal groups in ability. Clearly, there is the possibility that reference and focal groups will differ in average level of ability, and this may affect DIF detection rates by the methods.

Future Research

The CFA methods have shown themselves to be useful in the detection of DIF. The MI-Divided method was studied for the first time in the present research, and it showed detection rates equivalent to those of the IRT method. As mentioned previously, the MI method is a practical choice for detecting b DIF, because it requires only a single analysis. Future research should also consider a sequential strategy that uses both the MI

and MI-Divided methods. That is, the MI method could be used initially to screen items for uniform DIF and then the MI-Divided method could be used subsequently to screen items for nonuniform DIF. Such a strategy involves only three separate analyses and would provide excellent detection of DIF items.

In spite of the apparent advantages of the CFA methods, they showed the disadvantage of high false positive rates. The MI-Divided was particularly vulnerable to this problem. Future research should identify strategies to reduce the false positive rates. Oort's (1998) approach of adjusting the critical values for the modification index has shown promise, and further refinements may prove beneficial in controlling false positive rates.

As mentioned previously, several variables were fixed that have an impact on detection rates of the IRT method. Variables such as sample size, percentage of DIF items, and DIF size were fixed at levels that previous IRT research indicated were optimal for DIF detection. Little is known about the effects of these variables on the detection rates of the CFA methods. The CFA literature suggests that the modification index behaves quite well for sample sizes ranging from 200 to 400, and this result occurs for models more complex than the simple CFA model used for DIF detection (e.g., Marsh, Balla, &McDonald, 1988). Perhaps the CFA methods have an advantage over the IRT method for smaller sample sizes. Further, the MI and MI-Divided methods provide a modification index for each item separately, and these indexes may not be influenced by the percentage DIF items.

Finally, the use of CFA can be expanded to consider more complex models. Mellenbergh (1994) has described a general linear model approach to item response

theory that can be implemented in CFA by using multiple samples (e.g., reference and

focal groups) and mean structures. This general approach holds the promise of providing

separate modification indexes for uniform and nonuniform DIF. Oort (1996) has

discussed implementation of this approach using the multi-sample option of LISREL, but

no research has been done with this promising technique.

REFERENCES

Angoff, W. H. (1972, September). *A technique for the investigation of cultural differences.* Paper presented at the annual meeting of the American Psychological Association, Honolulu. (ERIC Document Reproduction Service No. ED 069 686)

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

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46,* 443-460.

Camilli, B., & Shepard, L. A. (1994). *Methods for identifying biased test items.* Thousand Oaks, CA: Sage.

Cleary, T. A., & Hilton, T. L. (1968). An investigation of item bias. *Educational and Psychological Measurement, 28,* 61-75.

Cohen, A. S., & Kim, S. H. (1993). A comparison of Lord's χ^2 and Raju's area measures in detection of DIF. *Applied Psychological Measurement, 17,* 39-52.

Cohen, A. S., Kim, S. H., & Baker, F.B. (1993). Detection of differential item functioning in the graded response model. *Applied Psychological Measurement, 17,* 335-350.

Cohen, A. S., Kim, S. H., & Wollack, J. A. (1996). An investigation of the likelihood ratio test for detection of differential item functioning. *Applied Psychological Measurement, 20,* 15-26.

Drasgow, F., & Parsons, C. K. (1983). Application of unidimensional item response theory models to multidimensional data. *Applied Psychological Measurement, 7,*

189-199.

Ellis, B. B. (1989). Differential item functioning: Implications for test translation. *Journal of Applied Psychology, 74,* 912-921.

Golden Rule Insurance Company et al. v. Washburn et al., 419-76 (stipulation for dismissal and order dismissing case, filed in the Circuit Court of the Seventh Judicial Circuit, Sangamon County, IL, 1984).

Hambleton, R. K., Swaminathan, H., & Roger, H. J. (1991). *Fundamentals of item response theory.* Newbury Park: Sage.

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

Holland, P. W., & Wainer, H. (1993). *Differential item functioning.* Hillsdale, NJ: Lawrence Erlbaum.

Huang, C. D., Church, A. T., & Katigbak, M. S. (1997). Identifying cultural differences in items and traits: Differential item functioning in the NEO personality inventory. *Journal of Cross-Cultural Psychology, 28,* 192-218.

Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). *Item response theory application to psychological measurement.* Homewood, IL: Dow Jones-Irwin.

Hulin, C. L., Lissak, R. I., & Drasgow, F. (1982). Recovery of two- and three-parameter logistic item characteristic curves: A monte carlo study. *Applied Psychological Measurement, 6,* 249-260.

Judd, C. M., & McClelland, G. H. (1989). *Data analysis: A model comparison approach.* San Diego, CA: Harcourt Brace Jovanovich.

Kim, S. H., & Cohen, A. S. (1992). Effects of linking methods on detection of DIF.

Journal of Educational Measurement, 29, 51-66.

Kim, S. H., Cohen, A. S., & Kim, H. O. (1994). An investigation of Lord's procedure for the detection of differential item functioning. *Applied Psychological Measurement, 18,* 217-228.

Linn, R. L., Levine, M. V., Hastings, C. N., & Wardrop, J. L. (1981). Item bias in a test of reading comprehension. *Applied Psychological Measurement, 5,* 159-173.

Maranon, P. P., Garcis, M. I. B., Costas, C. S. L. (1997). Identification of nonuniform DIF: A comparison of Mantel-Haenszel and IRT analysis procedure. *Educational and Psychological Measurement, 57(4),* 559-568.

Marsh, H.W., Balla, J.R., & McDonald, R.P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103,* 391-410.

Mazor, K. M., Clauser, B. E., & Hambleton, R. K. (1994). Identification of nonuniform differential item functioning using a variation of the Mantel-Haenszel procedure. *Educational and Psychological Measurement, 50(2),* 284-291.

McNemar, Q. (1962). *Psychological Statistics* (3rd ed.) New York: Wiley.

Mellenberg, G. J. (1982). Contingency table modes for assessing item bias. *Journal of Educational Statistics, 7,* 105-108.

Mellenbergh, G.J. (1994) Generalized linear item response theory. *Psychological Bulletin, 115,* 30

Narayanan, P., & Swaminathan, H. (1996). Identification of items that show nonuniform DIF. *Applied Psychological Measurement, 20,* 257-274.

Oort, F. J. (1992). Using restricted factor analysis to detect itemm bias. *Methodika, VI*, 150-166.

Oort, F. J. (1996). *Using restricted factor analysis in test construction.* Unpublished doctoral dissertation, University of Amsterdam, Amsterrdam, Netherlands.

Oort, F. J. (1998). Simulation study of item bias detection with restricted factor analysis. *Structural Equation modeling, 5*, 107-124.

Oshima, T. C., & Miller, D. M. (1992). Mulitdimensionality amd item bias in item response theory. *Applied Psychological Measurement, N6*, 237-248.

Raju, N. S. (1988). The area between two item characteristic curves. *Psychometrika, 53*, 495-502.

Ree, M. J., & Jensen, H. E. (1983). Effects of sample size on limear equating of item characteristic curve parameters. In D.J. Weiss (Ed.), *New horizons in testing: latent traits test theory and computerized adaptive testing.* New York: Academic Press.

Rogers, H. J., & Swaminathan, H. (1993). A comparison of logistic regression and Mantel-Haenszel procedures for detecting DIF. *Applied Psychological Measurement, 17*, 105-116.

SAS Institute. (1998). *Statistics user guide: Version 7.* Carey, NC: Author.

Suen, H. K. (1990). *Principles of test theories.* Hillsdale: Lawrence Erlbaum.

Swaminathan, H. & Gifford, J.A. (1983). Estimation of parameters in the three-parameter latent trait model. In D.J. Weiss (Ed.), *New Horizons in Testing: Latent Traits Test Theory and Computerized Adaptive Testing.* New York: Academic Press.

Swaminathan, H. & Rogers, H. J. (1990). Detecting DIF using Logistic Regression

procedures. *Journal of Educational Measurement*, 27(4), 361-370.

Thissen, D. (1991). *MULTILOG 6.30*. Scientific Software International: Lincolnwood,
IL.

APPENDIX A

SAS Programs for Data Generation

SAS Program: Generation of Ability Parameters

```
proc iml;
start a1;
lp: x1=normal(repeat(0,1000));
if any(x1>3) |any(x1<-3) then go to lp;
finish a1;
run a1;
create abi1 from x1[colname={abili1}];
append from x1;
;
proc iml;
start a2;
lp: x2=normal(repeat(0,1000));
if any(x2>3) |any(x2<-3) then go to lp;
finish a2;
run a2;
create abi2 from x2[colname={abili2}];
append from x2;
;
proc iml;
start a3;
lp: x3=normal(repeat(0,1000));
if any(x3>3) |any(x3<-3) then go to lp;
finish a3;
run a3;
create abi3 from x3[colname={abili3}];
append from x3;
;
proc iml;
start a4;
lp: x4=normal(repeat(0,1000));
if any(x4>3) |any(x4<-3) then go to lp;
finish a4;
run a4;
create abi4 from x4[colname={abili4}];
append from x4;
;
proc iml;
start a5;
lp: x5=normal(repeat(0,1000));
if any(x5>3) |any(x5<-3) then go to lp;
finish a5;
run a5;
create abi5 from x5[colname={abili5}];
append from x5;
```

```
;
proc iml;
start a6;
lp: x6=normal(repeat(0,1000));
if any(x6>3) |any(x6<-3) then go to lp;
finish a6;
run a6;
create abi6 from x6[colname={abili6}];
append from x6;
;
proc iml;
start a7;
lp: x7=normal(repeat(0,1000));
if any(x7>3) |any(x7<-3) then go to lp;
finish a7;
run a7;
create abi7 from x7[colname={abili7}];
append from x7;
;
proc iml;
start a8;
lp: x8=normal(repeat(0,1000));
if any(x8>3) |any(x8<-3) then go to lp;
finish a8;
run a8;
create abi8 from x8[colname={abili8}];
append from x8;
;
proc iml;
start a9;
lp: x9=normal(repeat(0,1000));
if any(x9>3) |any(x9<-3) then go to lp;
finish a9;
run a9;
create abi9 from x9[colname={abili9}];
append from x9;
;
proc iml;
start a10;
lp: x10=normal(repeat(0,1000));
if any(x10>3) |any(x10<-3) then go to lp;
finish a10;
run a10;
create abi10 from x10[colname={abili10}];
append from x10;
;
```

```
proc iml;
start a11;
lp: x11=normal(repeat(0,1000));
if any(x11>3) |any(x11<-3) then go to lp;
finish a11;
run a11;
create abi11 from x11[colname={abili11}];
append from x11;
;
proc iml;
start a12;
lp: x12=normal(repeat(0,1000));
if any(x12>3) |any(x12<-3) then go to lp;
finish a12;
run a12;
create abi12 from x12[colname={abili12}];
append from x12;
;
proc iml;
start a13;
lp: x13=normal(repeat(0,1000));
if any(x13>3) |any(x13<-3) then go to lp;
finish a13;
run a13;
create abi13 from x13[colname={abili13}];
append from x13;
;
proc iml;
start a14;
lp: x14=normal(repeat(0,1000));
if any(x14>3) |any(x14<-3) then go to lp;
finish a14;
run a14;
create abi14 from x14[colname={abili14}];
append from x14;
;
proc iml;
start a15;
lp: x15=normal(repeat(0,1000));
if any(x15>3) |any(x15<-3) then go to lp;
finish a15;
run a15;
create abi15 from x15[colname={abili15}];
append from x15;
;
proc iml;
```

```
start a16;
lp: x16=normal(repeat(0,1000));
if any(x16>3) |any(x16<-3) then go to lp;
finish a16;
run a16;
create abi16 from x16[colname={abili16}];
append from x16;
;
proc iml;
start a17;
lp: x17=normal(repeat(0,1000));
if any(x17>3) |any(x17<-3) then go to lp;
finish a17;
run a17;
create abi17 from x17[colname={abili17}];
append from x17;
;
proc iml;
start a18;
lp: x18=normal(repeat(0,1000));
if any(x18>3) |any(x18<-3) then go to lp;
finish a18;
run a18;
create abi18 from x18[colname={abili18}];
append from x18;
;
proc iml;
start a19;
lp: x19=normal(repeat(0,1000));
if any(x19>3) |any(x19<-3) then go to lp;
finish a19;
run a19;
create abi19 from x19[colname={abili19}];
append from x19;
;
proc iml;
start a20;
lp: x20=normal(repeat(0,1000));
if any(x20>3) |any(x20<-3) then go to lp;
finish a20;
run a20;
create abi20 from x20[colname={abili20}];
append from x20;
;
proc iml;
start a21;
```

```
lp: x21=normal(repeat(0,1000));
if any(x21>3) |any(x21<-3) then go to lp;
finish a21;
run a21;
create abi21 from x21[colname={abili21}];
append from x21;
;
proc iml;
start a22;
lp: x22=normal(repeat(0,1000));
if any(x22>3) |any(x22<-3) then go to lp;
finish a22;
run a22;
create abi22 from x22[colname={abili22}];
append from x22;
;
proc iml;
start a23;
lp: x23=normal(repeat(0,1000));
if any(x23>3) |any(x23<-3) then go to lp;
finish a23;
run a23;
create abi23 from x23[colname={abili23}];
append from x23;
;
proc iml;
start a24;
lp: x24=normal(repeat(0,1000));
if any(x24>3) |any(x24<-3) then go to lp;
finish a24;
run a24;
create abi24 from x24[colname={abili24}];
append from x24;
;
proc iml;
start a25;
lp: x25=normal(repeat(0,1000));
if any(x25>3) |any(x25<-3) then go to lp;
finish a25;
run a25;
create abi25 from x25[colname={abili25}];
append from x25;
;


libname abili1 'c:\my documents\my sas files\abili1';
    data abili1.ref1;
```

```
   set abi1;
proc standard mean=0 std=1;
var abili1;
run;
;
   data abili1.ref2;
   set abi2;
proc standard mean=0 std=1;
var abili2;
run;
;
   data abili1.ref3;
   set abi3;
proc standard mean=0 std=1;
var abili3;
run;
;
   data abili1.ref4;
   set abi4;
proc standard mean=0 std=1;
var abili4;
run;
;
   data abili1.ref5;
   set abi5;
proc standard mean=0 std=1;
var abili5;
run;
;
   data abili1.ref6;
   set abi6;
proc standard mean=0 std=1;
var abili6;
run;
;
   data abili1.ref7;
   set abi7;
proc standard mean=0 std=1;
var abili7;
run;
;
   data abili1.ref8;
   set abi8;
proc standard mean=0 std=1;
var abili8;
run;
```

```
;
    data abili1.ref9;
    set abi9;
proc standard mean=0 std=1;
var abili9;
run;
;
    data abili1.ref10;
    set abi10;
proc standard mean=0 std=1;
var abili10;
run;
;
    data abili1.ref11;
    set abi11;
proc standard mean=0 std=1;
var abili11;
run;
;
    data abili1.ref12;
    set abi12;
proc standard mean=0 std=1;
var abili12;
run;
;
    data abili1.ref13;
    set abi13;
proc standard mean=0 std=1;
var abili13;
run;
;
    data abili1.ref14;
    set abi14;
proc standard mean=0 std=1;
var abili14;
run;
;
    data abili1.ref15;
    set abi15;
proc standard mean=0 std=1;
var abili15;
run;
;
    data abili1.ref16;
    set abi16;
proc standard mean=0 std=1;
```

```
var abili16;
run;
;
   data abili1.ref17;
   set abi17;
proc standard mean=0 std=1;
var abili17;
run;
;
   data abili1.ref18;
   set abi18;
proc standard mean=0 std=1;
var abili18;
run;
;
   data abili1.ref19;
   set abi19;
proc standard mean=0 std=1;
var abili19;
run;
;
   data abili1.ref20;
   set abi20;
proc standard mean=0 std=1;
var abili20;
run;
;
   data abili1.ref21;
   set abi21;
proc standard mean=0 std=1;
var abili21;
run;
;
   data abili1.ref22;
   set abi22;
proc standard mean=0 std=1;
var abili22;
run;
;
   data abili1.ref23;
   set abi23;
proc standard mean=0 std=1;
var abili23;
run;
;
   data abili1.ref24;
```

```
    set abi24;
proc standard mean=0 std=1;
var abili24;
run;
;
   data abili1.ref25;
   set abi25;
proc standard mean=0 std=1;
var abili25;
run;
;

proc iml;
start a1;
lp: x1=normal(repeat(0,1000));
if any(x1>3) |any(x1<-3) then go to lp;
finish a1;
run a1;
create abi1 from x1[colname={abili1}];
append from x1;
;
proc iml;
start a2;
lp: x2=normal(repeat(0,1000));
if any(x2>3) |any(x2<-3) then go to lp;
finish a2;
run a2;
create abi2 from x2[colname={abili2}];
append from x2;
;
proc iml;
start a3;
lp: x3=normal(repeat(0,1000));
if any(x3>3) |any(x3<-3) then go to lp;
finish a3;
run a3;
create abi3 from x3[colname={abili3}];
append from x3;
;
proc iml;
start a4;
lp: x4=normal(repeat(0,1000));
if any(x4>3) |any(x4<-3) then go to lp;
finish a4;
run a4;
create abi4 from x4[colname={abili4}];
```

```
append from x4;
;
proc iml;
start a5;
lp: x5=normal(repeat(0,1000));
if any(x5>3) |any(x5<-3) then go to lp;
finish a5;
run a5;
create abi5 from x5[colname={abili5}];
append from x5;
;
proc iml;
start a6;
lp: x6=normal(repeat(0,1000));
if any(x6>3) |any(x6<-3) then go to lp;
finish a6;
run a6;
create abi6 from x6[colname={abili6}];
append from x6;
;
proc iml;
start a7;
lp: x7=normal(repeat(0,1000));
if any(x7>3) |any(x7<-3) then go to lp;
finish a7;
run a7;
create abi7 from x7[colname={abili7}];
append from x7;
;
proc iml;
start a8;
lp: x8=normal(repeat(0,1000));
if any(x8>3) |any(x8<-3) then go to lp;
finish a8;
run a8;
create abi8 from x8[colname={abili8}];
append from x8;
;
proc iml;
start a9;
lp: x9=normal(repeat(0,1000));
if any(x9>3) |any(x9<-3) then go to lp;
finish a9;
run a9;
create abi9 from x9[colname={abili9}];
append from x9;
```

```
;
proc iml;
start a10;
lp: x10=normal(repeat(0,1000));
if any(x10>3) |any(x10<-3) then go to lp;
finish a10;
run a10;
create abi10 from x10[colname={abili10}];
append from x10;
;
proc iml;
start a11;
lp: x11=normal(repeat(0,1000));
if any(x11>3) |any(x11<-3) then go to lp;
finish a11;
run a11;
create abi11 from x11[colname={abili11}];
append from x11;
;
proc iml;
start a12;
lp: x12=normal(repeat(0,1000));
if any(x12>3) |any(x12<-3) then go to lp;
finish a12;
run a12;
create abi12 from x12[colname={abili12}];
append from x12;
;
proc iml;
start a13;
lp: x13=normal(repeat(0,1000));
if any(x13>3) |any(x13<-3) then go to lp;
finish a13;
run a13;
create abi13 from x13[colname={abili13}];
append from x13;
;
proc iml;
start a14;
lp: x14=normal(repeat(0,1000));
if any(x14>3) |any(x14<-3) then go to lp;
finish a14;
run a14;
create abi14 from x14[colname={abili14}];
append from x14;
;
```

```
proc iml;
start a15;
lp: x15=normal(repeat(0,1000));
if any(x15>3) |any(x15<-3) then go to lp;
finish a15;
run a15;
create abi15 from x15[colname={abili15}];
append from x15;
;
proc iml;
start a16;
lp: x16=normal(repeat(0,1000));
if any(x16>3) |any(x16<-3) then go to lp;
finish a16;
run a16;
create abi16 from x16[colname={abili16}];
append from x16;
;
proc iml;
start a17;
lp: x17=normal(repeat(0,1000));
if any(x17>3) |any(x17<-3) then go to lp;
finish a17;
run a17;
create abi17 from x17[colname={abili17}];
append from x17;
;
proc iml;
start a18;
lp: x18=normal(repeat(0,1000));
if any(x18>3) |any(x18<-3) then go to lp;
finish a18;
run a18;
create abi18 from x18[colname={abili18}];
append from x18;
;
proc iml;
start a19;
lp: x19=normal(repeat(0,1000));
if any(x19>3) |any(x19<-3) then go to lp;
finish a19;
run a19;
create abi19 from x19[colname={abili19}];
append from x19;
;
proc iml;
```

```
start a20;
lp: x20=normal(repeat(0,1000));
if any(x20>3) |any(x20<-3) then go to lp;
finish a20;
run a20;
create abi20 from x20[colname={abili20}];
append from x20;
;
proc iml;
start a21;
lp: x21=normal(repeat(0,1000));
if any(x21>3) |any(x21<-3) then go to lp;
finish a21;
run a21;
create abi21 from x21[colname={abili21}];
append from x21;
;
proc iml;
start a22;
lp: x22=normal(repeat(0,1000));
if any(x22>3) |any(x22<-3) then go to lp;
finish a22;
run a22;
create abi22 from x22[colname={abili22}];
append from x22;
;
proc iml;
start a23;
lp: x23=normal(repeat(0,1000));
if any(x23>3) |any(x23<-3) then go to lp;
finish a23;
run a23;
create abi23 from x23[colname={abili23}];
append from x23;
;
proc iml;
start a24;
lp: x24=normal(repeat(0,1000));
if any(x24>3) |any(x24<-3) then go to lp;
finish a24;
run a24;
create abi24 from x24[colname={abili24}];
append from x24;
;
proc iml;
start a25;
```

```
lp: x25=normal(repeat(0,1000));
if any(x25>3) |any(x25<-3) then go to lp;
finish a25;
run a25;
create abi25 from x25[colname={abili25}];
append from x25;
;

libname abili1'c:\my documents\my sas files\abili1';
  data abili1.foc1;
  set abi1;
proc standard mean=0 std=1;
var abili1;
run;
;
  data abili1.foc2;
  set abi2;
proc standard mean=0 std=1;
var abili2;
run;
;
  data abili1.foc3;
  set abi3;
proc standard mean=0 std=1;
var abili3;
run;
;
  data abili1.foc4;
  set abi4;
proc standard mean=0 std=1;
var abili4;
run;
;
  data abili1.foc5;
  set abi5;
proc standard mean=0 std=1;
var abili5;
run;
;
  data abili1.foc6;
  set abi6;
proc standard mean=0 std=1;
var abili6;
run;
;
  data abili1.foc7;
```

```
    set abi7;
proc standard mean=0 std=1;
var abili7;
run;
    ;
  data abili1.foc8;
  set abi8;
proc standard mean=0 std=1;
var abili8;
run;
    ;
  data abili1.foc9;
  set abi9;
proc standard mean=0 std=1;
var abili9;
run;
    ;
  data abili1.foc10;
  set abi10;
proc standard mean=0 std=1;
var abili10;
run;
    ;
  data abili1.foc11;
  set abi11;
proc standard mean=0 std=1;
var abili11;
run;
    ;
  data abili1.foc12;
  set abi12;
proc standard mean=0 std=1;
var abili12;
run;
    ;
  data abili1.foc13;
  set abi13;
proc standard mean=0 std=1;
var abili13;
run;
    ;
  data abili1.foc14;
  set abi14;
proc standard mean=0 std=1;
var abili14;
run;
```

```
;
    data abili1.foc15;
    set abi15;
proc standard mean=0 std=1;
var abili15;
run;
;
    data abili1.foc16;
    set abi16;
proc standard mean=0 std=1;
var abili16;
run;
;
    data abili1.foc17;
    set abi17;
proc standard mean=0 std=1;
var abili17;
run;
;
    data abili1.foc18;
    set abi18;
proc standard mean=0 std=1;
var abili18;
run;
;
    data abili1.foc19;
    set abi19;
proc standard mean=0 std=1;
var abili19;
run;
;
    data abili1.foc20;
    set abi20;
proc standard mean=0 std=1;
var abili20;
run;
;
    data abili1.foc21;
    set abi21;
proc standard mean=0 std=1;
var abili21;
run;
;
    data abili1.foc22;
    set abi22;
proc standard mean=0 std=1;
```

```
var abili22;
run;
;
  data abili1.foc23;
  set abi23;
proc standard mean=0 std=1;
var abili23;
run;
;
  data abili1.foc24;
  set abi24;
proc standard mean=0 std=1;
var abili24;
run;
;
  data abili1.foc25;
  set abi25;
proc standard mean=0 std=1;
var abili25;
run;
;
```

SAS Program: Generation of a and b Parameters

```
proc iml;
reset noprint;
x={0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0};
do r=1 to 42;
do c=1 to 1;
lp:
x[r,c]=uniform(0)*3;
if abs(x[r,c])>2.0 then goto lp;
end;
end;
print x[format=6.4];
end;
proc iml;
start d;
lp:
x=normal(repeat(0,42));
if any(x>2.0) | any(x<-2) then goto lp;
```

```
finish d;
run d;
create dif1 from x [colname={x1}];
append from x;
proc standard out=bstd mean=0 std=1;
var x1;
run;
proc print;
var x1;
run;
```

SAS Program: Generation of Uniform Numbers

```
proc iml;        /* begin IML session*/
start g1;        /* begin module*/
step1:           /* create link statement in the case that the uniform number is 0 or 1 IML
                    will select the new uniform number*/

   x1=uniform(repeat(0,1000,50));   /* select the uniform number with seed number is 0,
                                        the number of observation is 1000, and the
                                        number of replication is 50 */
   if x1=0 | x1=1 then go to step1;

finish g1;       /* end module*/
run g1;
create uni1 from x1[colname={u1 u2 u3 u4 u5 u6 u7 u8 u9 u10 u11 u12 u13 u14 u15 u16
                    u17 u18 u19 u20 u21 u22 u23 u24 u25 u26 u27 u28 u29
                    u30 u31 u32 u33 u34 u35 u36 u37 u38 u39 u40 u41 u42
                    u43 u44 u45 u46 u47 u48 u49 u50}];
   append from x1;

/* create the dataset 'uni1' from matrix 'x1'. Colname define the name of variables in
   dataset 'uni1' */

proc iml;
start g2;
step2:
  x2=uniform(repeat(0,1000,50));
  if x2=0 | x2=1 then go to step2;
finish g2;
run g2;
create uni2 from x2[colname={u1 u2 u3 u4 u5 u6 u7 u8 u9 u10 u11 u12 u13 u14 u15 u16
                    u17 u18 u19 u20 u21 u22 u23 u24 u25 u26 u27 u28 u29
                    u30 u31 u32 u33 u34 u35 u36 u37 u38 u39 u40 u41 u42
                    u43 u44 u45 u46 u47 u48 u49 u50}];
   append from x2;
```

```
proc iml;
start g3;
step3:
  x3=uniform(repeat(0,1000,50));
  if x3=0 | x3=1 then go to step3;
finish g3;
run g3;
  create uni3 from x3[colname={u1 u2 u3 u4 u5 u6 u7 u8 u9 u10 u11 u12 u13 u14 u15
                              u16 u17 u18 u19 u20 u21 u22 u23 u24 u25 u26 u27 u28
                              u29 u30 u31 u32 u33 u34 u35 u36 u37 u38 u39 u40 u41
                              u42  u43 u44 u45 u46 u47 u48 u49 u50}];
  append from x3;

proc iml;
start g4;
step4:

  x4=uniform(repeat(0,1000,50));
  if x4=0 | x4=1 then go to step4;
finish g4;
run g4;
  create uni4 from x4[colname={u1 u2 u3 u4 u5 u6 u7 u8 u9 u10 u11 u12 u13 u14 u15
                              u16 u17 u18 u19 u20 u21 u22 u23 u24 u25 u26 u27 u28
                              u29 u30 u31 u32 u33 u34 u35 u36 u37 u38 u39 u40 u41
                              u42 u43 u44 u45 u46 u47 u48 u49 u50}];
  append from x4;

proc iml;
start g5;
step5:
  x5=uniform(repeat(0,1000,50));
  if x5=0 | x5=1 then go to step5;
finish g5;
run g5;
  create uni5 from x5[colname={u1 u2 u3 u4 u5 u6 u7 u8 u9 u10 u11 u12 u13 u14 u15
                              u16 u17 u18 u19 u20 u21 u22 u23 u24 u25 u26 u27 u28
                              u29 u30 u31 u32 u33 u34 u35 u36 u37 u38 u39 u40 u41
                              u42 u43 u44 u45 u46 u47 u48 u49 u50}];
  append from x5;

proc iml;
start g6;
step6:
  x6=uniform(repeat(0,1000,50));
  if x6=0 | x6=1 then go to step6;
```

```
finish g6;
run g6;
  create uni6 from x6[colname={u1 u2 u3 u4 u5 u6 u7 u8 u9 u10 u11 u12 u13 u14 u15
                             u16 u17 u18 u19 u20 u21 u22 u23 u24 u25 u26 u27 u28
                             u29 u30 u31 u32 u33 u34 u35 u36 u37 u38 u39 u40 u41
                             u42 u43 u44 u45 u46 u47 u48 u49 u50}];
  append from x6;


proc iml;
start g7;
step7:
  x7=uniform(repeat(0,1000,50));
  if x7=0 |x7=1 then go to step7;
finish g7;
run g7;
  create uni7 from x7[colname={u1 u2 u3 u4 u5 u6 u7 u8 u9 u10 u11 u12 u13 u14 u15
                             u16 u17 u18 u19 u20 u21 u22 u23 u24 u25 u26 u27 u28
                             u29 u30 u31 u32 u33 u34 u35 u36 u37 u38 u39 u40 u41
                             u42 u43 u44 u45 u46 u47 u48 u49 u50}];
  append from x7;


proc iml;
start g8;
step8:
  x8=uniform(repeat(0,1000,50));
  if x8=0 | x8=1 then go to step8;
finish g8;
run g8;
  create uni8 from x8[colname={u1 u2 u3 u4 u5 u6 u7 u8 u9 u10 u11 u12 u13 u14 u15
                             u16 u17 u18 u19 u20 u21 u22 u23 u24 u25 u26 u27 u28
                             u29 u30 u31 u32 u33 u34 u35 u36 u37 u38 u39 u40 u41
                             u42 u43 u44 u45 u46 u47 u48 u49 u50}];
  append from x8;


proc iml;
start g9;
step9:
  x9=uniform(repeat(0,1000,50));
  if x9=0 | x9=1 then go to step9;
finish g9;
run g9;
  create uni9 from x9[colname={u1 u2 u3 u4 u5 u6 u7 u8 u9 u10 u11 u12 u13 u14 u15
                             u16 u17 u18 u19 u20 u21 u22 u23 u24 u25 u26 u27 u28
                             u29 u30 u31 u32 u33 u34 u35 u36 u37 u38 u39 u40 u41
                             u42 u43 u44 u45 u46 u47 u48 u49 u50}];
  append from x9;
```

```
proc iml;
start g10;
step10:
   x10=uniform(repeat(0,1000,50));
   if x10=0 | x10=1 then go to step10;
finish g10;
run g10;
   create uni10 from x10[colname={u1 u2 u3 u4 u5 u6 u7 u8 u9 u10 u11 u12 u13 u14 u15
                               u16 u17 u18 u19 u20 u21 u22 u23 u24 u25 u26 u27 u28
                               u29 u30 u31 u32 u33 u34 u35 u36 u37 u38 u39 u40 u41
                               u42 u43 u44 u45 u46 u47 u48 u49 u50}];
   append from x10;

proc iml;
start g11;
step11:
   x11=uniform(repeat(0,1000,50));
   if x11=0 | x11=1 then go to step11;
finish g11;
run g11;
   create uni11 from x11[colname={u1 u2 u3 u4 u5 u6 u7 u8 u9 u10 u11 u12 u13 u14 u15
                               u16 u17 u18 u19 u20 u21 u22 u23 u24 u25 u26 u27 u28
                               u29 u30 u31 u32 u33 u34 u35 u36 u37 u38 u39 u40 u41
                               u42 u43 u44 u45 u46 u47 u48 u49 u50}];
   append from x11;

proc iml;
start g12;
step12:
   x12=uniform(repeat(0,1000,50));
   if x12=0 | x12=1 then go to step12;
finish g12;
run g12;
   create uni12 from x12[colname={u1 u2 u3 u4 u5 u6 u7 u8 u9 u10 u11 u12 u13 u14 u15
                               u16 u17 u18 u19 u20 u21 u22 u23 u24 u25 u26 u27 u28
                               u29 u30 u31 u32 u33 u34 u35 u36 u37 u38 u39 u40 u41
                               u42 u43 u44 u45 u46 u47 u48 u49 u50}];
   append from x12;

proc iml;
start g13;
step13:
   x13=uniform(repeat(0,1000,50));
   if x13=0 | x13=1 then go to step13;
finish g13;
```

```
run g13;
    create uni13 from x13[colname={u1 u2 u3 u4 u5 u6 u7 u8 u9 u10 u11 u12 u13 u14 u15
                                   u16 u17 u18 u19 u20 u21 u22 u23 u24 u25 u26 u27 u28
                                   u29 u30 u31 u32 u33 u34 u35 u36 u37 u38 u39 u40 u41
                                   u42 u43 u44 u45 u46 u47 u48 u49 u50}];
    append from x13;

proc iml;
start g14;
step14:
   x14=uniform(repeat(0,1000,50));
   if x14=0 | x14=1 then go to step14;
finish g14;
run g14;
    create uni14 from x14[colname={u1 u2 u3 u4 u5 u6 u7 u8 u9 u10 u11 u12 u13 u14 u15
                                   u16 u17 u18 u19 u20 u21 u22 u23 u24 u25 u26 u27 u28
                                   u29 u30 u31 u32 u33 u34 u35 u36 u37 u38 u39 u40 u41
                                   u42 u43 u44 u45 u46 u47 u48 u49 u50}];
    append from x14;

proc iml;
start g15;
step15:
   x15=uniform(repeat(0,1000,50));
   if x15=0 | x15=1 then go to step15;
finish g15;
run g15;
    create uni15 from x15[colname={u1 u2 u3 u4 u5 u6 u7 u8 u9 u10 u11 u12 u13 u14 u15
                                   u16 u17 u18 u19 u20 u21 u22 u23 u24 u25 u26 u27 u28
                                   u29 u30 u31 u32 u33 u34 u35 u36 u37 u38 u39 u40 u41
                                   u42 u43 u44 u45 u46 u47 u48 u49 u50}];
    append from x15;

proc iml;
start g16;
step16:
   x16=uniform(repeat(0,1000,50));
   if x16=0 | x16=1 then go to step16;
finish g16;
run g16;
    create uni16 from x16[colname={u1 u2 u3 u4 u5 u6 u7 u8 u9 u10 u11 u12 u13 u14 u15
                                   u16 u17 u18 u19 u20 u21 u22 u23 u24 u25 u26 u27 u28
                                   u29 u30 u31 u32 u33 u34 u35 u36 u37 u38 u39 u40 u41
                                   u42 u43 u44 u45 u46 u47 u48 u49 u50}];
    append from x16;
```

```
proc iml;
start g17;
step17:
  x17=uniform(repeat(0,1000,50));
  if x17=0 | x17=1 then go to step17;
finish g17;
run g17;
  create uni17 from x17[colname={u1 u2 u3 u4 u5 u6 u7 u8 u9 u10 u11 u12 u13 u14 u15
                      u16 u17 u18 u19 u20 u21 u22 u23 u24 u25 u26 u27 u28
                      u29 u30 u31 u32 u33 u34 u35 u36 u37 u38 u39 u40 u41
                      u42 u43 u44 u45 u46 u47 u48 u49 u50}];
  append from x17;

proc iml;
start g18;
step18:
  x18=uniform(repeat(0,1000,50));
  if x18=0 | x18=1 then go to step18;
finish g18;
run g18;
  create uni18 from x18[colname={u1 u2 u3 u4 u5 u6 u7 u8 u9 u10 u11 u12 u13 u14 u15
                      u16 u17 u18 u19 u20 u21 u22 u23 u24 u25 u26 u27 u28
                      u29 u30 u31 u32 u33 u34 u35 u36 u37 u38 u39 u40 u41
                      u42 u43 u44 u45 u46 u47 u48 u49 u50}];
  append from x18;

proc iml;
start g19;
step19:
  x19=uniform(repeat(0,1000,50));
  if x19=0 | x19=1 then go to step19;
finish g19;
run g19;
  create uni19 from x19[colname={u1 u2 u3 u4 u5 u6 u7 u8 u9 u10 u11 u12 u13 u14 u15
                      u16 u17 u18 u19 u20 u21 u22 u23 u24 u25 u26 u27 u28
                      u29 u30 u31 u32 u33 u34 u35 u36 u37 u38 u39 u40 u41
                      u42 u43 u44 u45 u46 u47 u48 u49 u50}];
  append from x19;

proc iml;
start g20;
step20:
  x20=uniform(repeat(0,1000,50));
  if x20=0 | x20=1 then go to step20;
finish g20;
run g20;
```

```
create uni20 from x20[colname={ul u2 u3 u4 u5 u6 u7 u8 u9 u10 u11 u12 u13 u14 u15
                                u16 u17 u18 u19 u20 u21 u22 u23 u24 u25 u26 u27 u28
                                u29 u30 u31 u32 u33 u34 u35 u36 u37 u38 u39 u40 u41
                                u42 u43 u44 u45 u46 u47 u48 u49 u50}];
append from x20;

proc iml;
start g21;
step21:
  x21=uniform(repeat(0,1000,50));
  if x21=0 | x21=1 then go to step21;
finish g21;
run g21;
  create uni21 from x21[colname={ul u2 u3 u4 u5 u6 u7 u8 u9 u10 u11 u12 u13 u14 u15
                                u16 u17 u18 u19 u20 u21 u22 u23 u24 u25 u26 u27 u28
                                u29 u30 u31 u32 u33 u34 u35 u36 u37 u38 u39 u40 u41
                                u42 u43 u44 u45 u46 u47 u48 u49 u50}];
append from x21;

proc iml;
start g22;
step22:
  x22=uniform(repeat(0,1000,50));
  if x22=0 | x22=1 then go to step22;
finish g22;
run g22;
  create uni22 from x22[colname={ul u2 u3 u4 u5 u6 u7 u8 u9 u10 u11 u12 u13 u14 u15
                                u16 u17 u18 u19 u20 u21 u22 u23 u24 u25 u26 u27 u28
                                u29 u30 u31 u32 u33 u34 u35 u36 u37 u38 u39 u40 u41
                                u42 u43 u44 u45 u46 u47 u48 u49 u50}];
append from x22;

proc iml;
start g23;
step23:
  x23=uniform(repeat(0,1000,50));
  if x23=0 | x23=1 then go to step23;
finish g23;
run g23;
  create uni23 from x23[colname={ul u2 u3 u4 u5 u6 u7 u8 u9 u10 u11 u12 u13 u14 u15
                                u16 u17 u18 u19 u20 u21 u22 u23 u24 u25 u26 u27 u28
                                u29 u30 u31 u32 u33 u34 u35 u36 u37 u38 u39 u40 u41
                                u42 u43 u44 u45 u46 u47 u48 u49 u50}];
append from x23;

proc iml;
```

```
start g24;
step24:
  x24=uniform(repeat(0,1000,50));
  if x24=0 | x24=1 then go to step24;
finish g24;
run g24;
  create uni24 from x24[colname={u1 u2 u3 u4 u5 u6 u7 u8 u9 u10 u11 u12 u13 u14 u15
                                 u16 u17 u18 u19 u20 u21 u22 u23 u24 u25 u26 u27 u28
                                 u29 u30 u31 u32 u33 u34 u35 u36 u37 u38 u39 u40 u41
                                 u42 u43 u44 u45 u46 u47 u48 u49 u50}];
  append from x24;

proc iml;
start g25;
step25:
  x25=uniform(repeat(0,1000,50));
  if x25=0 | x25=1 then go to step25;
finish g25;
run g25;
  create uni25 from x25[colname={u1 u2 u3 u4 u5 u6 u7 u8 u9 u10 u11 u12 u13 u14 u15
                                 u16 u17 u18 u19 u20 u21 u22 u23 u24 u25 u26 u27 u28
                                 u29 u30 u31 u32 u33 u34 u35 u36 u37 u38 u39 u40 u41
                                 u42 u43 u44 u45 u46 u47 u48 u49 u50}];
  append from x25;

libname unif12 'c:\my documents\my sas files\unif12'; /* unif12= uniform number for
reference group unidimensional dichotomous */
  data unif12.uni1;  /* unif12 has 25 set of 1000 rows 50 column for unidimensional
dichotomous reference group */
       set uni1;
run;
  data unif12.uni2;
       set uni2;
run;
  data unif12.uni3;
       set uni3;
run;
  data unif12.uni4;
       set uni4;
run;
  data unif12.uni5;
       set uni5;
run;
  data unif12.uni6;
       set uni6;
run;
```

```
    data unif12.uni7;
        set uni7;
run;
    data unif12.uni8;
        set uni8;
run;
    data unif12.uni9;
        set uni9;
run;
    data unif12.uni10;
        set uni10;
run;
    data unif12.uni11;
        set uni11;
run;
    data unif12.uni12;
        set uni12;
run;
    data unif12.uni13;
        set uni13;
run;
    data  unif12.uni14;
        set uni14;
run;
    data unif12.uni15;
        set uni15;
run;
    data unif12.uni16;
        set uni16;
run;
    data unif12.uni17;
        set uni17;
run;
    data unif12.uni18;
        set uni18;
run;
    data unif12.uni19;
        set uni19;
run;
    data unif12.uni20;
        set uni20;
run;
    data unif12.uni21;
        set uni21;
run;
    data unif12.uni22;
```

```
        set uni22;
run;
   data unif12.uni23;
        set uni23;
run;
   data unif12.uni24;
        set uni24;
run;
   data unif12.uni25;
        set uni25;
run;
```

SAS Program: Generation of Item Responses

Title 'this program is saved as unidi1.sas';
libname unidir1 'c:\my documents\my sas files\unidir1'; /* library name 'unidiref is created to keep the data from output of this program */
libname unif1 'c:\my documents\my sas files\unif1'; /* unif1 1 is the uniform number matrix for unidimensional dichotomous reference group */
libname abili1 'c:\my documents\my sas files\abili1'; /* library name 'ability' have ability parameter for reference group and focal group */
     data unidir1.group1; /* unidiref.group1 is created to keep the data of reference group 1 */
        set abili1.ref1; /* ability.ref1 contains 1000 ability parameter, variable in ref1 and foc1 is ability1 and so on */
        set unif1.uni1; /* unif1 is for unidimensional dichotomous data there are unif1.uni1-unif1.uni25 for each set of replication */

```
a1=    .5001    ;    b1=    -1.500;
a2=    .5001    ;    b2=    0      ;
a3=    .5001    ;    b3=    2.000 ;
a4=    1.000    ;    b4=    -1.500;
a5=    1.000    ;    b5=    0      ;
a6=    1.000    ;    b6=    2.000 ;
a7=    2.000    ;    b7=    -1.500;
a8=    2.000    ;    b8=    0      ;
a9=    2.000    ;    b9=    2.000 ;
a10=   0.5584 ;      b10=   -0.7024 ;
a11=   1.573  ;      b11=   -1.1617 ;
a12=   0.5724 ;      b12=   -0.5478 ;
a13=   1.4023 ;      b13=   0.3206  ;
a14=   1.8635 ;      b14=   -1.116  ;
a15=   0.9819 ;      b15=   0.645   ;
a16=   0.4035 ;      b16=   -1.1824 ;
a17=   0.5464 ;      b17=   -0.1218 ;
a18=   1.0219 ;      b18=   -0.4656 ;
```

```
a19=   0.4716 ;      b19=   -0.7483 ;
a20=   0.2022 ;      b20=   0.4394 ;
a21=   0.3338 ;      b21=   -1.7195 ;
a22=   1.8987 ;      b22=   -1.1037 ;
a23=   0.9989 ;      b23=   -0.2489 ;
a24=   0.2139 ;      b24=   1.1192 ;
a25=   1.1698 ;      b25=   0.15   ;
a26=   0.4043 ;      b26=   0.1201 ;
a27=   0.7342 ;      b27=   -0.4323 ;
a28=   1.3684 ;      b28=   0.9613 ;
a29=   1.92   ;      b29=   -0.3927 ;
a30=   1.1844 ;      b30=   -1.4364 ;
a31=   1.095  ;      b31=   -0.6721 ;
a32=   0.5358 ;      b32=   1.4818 ;
a33=   0.2297 ;      b33=   0.2384 ;
a34=   0.8075 ;      b34=   1.4778 ;
a35=   0.2126 ;      b35=   -0.1572 ;
a36=   1.2376 ;      b36=   -0.0476 ;
a37=   0.8673 ;      b37=   0.702  ;
a38=   0.2628 ;      b38=   0.5566 ;
a39=   0.8248 ;      b39=   -0.2334 ;
a40=   1.1579 ;      b40=   0.2141 ;
a41=   1.4277 ;      b41=   1.5473 ;
a42=   1.1839 ;      b42=   1.2567 ;
a43=   0.4158 ;      b43=   1.1952 ;
a44=   0.4981 ;      b44=   -0.786 ;
a45=   1.6642 ;      b45=   -1.7253 ;
a46=   0.3765 ;      b46=   -0.3399 ;
a47=   1.2371 ;      b47=   -0.542 ;
a48=   0.8993 ;      b48=   -0.7774 ;
a49=   1.477  ;      b49=   1.9439 ;
a50=   0.3341 ;      b50=   2.2909 ;

array a(50) a1-a50;
array b(50) b1-b50;
array p(50) p1-p50;
array u(50) u1-u50;
array item(50) item1-item50;
do i=1 to 50;
  p(i) = 1/(1+exp((-1.7*a(i))*(abili1-b(i))));  /* this is the two parameter model */
      if p(i) >= u(i) then item(i)=1;  /* compare the probability with the uniform number
to create measurement error */
            else item(i)=0;
      end;
 drop i;
run;
```

libname unidif1 'c:\my documents\my sas files\unidif1';
libname unif2 'c:\my documents\my sas files\unif2'; /*unif2 contains uni1-uni25 which is the uniform number for unidimensional dichotomous focal group*/
libname abili1 'c:\my documents\my sas files\abili1'; /*ability contains ref1-ref25 foc1-foc25 which is the ability parameter for each group i.e. ref1 and foc1 contain ability 1, ref2 and foc2 contain ability2*/
    data unidif1.group1;
        set abili1.foc1; /* foc1 contains only ability1 */
        set unif2.uni1; /* unif2 for unidimensional dichotomous focal group */

a1=	0.3830 ;	b1=	-1.700 ;	
a2=	0.3830 ;	b2=	-0.200 ;	
a3=	0.3830 ;	b3=	1.800 ;	
a4=	0.6200 ;	b4=	-1.700 ;	
a5=	0.6200 ;	b5=	-0.200 ;	
a6=	0.6200 ;	b6=	1.800 ;	
a7=	0.8990 ;	b7=	-1.700 ;	
a8=	0.8992 ;	b8=	-0.200 ;	
a9=	0.8990 ;	b9=	1.800 ;	
a10=	0.5584 ;	b10=	-0.7024 ;	
a11=	1.573 ;	b11=	-1.1617 ;	
a12=	0.5724 ;	b12=	-0.5478 ;	
a13=	1.4023 ;	b13=	0.3206 ;	
a14=	1.8635 ;	b14=	-1.116 ;	
a15=	0.9819 ;	b15=	0.645 ;	
a16=	0.4035 ;	b16=	-1.1824 ;	
a17=	0.5464 ;	b17=	-0.1218 ;	
a18=	1.0219 ;	b18=	-0.4656 ;	
a19=	0.4716 ;	b19=	-0.7483 ;	
a20=	0.2022 ;	b20=	0.4394 ;	
a21=	0.3338 ;	b21=	-1.7195 ;	
a22=	1.8987 ;	b22=	-1.1037 ;	
a23=	0.9989 ;	b23=	-0.2489 ;	
a24=	0.2139 ;	b24=	1.1192 ;	
a25=	1.1698 ;	b25=	0.15 ;	
a26=	0.4043 ;	b26=	0.1201 ;	
a27=	0.7342 ;	b27=	-0.4323 ;	
a28=	1.3684 ;	b28=	0.9613 ;	
a29=	1.92 ;	b29=	-0.3927 ;	
a30=	1.1844 ;	b30=	-1.4364 ;	
a31=	1.095 ;	b31=	-0.6721 ;	
a32=	0.5358 ;	b32=	1.4818 ;	
a33=	0.2297 ;	b33=	0.2384 ;	
a34=	0.8075 ;	b34=	1.4778 ;	
a35=	0.2126 ;	b35=	-0.1572 ;	

```
a36=   1.2376 ;     b36=   -0.0476 ;
a37=   0.8673 ;     b37=   0.702  ;
a38=   0.2628 ;     b38=   0.5566 ;
a39=   0.8248 ;     b39=   -0.2334 ;
a40=   1.1579 ;     b40=   0.2141 ;
a41=   1.4277 ;     b41=   1.5473 ;
a42=   1.1839 ;     b42=   1.2567 ;
a43=   0.4158 ;     b43=   1.1952 ;
a44=   0.4981 ;     b44=   -0.786 ;
a45=   1.6642 ;     b45=   -1.7253 ;
a46=   0.3765 ;     b46=   -0.3399 ;
a47=   1.2371 ;     b47=   -0.542 ;
a48=   0.8993 ;     b48=   -0.7774 ;
a49=   1.477  ;     b49=   1.9439 ;
a50=   0.3341 ;     b50=   2.2909 ;


array a(50) a1-a50;
array b(50) b1-b50;
array p(50) p1-p50;
array u(50) u1-u50;
array item(50) item1-item50;
do i=1 to 50;
  p(i) = 1/(1+exp((-1.7*a(i))*(abili1-b(i))));
        if p(i) >= u(i) then item(i)=1;
             else item(i)=0;
        end;
 drop i;
run;


libname itemud 'c:\my documents\my sas files\itemud';/* itemud=item of unidimensional
dichotomous*/
data itemud.rep1;
        set unidir1.group1(in=ref1 keep=item1-item50 drop=a1-a50 b1-b50 u1-u50 p1-p50
abili1)
        unidif1.group1(in=foc1 keep=item1-item50 drop=a1-a50 b1-b50 u1-u50 p1-p50
abili1);

if ref1 then group='01';
   else if foc1 then group='02';
proc print;
   var group item1-item50;
run;
```

APPENDIX B

ICCs of Items Showing DIF

Figure B.1. Parameter values and ICCs of item 1 for *a* DIF

Item Parameter Values: Low *a* (discrimination) and Low *b* (difficulty)

	a parameter	*b* parameter
Reference group	0.5001	-1.5000
Focal group	0.3830	-1.5000

Item 1 for *a* DIF
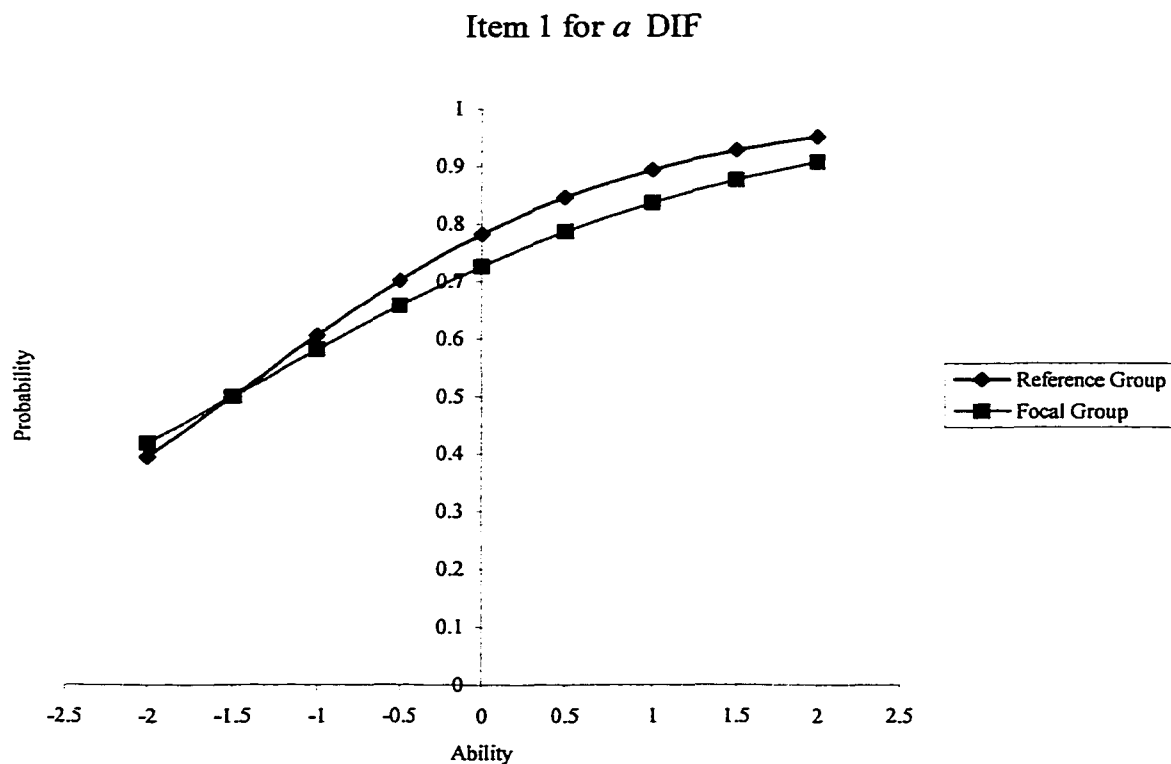
Figure B.2. Parameter values and ICCs of item 2 for *a* DIF

Item Parameter Values: Low *a* (discrimination) and medium *b* (difficulty)

	a parameter	*b* parameter
Reference group	0.5001	0.0000
Focal group	0.3830	0.0000

Item 2 for *a* DIF
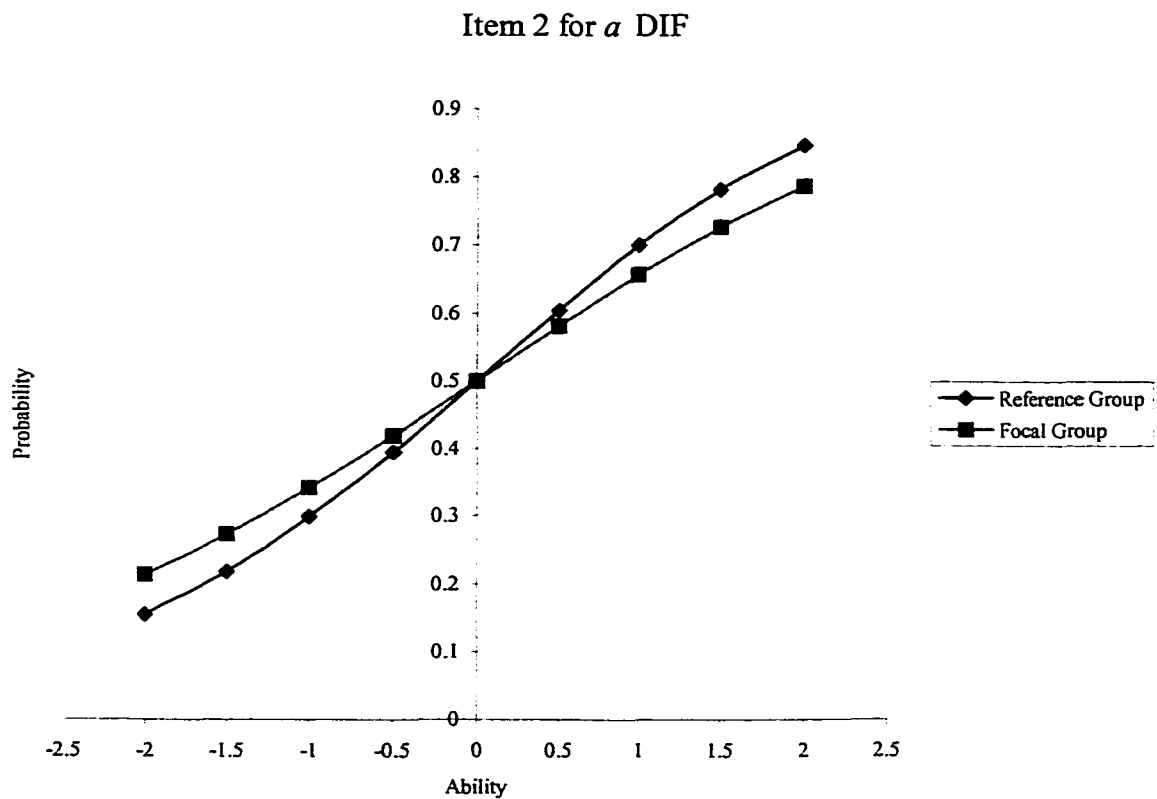
Figure B.3. Parameter values and ICCs of item 3 for a DIF

Item Parameter Values: Low a (discrimination) and high b (difficulty)

	a parameter	b parameter
Reference group	0.5001	2.0000
Focal group	0.3830	2.0000

Item 3 for a DIF
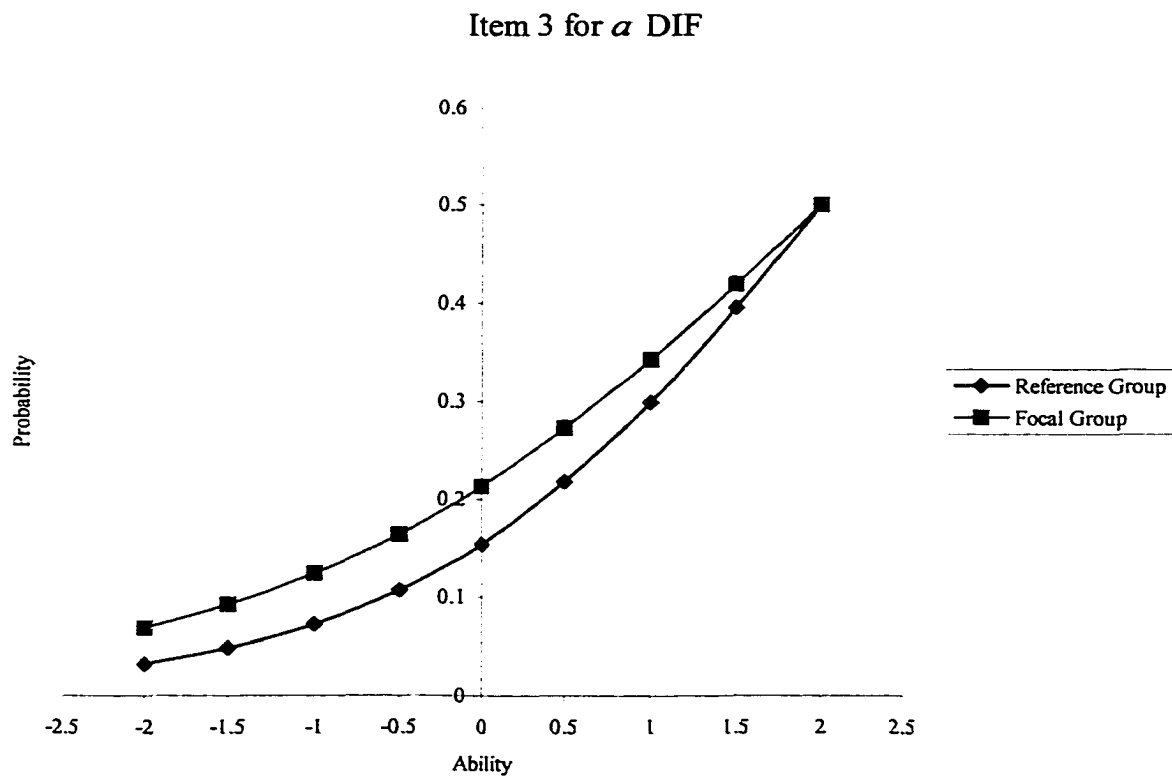
Figure B.4. Parameter values and ICCs of item 4 for *a* DIF

Item Parameter Values: medium *a* (discrimination) and low *b* (difficulty)

	a parameter	*b* parameter
Reference group	1.0000	-1.5000
Focal group	0.6200	-1.5000

Item 4 for *a* DIF
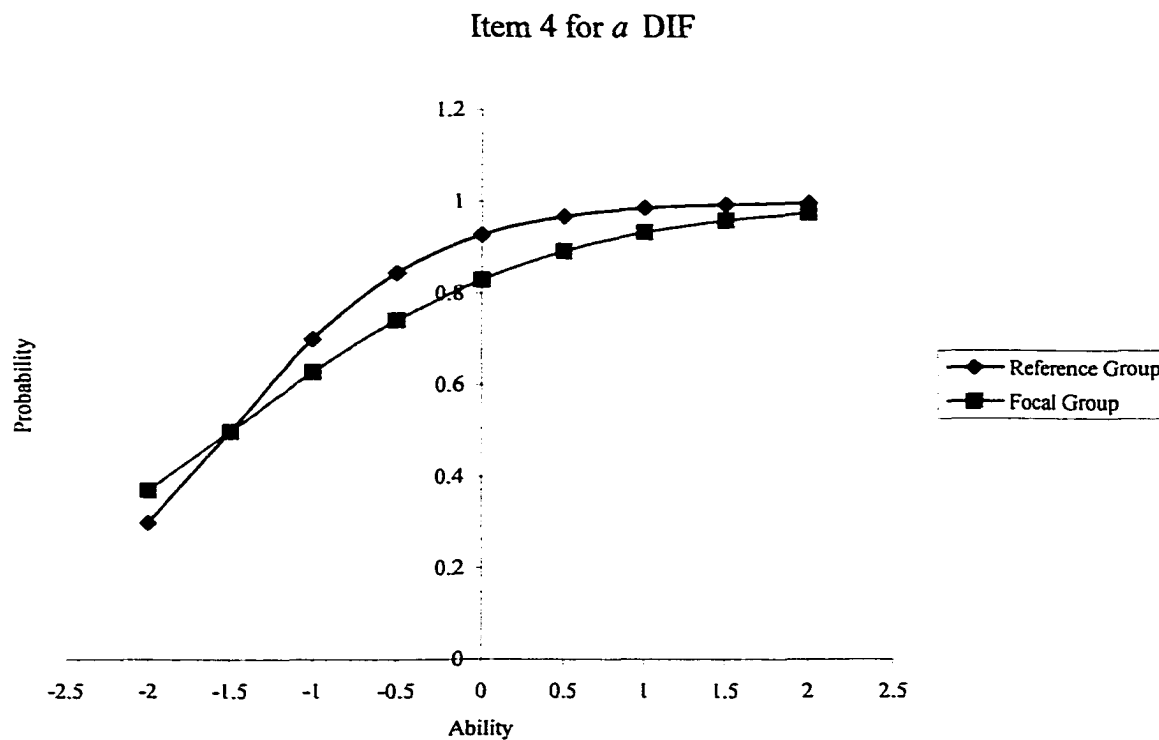
Figure B.5. Parameter values and ICCs of item 5 for *a* DIF

Item Parameter Values: medium *a* (discrimination) and medium *b* (difficulty)

	a parameter	*b* parameter
Reference group	1.0000	0.0000
Focal group	0.6200	0.0000

Item 5 for *a* DIF
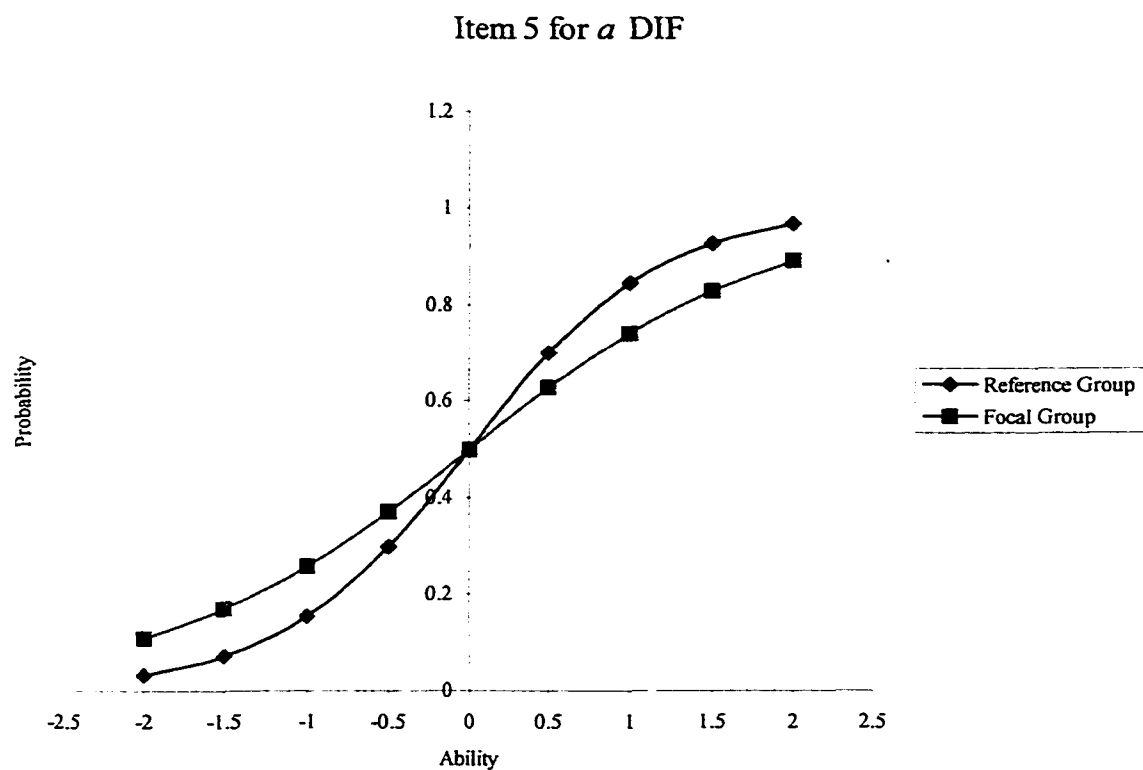
Figure B.6. Parameter values and ICCs of item 6 for *a* DIF

Item Parameter Values: medium *a* (discrimination) and high *b* (difficulty)

	a parameter	*b* parameter
Reference group	1.0000	2.0000
Focal group	0.6200	2.0000
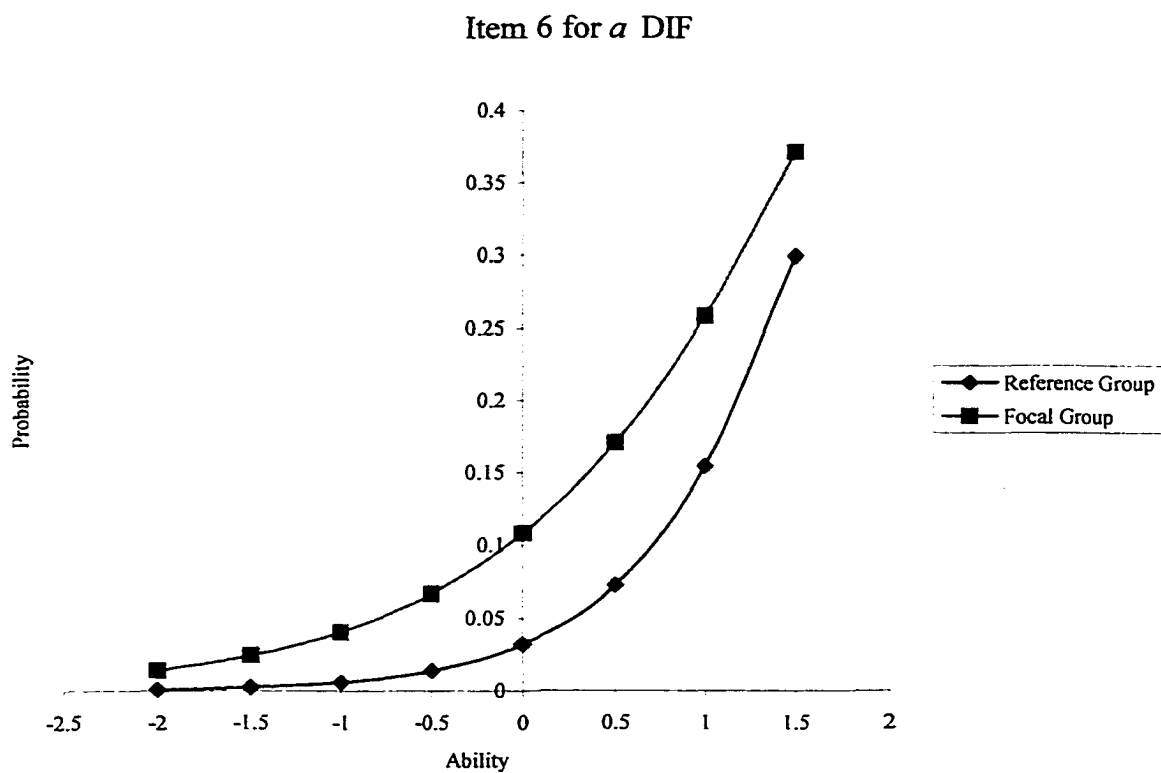
Item 6 for *a* DIF

Figure B.7. Parameter values and ICCs of item 7 for a DIF

Item Parameter Values: high a (discrimination) and low b (difficulty)

	a parameter	b parameter
Reference group	2.0000	-1.5000
Focal group	0.8990	-1.5000

Item 7 for a DIF

Figure B.7. Parameter values and ICCs of item 7 for a DIF

Item Parameter Values: high a (discrimination) and low b (difficulty)

	a parameter	b parameter
Reference group	2.0000	-1.5000
Focal group	0.8990	-1.5000

Item 7 for a DIF
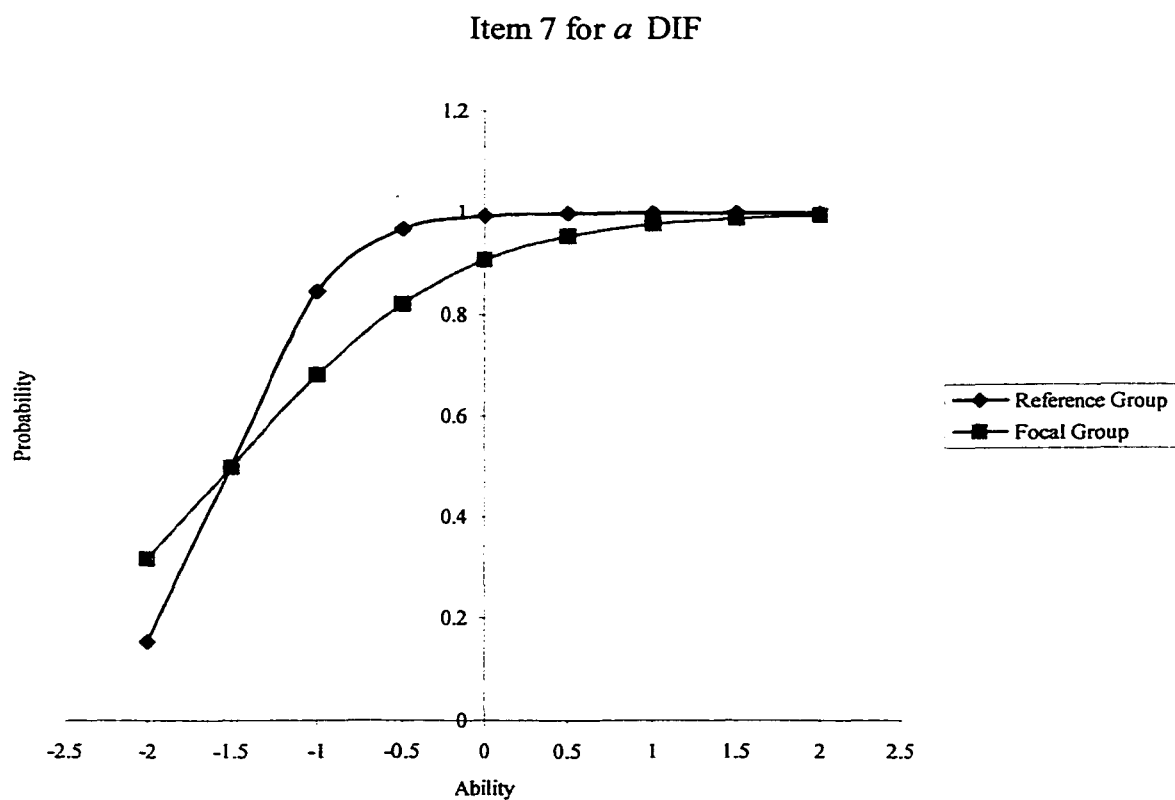
Figure B.8. Parameter values and ICCs of item 8 for *a* DIF

Item Parameter Values: high *a* (discrimination) and medium *b* (difficulty)

	a parameter	*b* parameter
Reference group	2.0000	0.0000
Focal group	0.8990	0.0000

Item 8 for *a* DIF
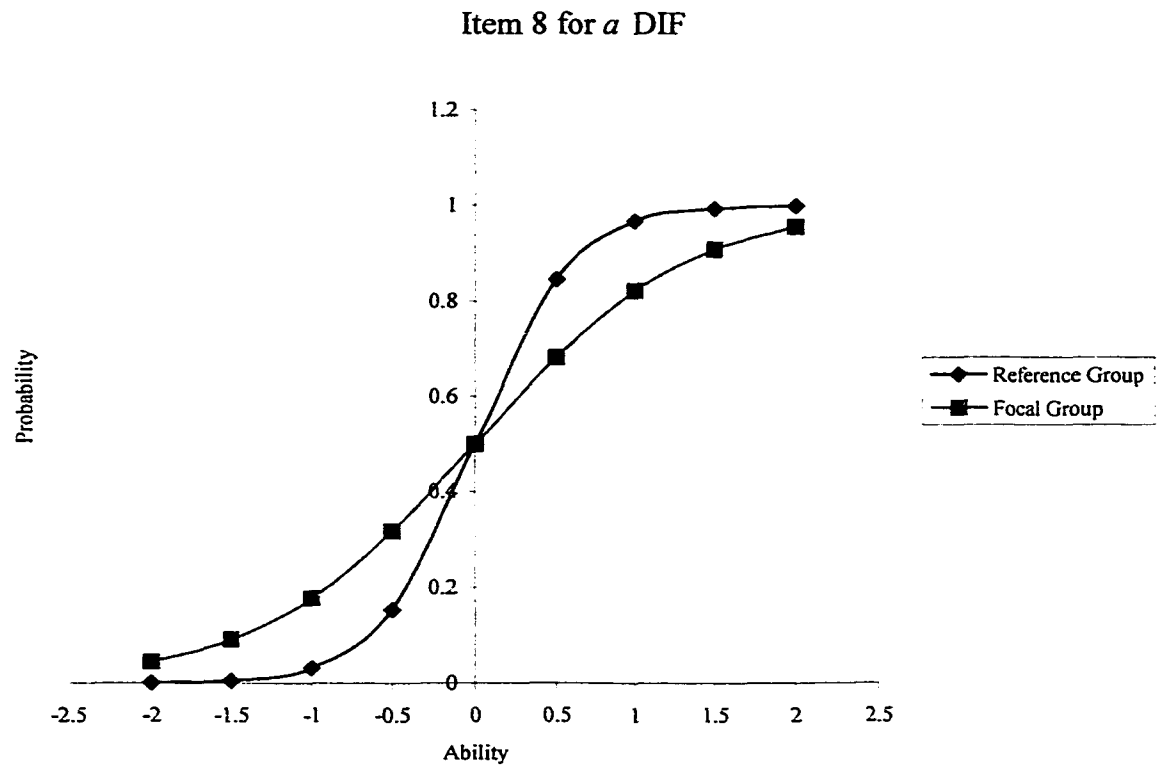
Figure B.9. Parameter values and ICCs of item 9 for *a* DIF

Item Parameters Values: high *a* (discrimination) and high *b* (difficulty)

	a parameter	*b* parameter
Reference group	2.0000	2.0000
Focal group	0.8990	2.0000

Item 9 for *a* DIF
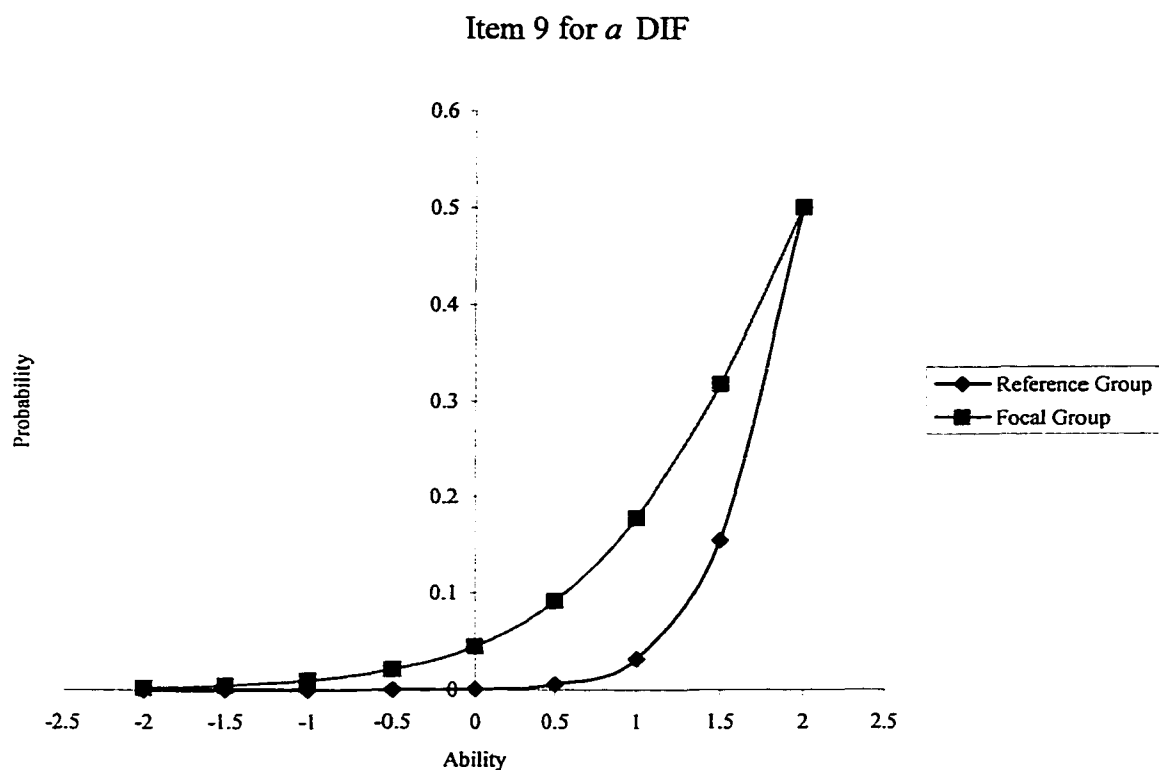
Figure B.10. Parameter values and ICCs of item 1 for *ab* DIF

Item Parameter Values: low *a* (discrimination) and low *b* (difficulty)

	a parameter	*b* parameter
Reference group	0.5001	-1.5000
Focal group	0.3880	-1.7000
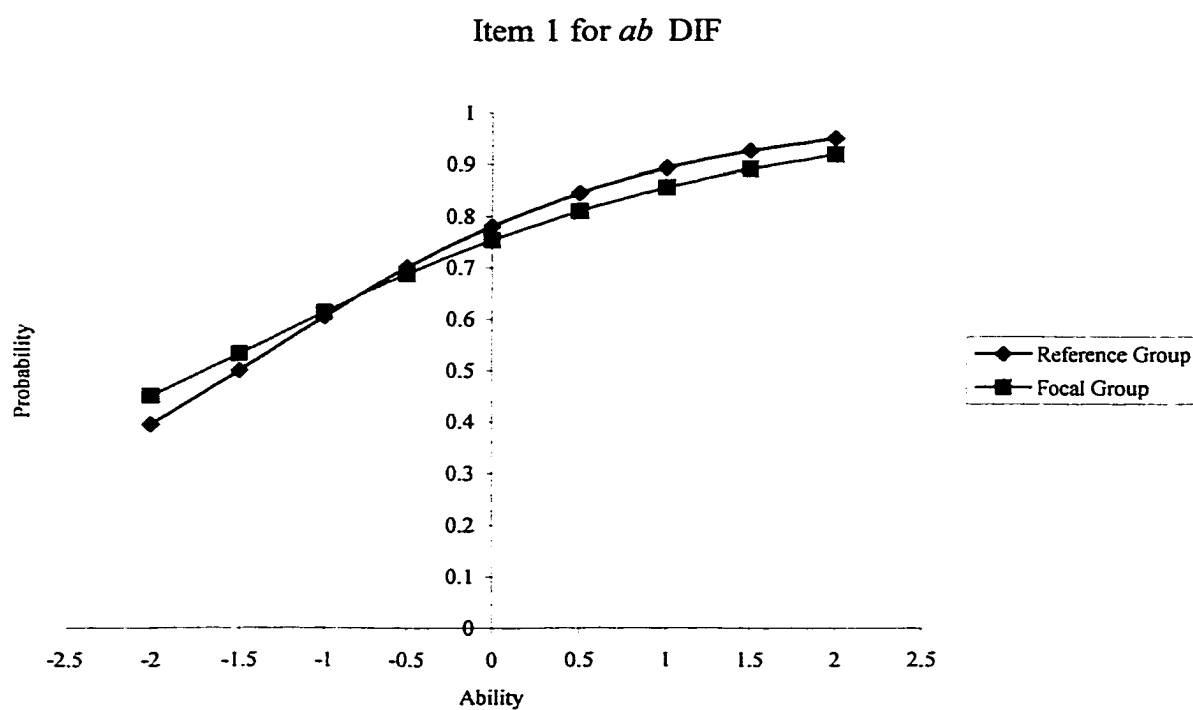
Item 1 for *ab* DIF

Figure B.11. Parameter values and ICCs of item 2 for *ab* DIF

Item Parameter Values: low *a* (discrimination) and medium *b* (difficulty)

	a parameter	*b* parameter
Reference group	0.5001	0.0000
Focal group	0.3880	-0.2000
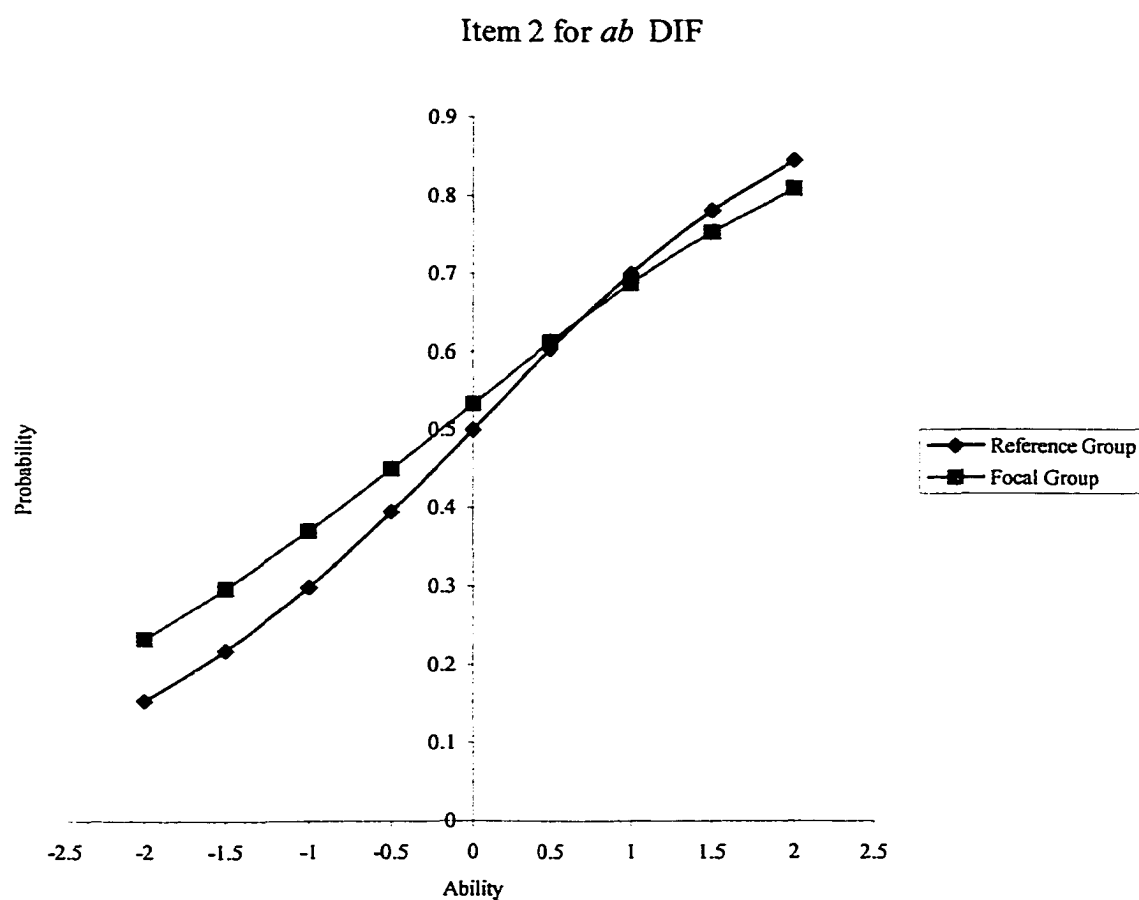
Item 2 for *ab* DIF

Figure B.12. Parameter values and ICCs of item 3 for *ab* DIF

Item Parameter Values: low *a* (discrimination) and high *b* (difficulty)

	a parameter	*b* parameter
Reference group	0.5001	2.0000
Focal group	0.3880	1.8000
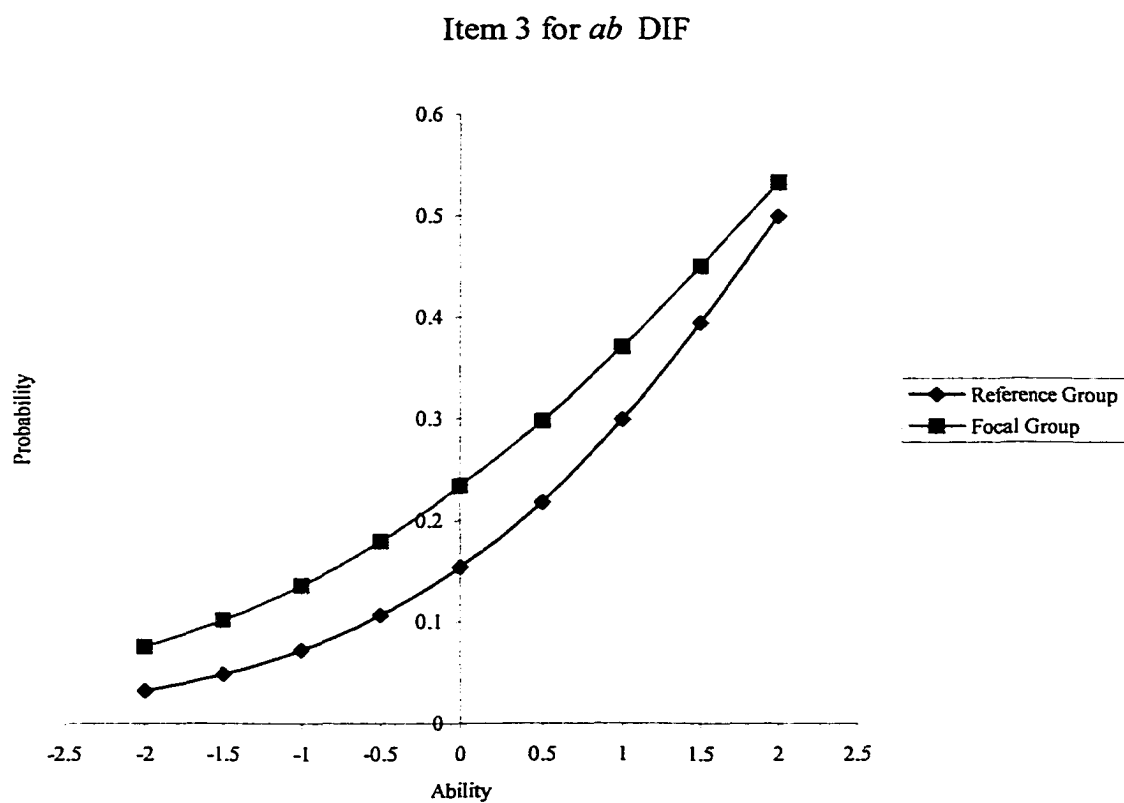
Item 3 for *ab* DIF

Figure B.13. Parameter values and ICCs of item 4 for *ab* DIF

Item Parameter Values: medium *a* (discrimination) and low *b* (difficulty)

	a parameter	*b* parameter
Reference group	1.0000	-1.5000
Focal group	0.6340	-1.7000
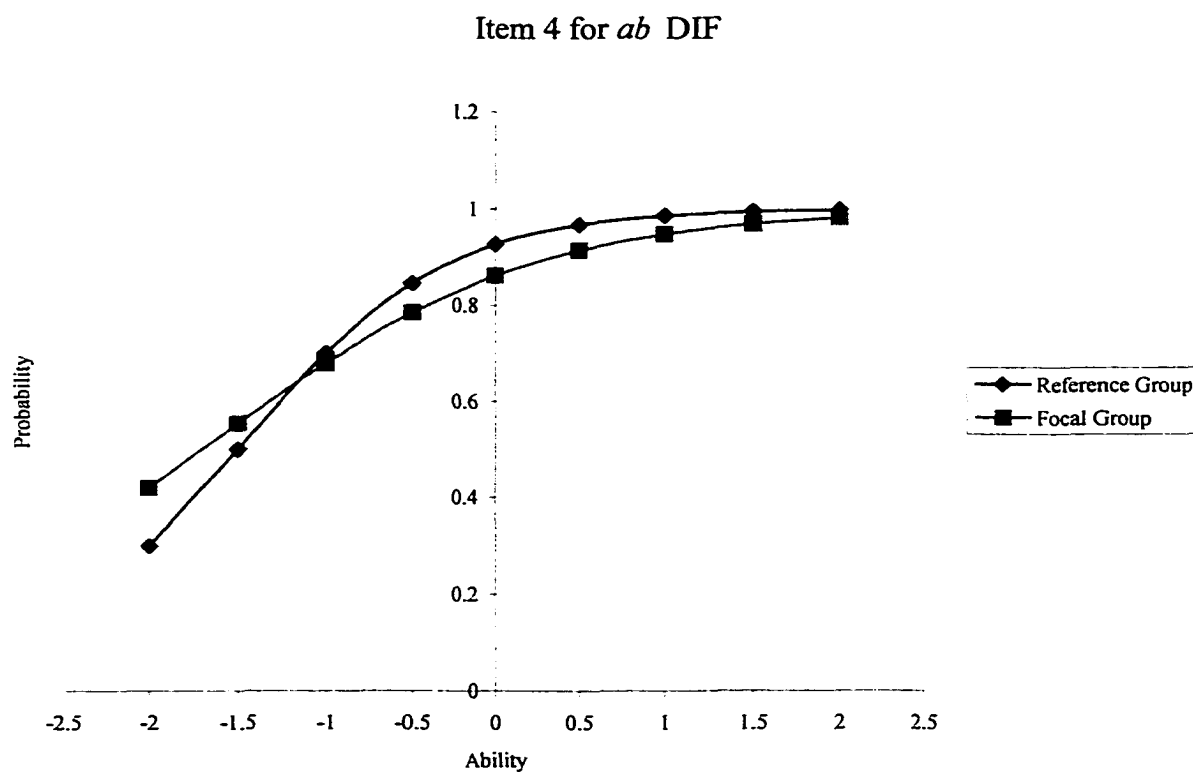
Item 4 for *ab* DIF

Figure B.14. Parameter values and ICCs of item 5 for *ab* DIF

Item Parameter Values: medium *a* (discrimination) and medium *b* (difficulty)

	a parameter	*b* parameter
Reference group	1.0000	0.0000
Focal group	0.6340	-0.2000
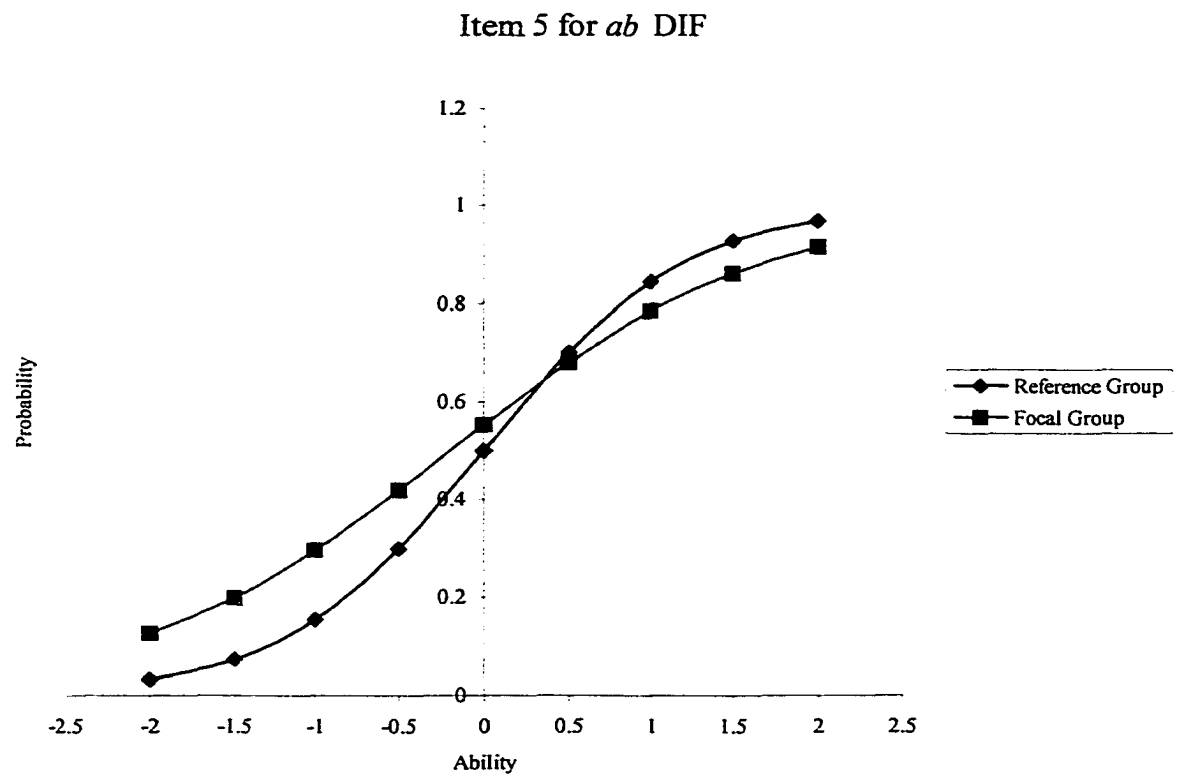
Item 5 for *ab* DIF

Figure B.15. Parameter values and ICCs of item 6 for *ab* DIF

Item Parameter Values: medium *a* (discrimination) and high *b* (difficulty)

	a parameter	*b* parameter
Reference group	1.0000	2.0000
Focal group	0.6340	1.8000
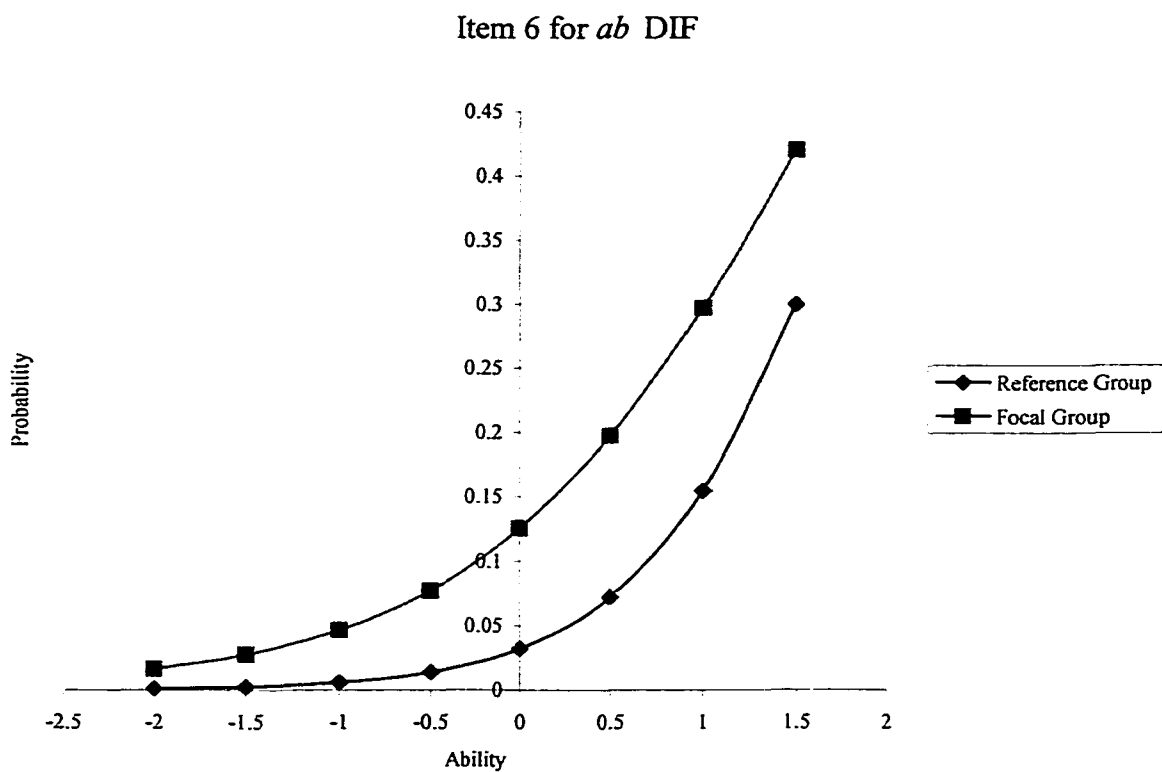
Item 6 for *ab* DIF

Figure B.16. Parameter values and ICCs of item 7 for *ab* DIF

Item Parameter Values: high *a* (discrimination) and low *b* (difficulty)

	a parameter	*b* parameter
Reference group	2.0000	-1.5000
Focal group	0.9280	-1.7000
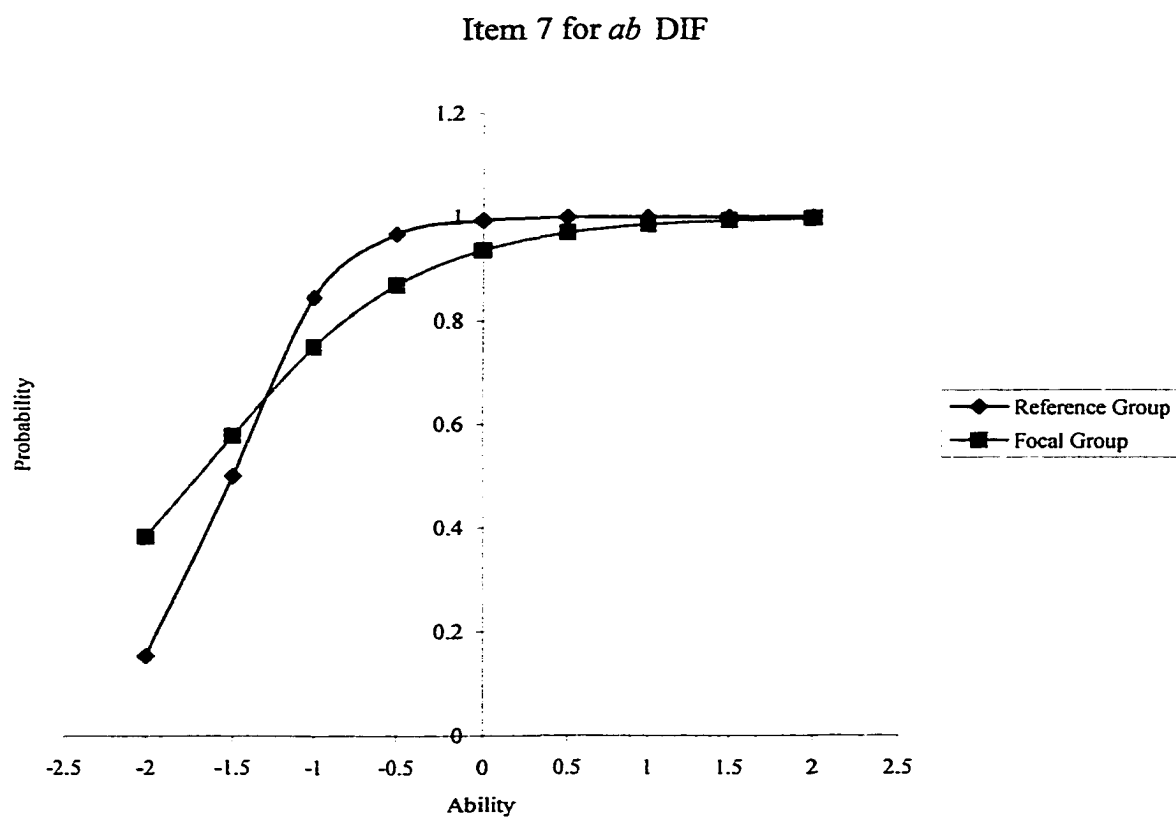
Item 7 for *ab* DIF

Figure B.17. Parameter values and ICCs of item 8 for *ab* DIF

Item Parameter Values: high *a* (discrimination) and medium *b* (difficulty)

	a parameter	*b* parameter
Reference group	2.0000	0.0000
Focal group	0.9280	-0.2000
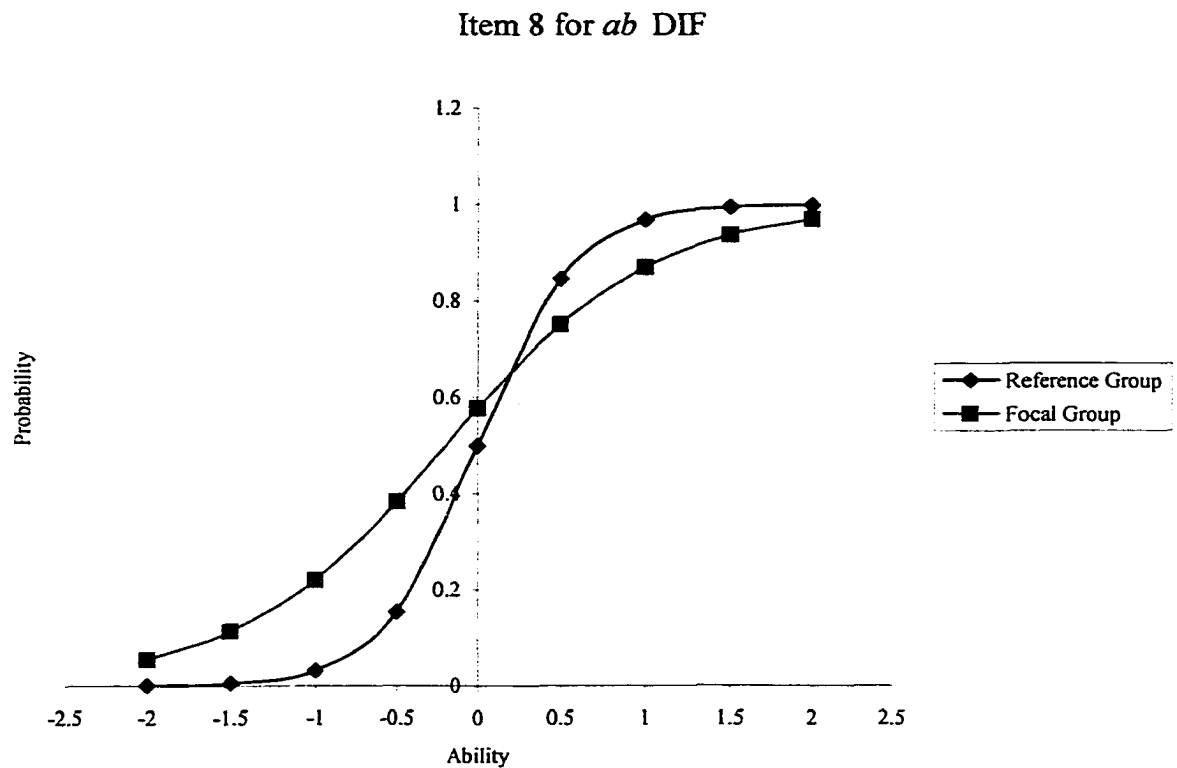
Item 8 for *ab* DIF

Figure B.18. Parameter values and ICCs of item 9 for *ab* DIF

Item Parameter Values: high *a* (discrimination) and high *b* (difficulty)

	a parameter	*b* parameter
Reference group	2.0000	2.0000
Focal group	0.9280	1.8000

Item 9 for *ab* DIF

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
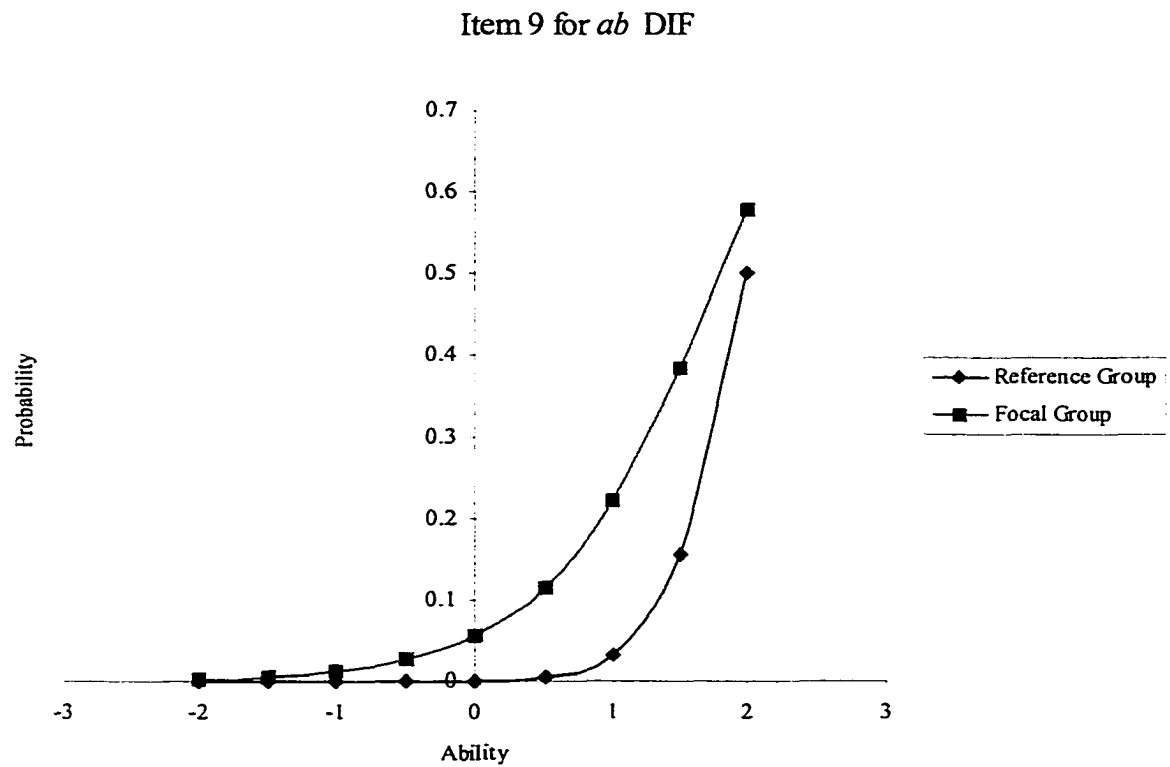
Figure B.19. Parameter values and ICCs of item 1 for *b* DIF

Item Parameter Values: low *a* (discrimination) and low *b* (difficulty)

	a parameter	*b* parameter
Reference group	0.5001	-1.5000
Focal group	0.5001	-2.0000

Item 1 for *b* DIF
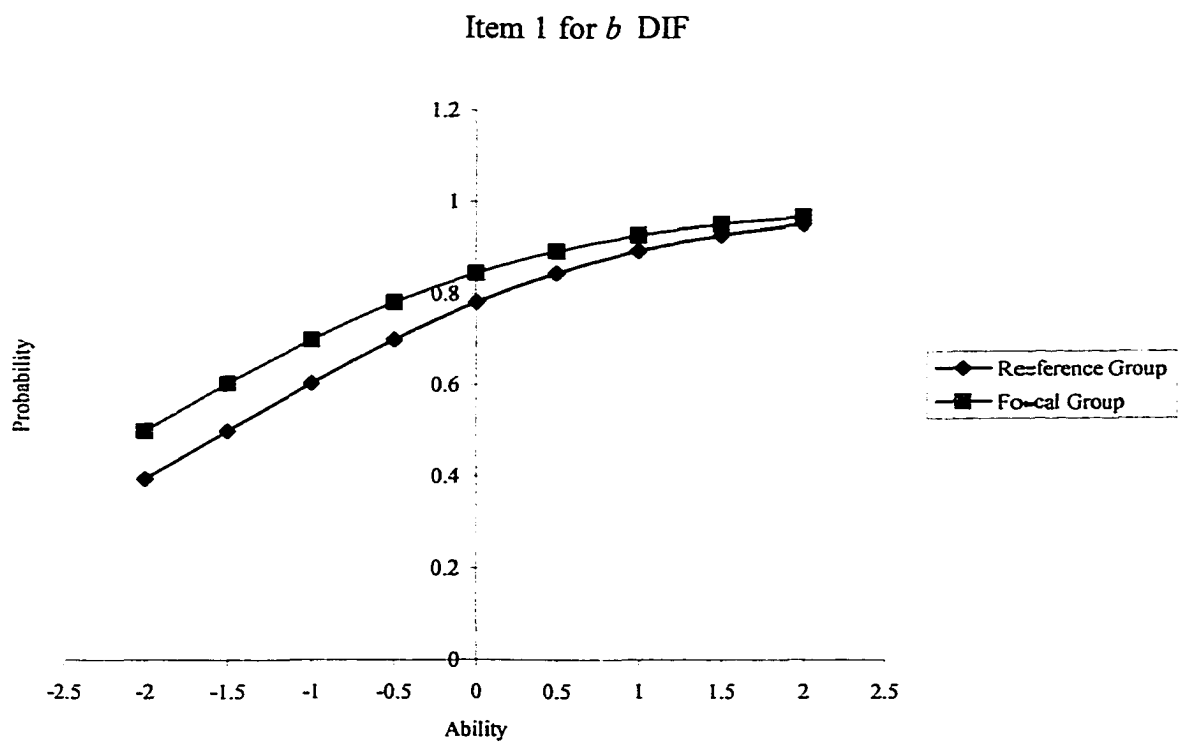
Figure B.20. Parameter values and ICCs of item 2 for *b* DIF

Item Parameter Values: low *a* (discrimination) and medium *b* (difficulty)

	a parameter	*b* parameter
Reference group	0.5001	0.0000
Focal group	0.5001	-0.5000

Item 2 for *b* DIF
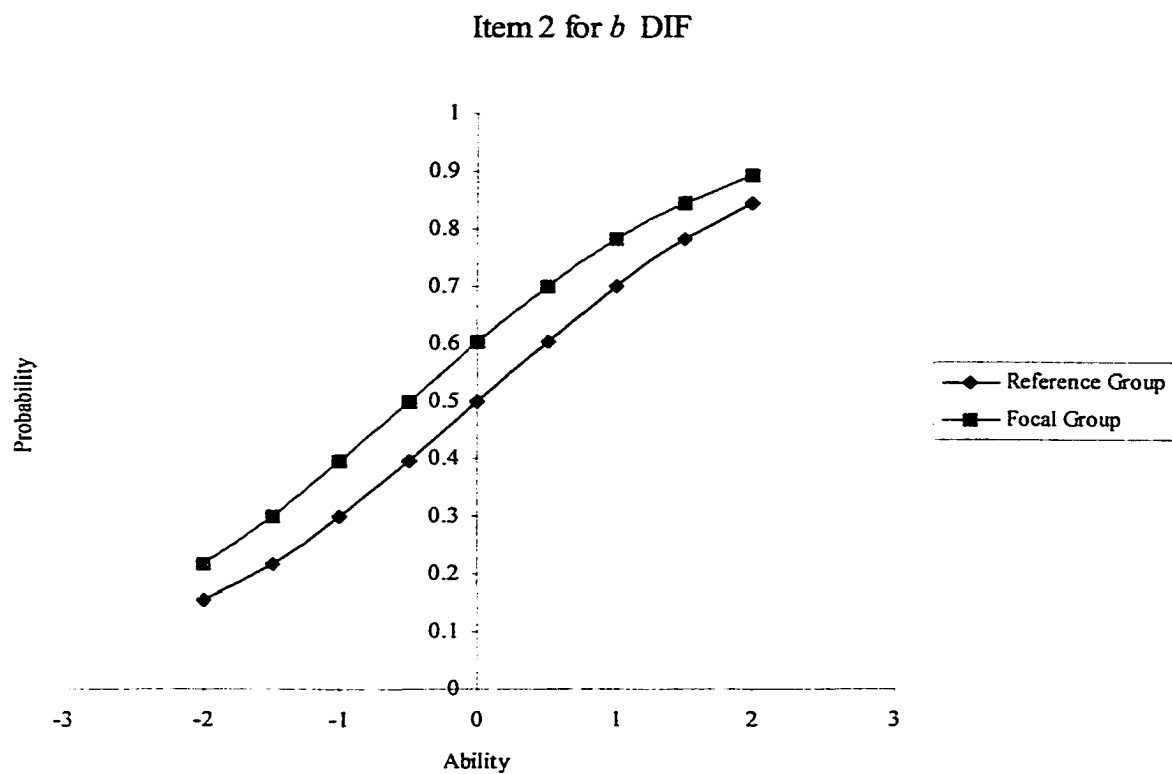
Figure B.21. Parameter values and ICCs of item 3 for *b* DIF

Item Parameter Values: low *a* (discrimination) and high *b* (difficulty)

	a parameter	*b* parameter
Reference group	0.5001	2.0000
Focal group	0.5001	1.5000

Item 3 for *b* DIF
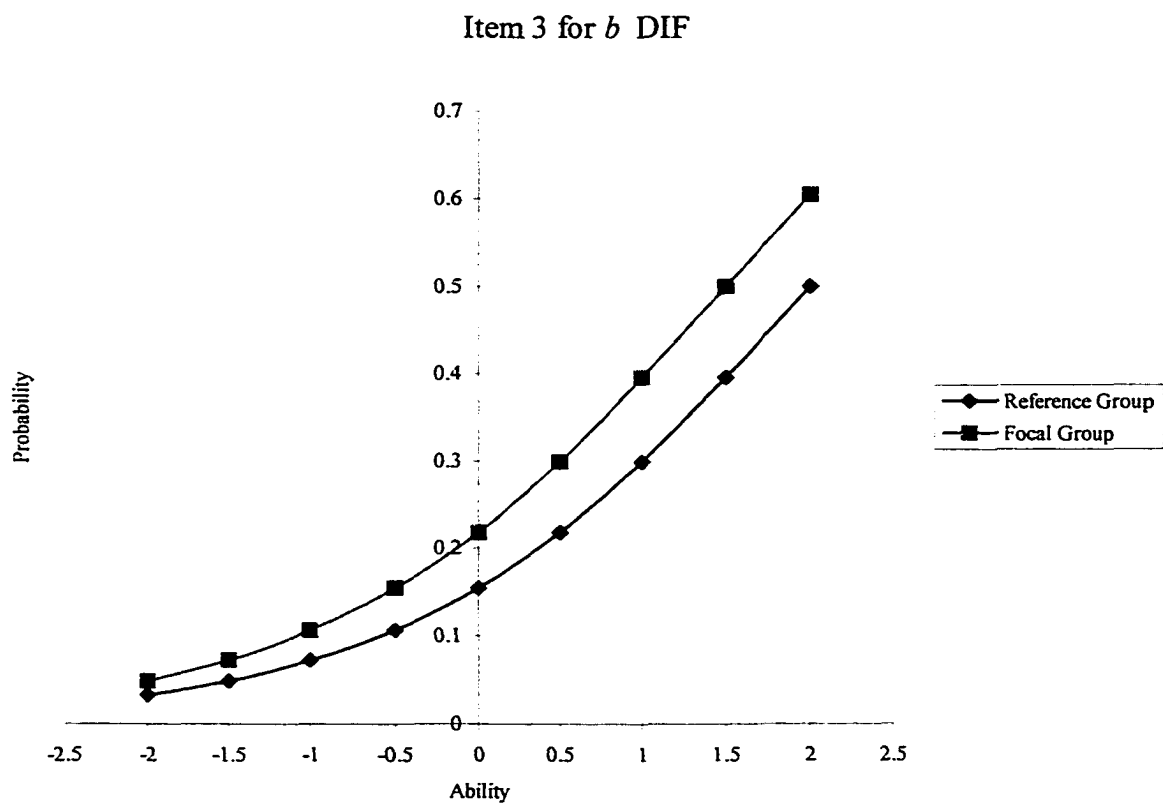
Figure B.22. Parameter values and ICCs of item 4 for *b* DIF

Item Parameter Values: medium *a* (discrimination and low *b* (difficulty)

	a parameter	*b* parameter
Reference group	1.0000	-1.5000
Focal group	1.0000	-2.0000

Item 4 for *b* DIF
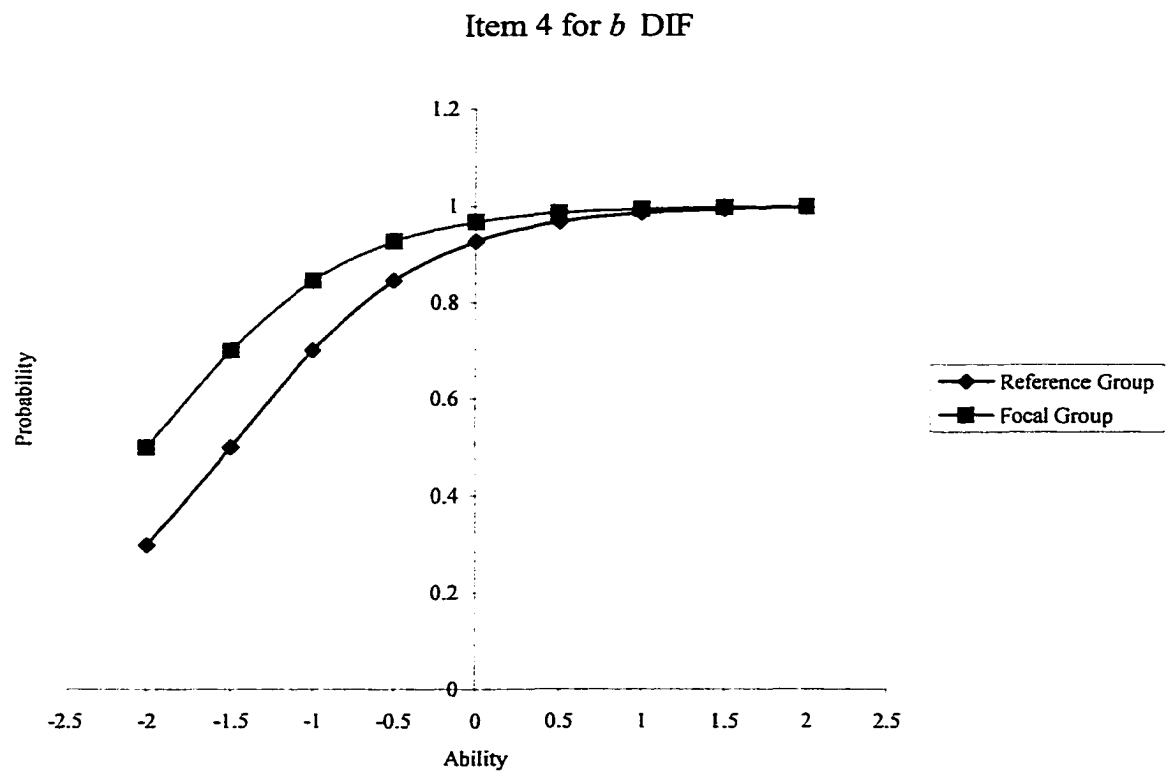
Figure B.23. Parameter values and ICCs of item 5 for *b* DIF

Item Parameter Values: medium *a* (discrimination) and medium *b* (difficulty)

	a parameter	*b* parameter
Reference group	1.0000	0.0000
Focal group	1.0000	-0.5000

Item 5 for *b* DIF
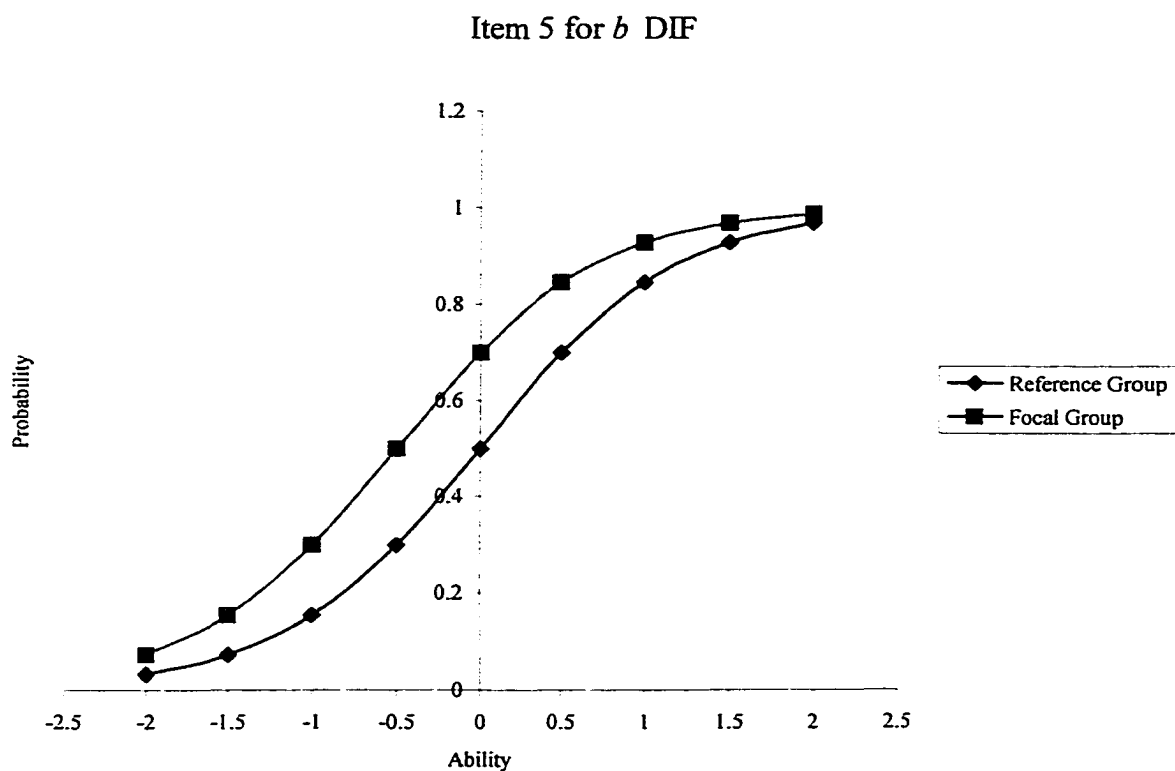
Figure B.24. Parameter values and ICCs of item 6 for *b* DIF

Item Parameter Values: medium *a* (discrimination) and high *b* (difficulty)

	a parameter	*b* parameter
Reference group	1.0000	2.0000
Focal group	1.0000	1.5000

Item 6 for *b* DIF
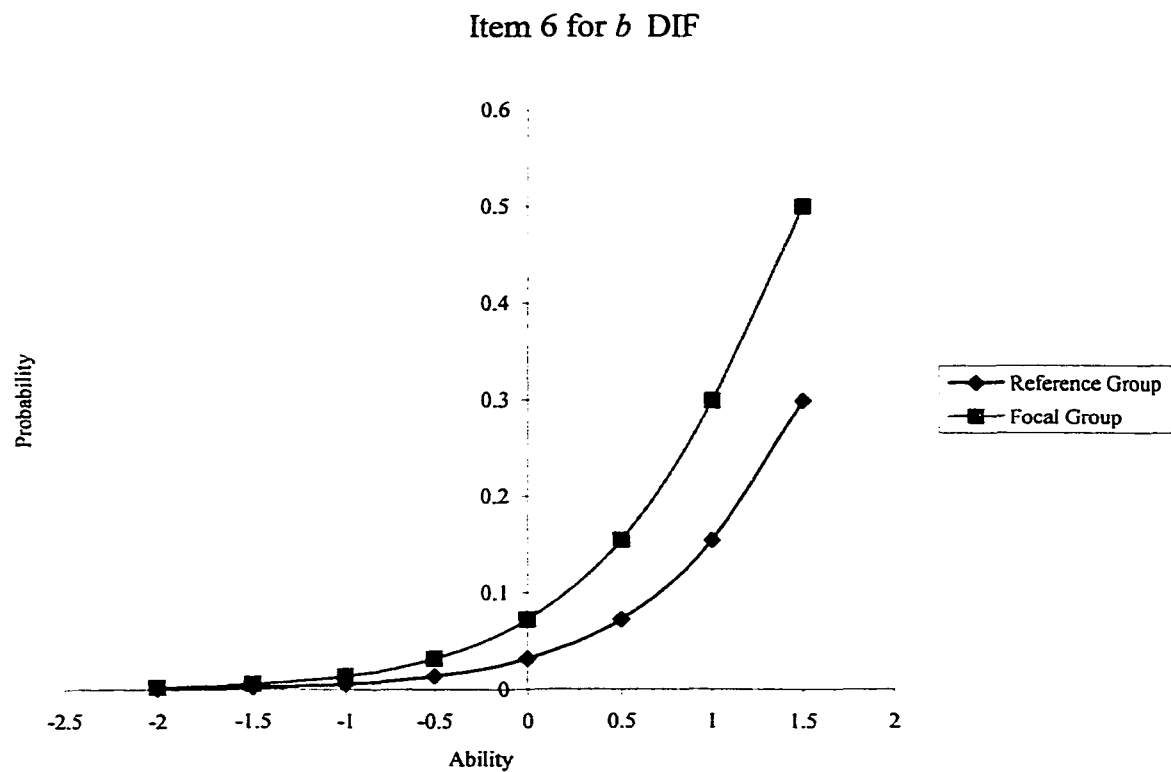
Figure B.25. Parameter values and ICCs of item 7 for *b* DIF

Item Parameter Values: high *a* (discrimination) and low *b* (difficulty)

	a parameter	*b* parameter
Reference group	2.0000	-1.5000
Focal group	2.0000	-2.0000

Item 7 for *b* DIF
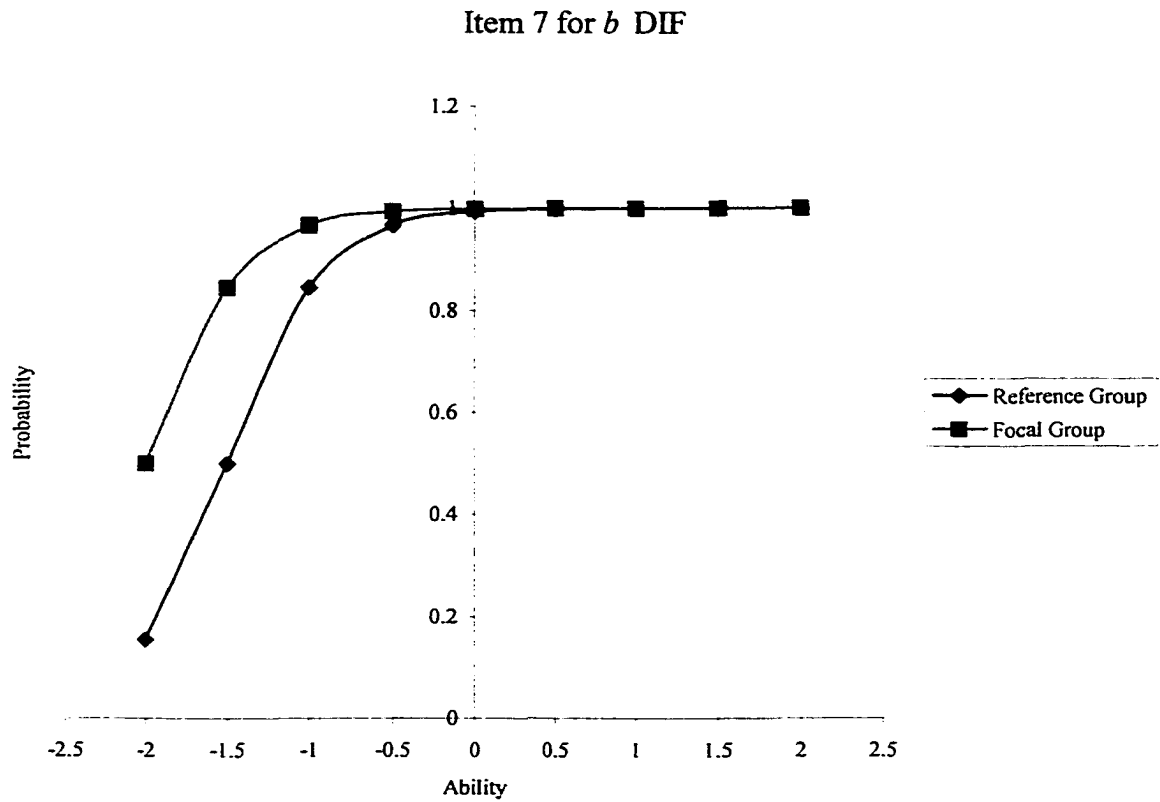
Figure B.26. Parameter values and ICCs of item 8 for *b* DIF

Item Parameter Values: high *a* (discrimination) and medium *b* (difficulty)

	a parameter	*b* parameter
Reference group	2.0000	0.0000
Focal group	2.0000	-0.5000

Item 8 for *b* DIF
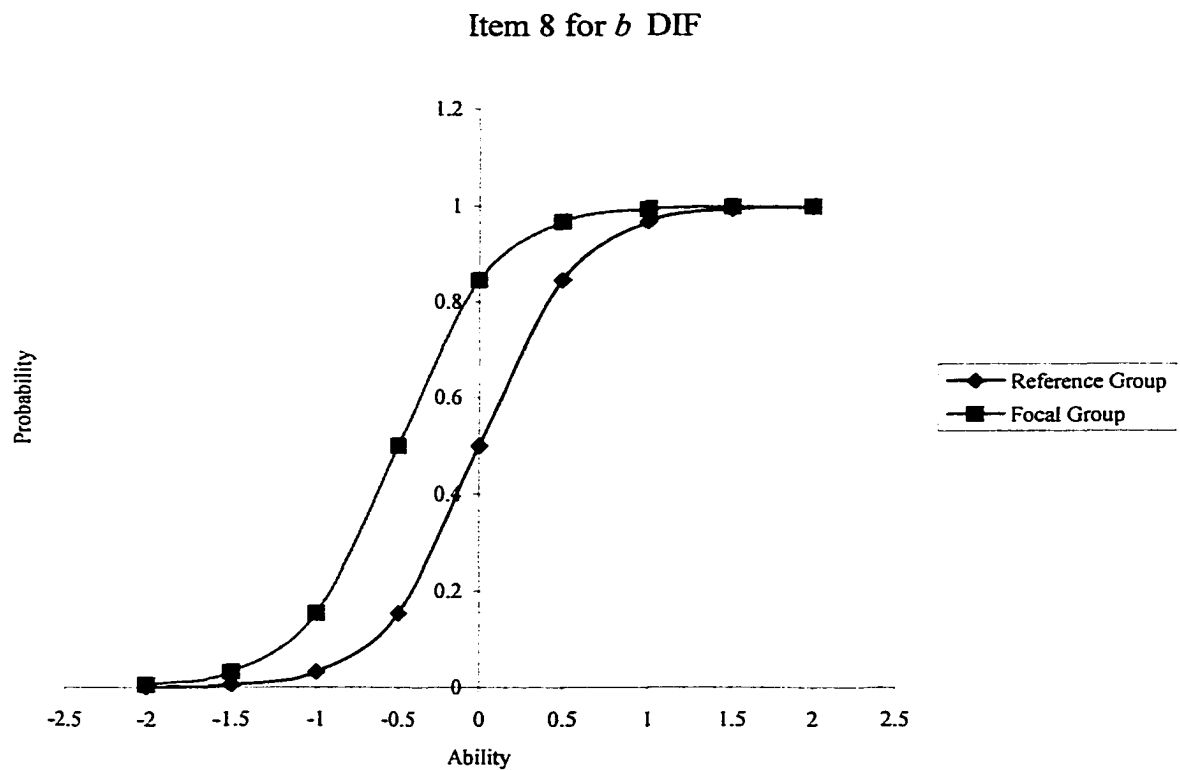
Figure B.27. Parameter values and ICCs of item 9 for *b* DIF

Item Parameter Values: high *a* (discrimination) and high *b* (difficulty)

	a parameter	*b* parameter
Reference group	2.0000	2.0000
Focal group	2.0000	1.5000

Item 9 for *b* DIF
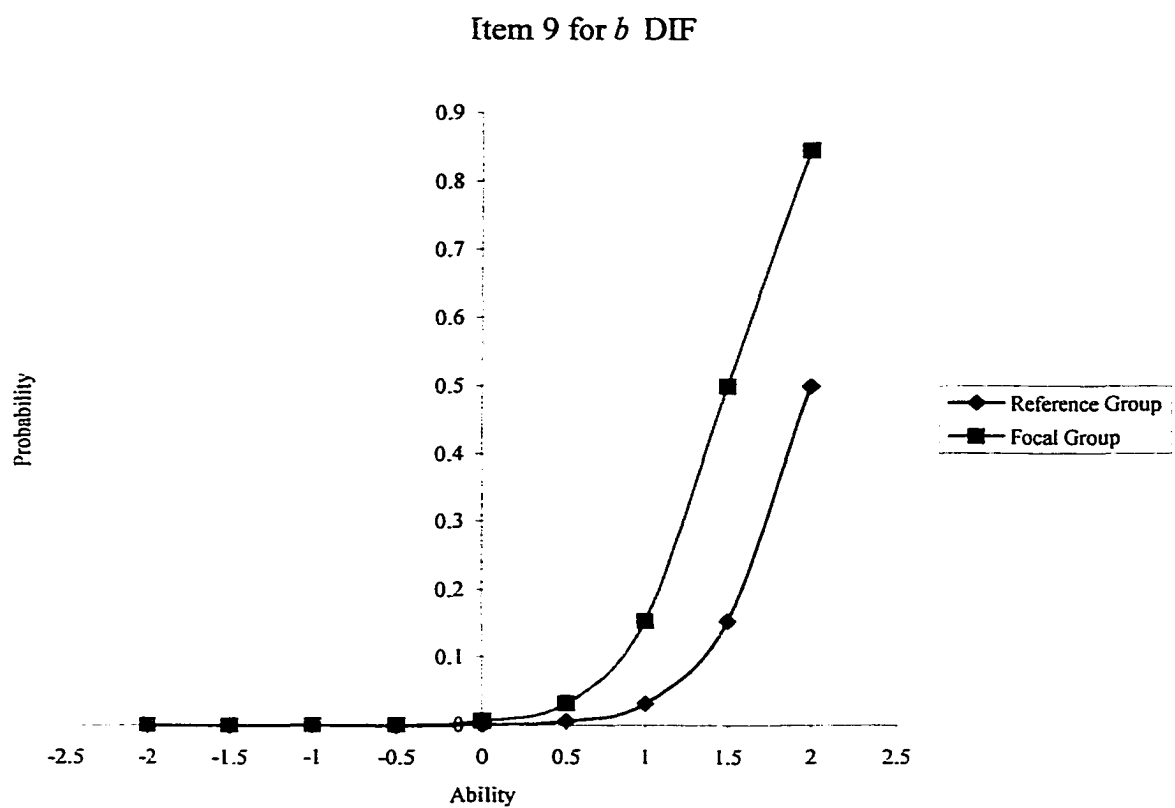
APPENDIX C

Chi-Square Goodness-of-Fit Statistic G(1) and G(2) of IRT

Table C.1

Chi-square Goodness-of-Fit Statistic G(1) and G(2)and their Difference for a DIF

Replication 1	G(1)	G(2)	Difference
Item 1	56837.9	56842.7	4.8
Item 2	57065.8	57070.7	4.9
Item 3	56297.0	56311.3	14.3*
Item 4	56445.0	56458.8	13.8*
Item 5	56762.9	56770.4	7.5*
Item 6	55686.1	55698.6	12.5*
Item 7	56128.8	56150.1	21.3*
Item 8	56302.2	56343.1	40.9*
Item 9	55162.1	55196.8	34.7*
Replication 2	G(1)	G(2)	Difference
Item 1	62085.8	62090.5	4.7
Item 2	62361.8	62366.5	4.7
Item 3	61673.6	61688.1	14.5*
Item 4	61693.7	61710.3	16.6*
Item 5	62054.8	62073.9	19.1*
Item 6	60937.2	60969.1	31.9*
Item 7	61375.1	61401.5	26.4*
Item 8	61570.7	61622.5	51.8*
Item 9	60417.0	60458.6	41.6*
Replication 3	G(1)	G(2)	Difference
Item 1	58625.1	58626.7	1.6
Item 2	58857.8	58862.2	4.4
Item 3	58064.4	58072.1	7.7*
Item 4	58185.8	58220.4	34.6*
Item 5	58531.8	58560.1	28.3*
Item 6	57399.1	57424.8	25.7*
Item 7	57978.5	57994.6	16.1*
Item 8	58189.0	58240.1	51.1*
Item 9	56902.7	56941.9	39.2*

Table C.1 *(continued)*

Replication 4	G(1)	G(2)	Difference
Item 1	55876.4	55882.3	5.9
Item 2	56139.3	56144.4	5.1
Item 3	55318.7	55321.4	2.7
Item 4	55505.0	55520.3	15.3*
Item 5	55751.4	55759.2	7.8*
Item 6	54724.7	54760.6	35.9*
Item 7	55233.5	55253.8	20.3*
Item 8	55421.6	55457.5	35.9*
Item 9	54239.1	54274.0	34.9*
Replication 5	**G(1)**	**G(2)**	**Difference**
Item 1	55939.1	55940.2	1.1
Item 2	56097.4	56101.5	4.1
Item 3	55395.9	55402.0	6.1*
Item 4	55489.0	55510.3	21.3*
Item 5	55785.9	55803.2	17.3*
Item 6	54694.1	54728.9	34.8*
Item 7	55163.7	55184.0	20.3*
Item 8	55417.4	55463.7	46.3*
Item 9	54278.0	54335.4	57.4*
Replication 6	**G(1)**	**G(2)**	**Difference**
Item 1	56617.9	56621.6	3.7
Item 2	56768.0	56771.8	3.8
Item 3	56045.9	56052.7	6.8*
Item 4	56161.2	56178.7	17.5*
Item 5	56524.8	56568.7	43.9*
Item 6	55433.4	55471.2	37.8*
Item 7	55883.3	55890.9	7.6*
Item 8	56178.1	56216.8	38.7*
Item 9	54966.1	55014.2	48.1*
Replication 7	**G(1)**	**G(2)**	**Difference**
Item 1	58466.6	58475.8	9.2*
Item 2	58686.2	58689.6	3.4
Item 3	58011.6	58017.5	5.9
Item 4	58092.0	58105.8	13.8*
Item 5	58425.2	58434.6	9.4*
Item 6	57289.7	57318.6	28.9*
Item 7	57710.2	57725.7	15.5*
Item 8	57868.4	57916.0	47.6*
Item 9	56805.8	56839.6	33.8*

Table C.1 *(continued)*

Replication 8	G(1)	G(2)	Difference
Item 1	56964.5	56972.5	8.0*
Item 2	57220.3	57228.5	8.2*
Item 3	56526.5	56544.1	17.6*
Item 4	56580.6	56595.9	15.3*
Item 5	56836.3	56849.5	13.2*
Item 6	55746.4	55787.1	40.7*
Item 7	56259.0	56281.9	22.9*
Item 8	56503.2	56562.8	59.6*
Item 9	55331.4	55381.8	50.4*
Replication 9	G(1)	G(2)	Difference
Item 1	56765.2	56769.5	4.3
Item 2	57007.9	57009.1	1.2
Item 3	56287.0	56298.4	11.4*
Item 4	56424.5	56441.1	16.6*
Item 5	56621.9	56643.7	21.8*
Item 6	55687.3	55707.8	20.5*
Item 7	56122.8	56139.9	17.1*
Item 8	56225.1	56263.2	38.1*
Item 9	55204.3	55226.2	21.9*
Replication 10	G(1)	G(2)	Difference
Item 1	59359.9	59364.8	4.9
Item 2	59635.4	59640.3	4.9
Item 3	58866.4	58893.6	27.2*
Item 4	58969.3	58979.5	10.2*
Item 5	59337.5	59351.0	13.5*
Item 6	58294.3	58320.4	26.1*
Item 7	58742.5	58762.8	20.3*
Item 8	58806.4	58877.4	71.0*
Item 9	57698.5	57749.0	50.5*
Replication 11	G(1)	G(2)	Difference
Item 1	54886.6	54894.1	7.5*
Item 2	55172.7	55183.2	10.5*
Item 3	54511.2	54524.2	13.0*
Item 4	54595.0	54602.2	7.2*
Item 5	54875.3	54892.5	17.2*
Item 6	53803.9	53841.1	37.2*
Item 7	54286.9	54300.9	14.0*
Item 8	54512.3	54574.0	61.7*
Item 9	53329.7	53371.6	41.9*

Table C.1 *(continued)*

Replication 12	G(1)	G(2)	Difference
Item 1	56229.7	56237.9	8.2*
Item 2	56464.1	56467.7	3.6
Item 3	55800.8	55817.0	16.2*
Item 4	55734.6	55743.8	9.2*
Item 5	56172.2	56185.7	13.5*
Item 6	55079.0	55107.9	28.9*
Item 7	55519.6	55525.5	5.9
Item 8	55778.4	55813.8	35.4*
Item 9	54638.7	54687.8	49.1*
Replication 13	G(1)	G(2)	Difference
Item 1	55023.7	55032.3	8.6*
Item 2	55331.7	55333.9	2.2
Item 3	54640.0	54649.5	9.5*
Item 4	54661.0	54676.9	15.9*
Item 5	54903.4	54929.1	25.7*
Item 6	53924.4	53971.7	47.3*
Item 7	54373.4	54389.4	16.0*
Item 8	54551.3	54605.4	54.1*
Item 9	53442.6	53506.8	64.2*
Replication 14	G(1)	G(2)	Difference
Item 1	60295.0	60296.9	1.9
Item 2	60538.7	60546.2	7.5*
Item 3	59800.2	59810.8	10.6*
Item 4	59941.0	59952.3	11.3*
Item 5	60217.7	60248.2	30.5*
Item 6	59097.4	59148.9	51.5*
Item 7	59654.4	59667.4	13.0*
Item 8	59816.7	59848.0	31.3*
Item 9	58596.0	58641.5	45.5*
Replication 15	G(1)	G(2)	Difference
Item 1	58020.2	58021.5	1.3
Item 2	58320.2	58323.0	2.8
Item 3	57604.1	57622.0	17.9*
Item 4	57576.6	57588.6	12.0*
Item 5	58004.9	58026.7	21.8*
Item 6	56871.7	56887.4	15.7*
Item 7	57407.0	57430.6	23.6*
Item 8	57609.6	57678.2	68.6*
Item 9	56370.0	56413.7	43.7*

Table C.1 *(continued)*

Replication 16	G(1)	G(2)	Difference
Item 1	51687.1	51694.7	7.6*
Item 2	51892.9	51899.4	6.5*
Item 3	51073.9	51081.5	7.6*
Item 4	51222.2	51248.7	26.5*
Item 5	51514.7	51556.8	42.1*
Item 6	50414.9	50435.8	20.9*
Item 7	50982.6	50996.1	13.5*
Item 8	51147.3	51228.8	81.5*
Item 9	49980.8	50012.0	31.2*
Replication 17	G(1)	G(2)	Difference
Item 1	56168.6	56170.8	2.2
Item 2	56399.0	56401.3	2.3
Item 3	55787.7	55791.7	4.0
Item 4	55790.5	55793.0	2.5
Item 5	56127.0	56149.8	22.8*
Item 6	55114.5	55150.8	36.3*
Item 7	55563.1	55573.6	10.5*
Item 8	55766.0	55814.6	48.6*
Item 9	54615.6	54671.5	55.9*
Replication 18	G(1)	G(2)	Difference
Item 1	59643.1	59652.5	9.4*
Item 2	59913.5	59914.1	0.6
Item 3	59021.4	59030.5	9.1*
Item 4	59260.1	59274.4	14.3*
Item 5	59534.2	59543.3	9.1*
Item 6	58515.3	58554.8	39.5*
Item 7	58906.3	58913.8	7.5*
Item 8	59151.4	59238.2	86.8*
Item 9	57978.3	58025.2	46.9*
Replication 19	G(1)	G(2)	Difference
Item 1	57028.3	57034.5	6.2*
Item 2	57293.5	57295.7	2.2
Item 3	56534.2	56534.3	0.1
Item 4	56547.0	56550.0	3.0
Item 5	56957.9	56979.1	21.2*
Item 6	55836.6	55869.6	33.0*
Item 7	56312.1	56328.9	16.8*
Item 8	56528.6	56567.2	38.6*
Item 9	55399.6	55445.6	46.0*

Table C.1 *(continued)*

Replication 20	G(1)	G(2)	Difference
Item 1	58649.0	58655.2	6.2*
Item 2	58880.5	58898.1	17.6*
Item 3	58131.2	58139.6	8.4*
Item 4	58246.9	58270.7	23.8*
Item 5	58474.6	58488.2	13.6*
Item 6	57490.1	57528.6	38.5*
Item 7	57941.9	57953.5	11.6*
Item 8	58065.5	58115.5	50.0*
Item 9	56996.2	57042.3	46.1*
Replication 21	G(1)	G(2)	Difference
Item 1	57615.4	57621.4	6.0*
Item 2	57801.3	57804.6	3.3
Item 3	57055.9	57073.5	17.6*
Item 4	57209.2	57221.2	12.0*
Item 5	57510.8	57534.5	23.7*
Item 6	56430.5	56460.5	30.0*
Item 7	56907.3	56927.1	19.8*
Item 8	57117.4	57159.8	42.4*
Item 9	56050.6	56098.7	48.1*
Replication 22	G(1)	G(2)	Difference
Item 1	56320.3	56325.7	5.4
Item 2	56526.9	56532.6	5.7
Item 3	55780.4	55791.8	11.4*
Item 4	55843.0	55865.1	22.1*
Item 5	56162.1	56175.5	13.4*
Item 6	55076.6	55115.3	38.7*
Item 7	55553.9	55575.5	21.6*
Item 8	55703.8	55754.2	50.4*
Item 9	54721.1	54771.6	50.5*
Replication 23	G(1)	G(2)	Difference
Item 1	61668.2	61673.7	5.5
Item 2	61924.0	61924.7	0.7
Item 3	61244.1	61253.6	9.5*
Item 4	61238.2	61252.9	14.7*
Item 5	61573.8	61587.3	13.5*
Item 6	60463.4	60474.7	11.3*
Item 7	60927.9	60943.7	15.8*
Item 8	61190.6	61227.8	37.2*
Item 9	60003.9	60061.5	57.6*

Table C.1 *(continued)*

Replication 24	G(1)	G(2)	Difference
Item 1	54224.6	54234.0	9.4*
Item 2	54487.7	54496.0	8.3*
Item 3	53858.6	53861.1	2.5
Item 4	53862.7	53887.8	25.1*
Item 5	54156.2	54170.2	14.0*
Item 6	53058.7	53094.8	36.1*
Item 7	53562.5	53586.0	23.5*
Item 8	53722.9	53770.8	47.9*
Item 9	52680.2	52712.3	32.1*
Replication 25	G(1)	G(2)	Difference
Item 1	62312.6	62325.9	13.3*
Item 2	62476.8	62480.6	3.8
Item 3	61787.0	61797.7	10.7*
Item 4	61847.1	61860.4	13.3*
Item 5	62183.7	62194.7	11.0*
Item 6	61071.6	61108.1	36.5*
Item 7	61587.2	61606.4	19.2*
Item 8	61793.4	61852.7	59.3*
Item 9	60741.8	60793.9	52.1*

Note. The Critical values is 5.991 for chi-square with df = 2.
* $p<.05$

Table C.2

Chi-square Goodness-of-Fit Statistic G(1) and G(2) and their Difference for ab DIF

Replication 1	G(1)	G(2)	Difference
Item 1	57065.2	57065.4	0.2
Item 2	57338.7	57343.4	4.7
Item 3	56626.0	56641.0	15.0*
Item 4	56688.7	56693.2	4.5
Item 5	57094.6	57107.2	12.6*
Item 6	56014.5	56030.5	16.0*
Item 7	56322.8	56327.6	4.8
Item 8	56583.4	56618.6	35.2*
Item 9	55533.1	55590.5	57.4*
Replication 2	G(1)	G(2)	Difference
Item 1	61816.1	61823.4	7.3*
Item 2	62030.4	62032.0	1.6
Item 3	61367.8	61385.0	17.2*
Item 4	61296.4	61301.7	5.3
Item 5	61750.2	61776.6	26.4*
Item 6	60675.4	60719.4	44.0*
Item 7	60973.1	60980.9	7.8*
Item 8	61269.2	61344.7	75.5*
Item 9	60219.3	60307.4	88.1*
Replication 3	G(1)	G(2)	Difference
Item 1	58404.8	58409.8	5.0
Item 2	58652.0	58659.5	7.5*
Item 3	57894.2	57908.6	14.4*
Item 4	57827.9	57834.4	6.5*
Item 5	58333.0	58374.4	41.4*
Item 6	57292.9	57347.4	54.5*
Item 7	57659.5	57661.3	1.8
Item 8	57964.6	58020.0	55.4*
Item 9	56800.8	56870.8	70.0*

Table C.2 *(continued)*

Replication 4	G(1)	G(2)	Difference
Item 1	55989.0	55990.8	1.8
Item 2	56274.5	56275.4	0.9
Item 3	55546.0	55557.7	11.7*
Item 4	55578.6	55582.9	4.3
Item 5	55940.2	55956.3	16.1*
Item 6	54864.2	54904.9	40.7*
Item 7	55324.9	55330.1	5.2
Item 8	55590.3	55641.1	50.8*
Item 9	54462.6	54516.6	54.0*
Replication 5	G(1)	G(2)	Difference
Item 1	56045.4	56047.2	1.8
Item 2	56215.0	56219.6	4.6
Item 3	55580.7	55597.4	16.7*
Item 4	55524.0	55531.5	7.5*
Item 5	55914.1	55937.8	23.7*
Item 6	54889.3	54945.7	56.4*
Item 7	55206.5	55213.3	6.8*
Item 8	55507.1	55552.6	45.5*
Item 9	54424.7	54501.7	77.0*
Replication 6	G(1)	G(2)	Difference
Item 1	56441.0	56441.6	0.6
Item 2	56609.2	56611.8	2.6
Item 3	55986.3	56007.6	21.3*
Item 4	55962.7	55971.7	9.0*
Item 5	56295.5	56324.2	28.7*
Item 6	55230.2	55256.2	26.0*
Item 7	55692.1	55697.3	5.2
Item 8	56009.9	56047.6	37.7*
Item 9	54876.9	54940.4	63.5*
Replication 7	G(1)	G(2)	Difference
Item 1	58133.9	58136.7	2.8
Item 2	58446.2	58450.5	4.3
Item 3	57725.7	57729.9	4.2
Item 4	57775.4	57781.2	5.8
Item 5	58192.5	58218.8	26.3*
Item 6	57073.4	57103.7	30.3*
Item 7	57507.8	57520.5	12.7*
Item 8	57694.2	57774.5	80.3*
Item 9	56652.4	56710.0	57.6*

Table C.2 *(continued)*

Replication 8	G(1)	G(2)	Difference
Item 1	57144.7	57145.8	1.1
Item 2	57418.6	57419.0	0.4
Item 3	56797.3	56810.7	13.4*
Item 4	56760.6	56766.0	5.4
Item 5	57071.3	57082.9	11.6*
Item 6	56078.4	56139.7	61.3*
Item 7	56452.0	56459.6	7.6*
Item 8	56790.8	56867.8	77.0*
Item 9	55650.2	55729.0	78.8*
Replication 9	G(1)	G(2)	Difference
Item 1	56961.0	56971.1	10.1*
Item 2	57240.8	57246.1	5.3
Item 3	56489.7	56500.2	10.5*
Item 4	56549.6	56558.6	9.0*
Item 5	56893.1	56943.1	50.0*
Item 6	55885.7	55905.0	19.3*
Item 7	56301.4	56312.1	10.7*
Item 8	56482.2	56537.5	55.3*
Item 9	55554.7	55608.2	53.5*
Replication 10	G(1)	G(2)	Difference
Item 1	59353.9	59360.2	6.3*
Item 2	59621.2	59633.3	12.1*
Item 3	58894.7	58915.7	21.0*
Item 4	58883.8	58887.9	4.1
Item 5	59284.8	59292.0	7.2*
Item 6	58308.5	58339.0	30.5*
Item 7	58576.0	58577.3	1.3
Item 8	58853.3	58967.8	114.5*
Item 9	57802.0	57871.1	69.1*
Replication 11	G(1)	G(2)	Difference
Item 1	54995.2	55008.6	13.4*
Item 2	55272.9	55288.5	15.6*
Item 3	54661.5	54682.0	20.5*
Item 4	54587.3	54588.0	0.7
Item 5	54967.2	54987.9	20.7*
Item 6	53963.7	54028.0	64.3*
Item 7	54309.9	54314.0	4.1
Item 8	54576.3	54650.1	73.8*
Item 9	53454.1	53518.9	64.8*

Table C.2 *(continued)*

Replication 12	G(1)	G(2)	Difference
Item 1	55973.6	55974.8	1.2
Item 2	56263.1	56274.4	11.3*
Item 3	55575.4	55592.2	16.8*
Item 4	55543.1	55548.8	5.7
Item 5	55945.7	55960.4	14.7*
Item 6	54898.3	54934.7	36.4*
Item 7	55277.2	55284.3	7.1*
Item 8	55560.2	55609.6	49.4*
Item 9	54401.7	54446.9	45.2*
Replication 13	G(1)	G(2)	Difference
Item 1	54989.0	54999.0	10.0*
Item 2	55301.1	55303.5	2.4
Item 3	54641.2	54654.8	13.6*
Item 4	54585.6	54596.2	10.6*
Item 5	54938.0	54990.9	52.9*
Item 6	53911.6	53967.2	55.6*
Item 7	54317.2	54340.3	23.1*
Item 8	54490.5	54546.8	56.3*
Item 9	53446.8	53538.9	92.1*
Replication 14	G(1)	G(2)	Difference
Item 1	60191.7	60195.4	3.7
Item 2	60409.1	60413.4	4.3
Item 3	59708.8	59724.0	15.2*
Item 4	59781.0	59790.9	9.9*
Item 5	60061.7	60082.5	20.8*
Item 6	59098.1	59184.4	86.3*
Item 7	59457.5	59461.0	3.5
Item 8	59695.0	59733.0	38.0*
Item 9	58543.0	58612.9	69.9*
Replication 15	G(1)	G(2)	Difference
Item 1	58119.6	58124.6	5.0
Item 2	58339.1	58341.0	1.9
Item 3	57582.8	57587.3	4.5
Item 4	57627.6	57636.1	8.5*
Item 5	58016.7	58029.5	12.8*
Item 6	57035.2	57066.4	31.2*
Item 7	57330.1	57334.0	3.9
Item 8	57634.5	57695.6	61.1*
Item 9	56518.9	56588.1	69.2*

Table C.2 *(continued)*

Replication 16	G(1)	G(2)	Difference
Item 1	51555.2	51562.4	7.2*
Item 2	51818.9	51826.6	7.7*
Item 3	51049.7	51062.3	12.6*
Item 4	51082.7	51099.9	17.2*
Item 5	51390.6	51417.4	26.8*
Item 6	50490.5	50542.8	52.3*
Item 7	50830.3	50835.7	5.4
Item 8	51018.7	51084.5	65.8*
Item 9	49988.7	50046.3	57.6*
Replication 17	G(1)	G(2)	Difference
Item 1	56136.3	56138.0	1.7
Item 2	56396.3	56399.9	3.6
Item 3	55761.8	55764.9	3.1
Item 4	55749.1	55751.9	2.8
Item 5	56101.2	56123.0	21.8*
Item 6	55063.9	55093.6	29.7*
Item 7	55462.1	55465.3	3.2
Item 8	55733.4	55786.0	52.6*
Item 9	54546.5	54597.1	50.6*
Replication 18	G(1)	G(2)	Difference
Item 1	59781.8	59786.0	4.2
Item 2	60101.0	60103.1	2.1
Item 3	59283.8	59307.3	23.5*
Item 4	59388.4	59397.0	8.6*
Item 5	59791.4	59821.9	30.5*
Item 6	58736.7	58789.3	52.6*
Item 7	59078.7	59086.4	7.7*
Item 8	59237.5	59291.6	54.1*
Item 9	58269.4	58348.4	79.0*
Replication 19	G(1)	G(2)	Difference
Item 1	57045.4	57046.7	1.3
Item 2	57308.5	57309.8	1.3
Item 3	56633.8	56635.7	1.9
Item 4	56624.0	56630.6	6.6*
Item 5	57001.4	57020.5	19.1*
Item 6	55961.6	56010.4	48.8*
Item 7	56332.0	56338.6	6.6*
Item 8	56584.5	56624.8	40.3*
Item 9	55478.9	55539.8	60.9*

Table C.2 *(continued)*

Replication 20	G(1)	G(2)	Difference
Item 1	58649.0	58654.2	5.2
Item 2	58871.5	58875.3	3.8
Item 3	58182.6	58193.0	10.4*
Item 4	58141.7	58149.9	8.2*
Item 5	58551.5	58585.1	33.6*
Item 6	57524.3	57572.9	48.6*
Item 7	57893.9	57898.5	4.6
Item 8	58106.6	58162.9	56.3*
Item 9	57039.7	57090.2	50.5*
Replication 21	G(1)	G(2)	Difference
Item 1	57657.5	57661.3	3.8
Item 2	57868.7	57871.9	3.2
Item 3	57155.7	57174.8	19.1*
Item 4	57267.7	57274.9	7.2*
Item 5	57574.3	57595.7	21.4*
Item 6	56627.0	56686.3	59.3*
Item 7	56896.6	56903.4	6.8*
Item 8	57258.9	57325.1	66.2*
Item 9	56118.3	56177.2	58.9*
Replication 22	G(1)	G(2)	Difference
Item 1	56662.2	56670.9	8.7*
Item 2	56840.3	56843.5	3.2
Item 3	56190.3	56218.0	27.7*
Item 4	56103.7	56114.7	11.0*
Item 5	56567.7	56614.4	46.7*
Item 6	55490.0	55554.4	64.4*
Item 7	55825.8	55837.0	11.2*
Item 8	56103.4	56183.7	80.3*
Item 9	55050.5	55093.4	42.9*
Replication 23	G(1)	G(2)	Difference
Item 1	61740.6	61747.1	6.5*
Item 2	61998.8	62000.2	1.4
Item 3	61367.9	61388.3	20.4*
Item 4	61226.1	61230.6	4.5
Item 5	61674.5	61697.8	23.3*
Item 6	60689.6	60732.8	43.2*
Item 7	60970.7	60982.5	11.8*
Item 8	61269.3	61310.9	41.6*
Item 9	60117.2	60189.3	72.1*

Table C.2 *(continued)*

Replication 24	G(1)	G(2)	Difference
Item 1	54688.6	54697.8	9.2*
Item 2	54931.9	54934.9	3.0
Item 3	54350.8	54355.2	4.4
Item 4	54280.6	54298.5	17.9*
Item 5	54653.9	54675.4	21.5*
Item 6	53577.3	53635.9	58.6*
Item 7	53985.2	53999.3	14.1*
Item 8	54267.5	54355.1	87.6*
Item 9	53237.9	53284.4	46.5*
Replication 25	G(1)	G(2)	Difference
Item 1	62007.3	62012.6	5.3
Item 2	62269.6	62282.9	13.3*
Item 3	61563.8	61574.7	10.9*
Item 4	61563.7	61570.2	6.5*
Item 5	61985.0	62008.7	23.7*
Item 6	60879.3	60934.1	54.8*
Item 7	61333.2	61351.8	18.6*
Item 8	61538.7	61596.4	57.7*
Item 9	60490.1	60548.6	58.5*

Note. The Critical values is 5.991 for chi-square with df = 2.
* $p<.05$

Table C.3

Chi-square Goodness-of-Fit Statistic G(1) and G(2) and their Difference for b DIF

Replication 1	G(1)	G(2)	Difference
Item 1	56911.7	56921.7	10.0*
Item 2	57203.0	57229.9	26.9*
Item 3	56548.5	56563.2	14.7*
Item 4	56394.2	56403.0	8.8*
Item 5	56858.4	56894.0	35.6*
Item 6	55882.3	55894.2	11.9*
Item 7	56022.5	56065.3	42.8*
Item 8	56288.6	56381.6	93.0*
Item 9	55296.8	55323.1	26.3*
Replication 2	G(1)	G(2)	Difference
Item 1	61829.9	61841.9	12.0*
Item 2	62231.6	62238.6	7.0*
Item 3	61562.4	61573.5	11.1*
Item 4	61273.7	61288.7	15.0*
Item 5	61736.9	61770.5	33.6*
Item 6	60782.9	60809.8	26.9*
Item 7	60960.3	60983.6	23.3*
Item 8	61253.1	61374.5	121.4*
Item 9	60297.4	60372.7	75.3*
Replication 3	G(1)	G(2)	Difference
Item 1	58404.6	58423.9	19.3*
Item 2	58831.1	58857.6	26.5*
Item 3	58131.7	58151.2	19.5*
Item 4	57853.1	57867.6	14.5*
Item 5	58372.6	58416.7	44.1*
Item 6	57344.2	57379.1	34.9*
Item 7	57565.4	57596.4	31.0*
Item 8	57865.0	57940.0	75.0*
Item 9	56873.0	56922.0	49.0*

Table C.3 *(continued)*

Replication 4	G(1)	G(2)	Difference
Item 1	55795.9	55800.8	4.9
Item 2	56141.7	56156.2	14.5*
Item 3	55483.3	55500.8	17.5*
Item 4	55285.8	55300.9	15.1*
Item 5	55751.8	55778.8	27.0*
Item 6	54667.3	54690.1	22.8*
Item 7	54958.0	54984.2	26.2*
Item 8	55220.1	55319.2	99.1*
Item 9	54206.3	54230.1	23.8*
Replication 5	G(1)	G(2)	Difference
Item 1	55872.0	55881.8	9.8*
Item 2	56122.1	56137.3	15.2*
Item 3	55412.4	55418.8	6.4*
Item 4	55194.2	55219.1	24.9*
Item 5	55739.0	55783.3	44.3*
Item 6	54610.1	54626.0	15.9*
Item 7	54878.3	54929.6	51.3*
Item 8	55158.1	55218.9	60.8*
Item 9	54213.5	54268.9	55.4*
Replication 6	G(1)	G(2)	Difference
Item 1	56377.8	56394.8	17.0*
Item 2	56657.8	56676.4	18.6*
Item 3	56079.3	56107.8	28.5*
Item 4	55793.2	55804.1	10.9*
Item 5	56199.0	56267.5	68.5*
Item 6	55263.4	55285.4	22.0*
Item 7	55489.2	55544.9	55.7*
Item 8	55825.0	55896.2	71.2*
Item 9	54817.6	54873.7	56.1*
Replication 7	G(1)	G(2)	Difference
Item 1	58178.5	58188.6	10.1*
Item 2	58589.5	58606.1	16.6*
Item 3	57908.7	57917.6	8.9*
Item 4	57647.3	57689.1	41.8*
Item 5	58093.0	58139.2	46.2*
Item 6	57143.9	57160.4	16.5*
Item 7	57333.9	57363.9	30.0*
Item 8	57571.2	57667.0	95.8*
Item 9	56631.0	56648.9	17.9*

Table C.3 *(continued)*

Replication 8	G(1)	G(2)	Difference
Item 1	56902.0	56913_3	11.3*
Item 2	57270.8	57295_3	24.5*
Item 3	56575.2	56579_0	3.8
Item 4	56357.8	56388_3	30.5*
Item 5	56855.6	56884_7	29.1*
Item 6	55770.1	55803_7	33.6*
Item 7	56033.7	56055_9	22.2*
Item 8	56265.3	56369_8	104.5*
Item 9	55355.1	55438_9	83.8*

Replication 9	G(1)	G(2)	Difference
Item 1	56708.8	56724#.5	15.7*
Item 2	57008.2	57027#.8	19.6*
Item 3	56340.5	56351_.5	11.0*
Item 4	56178.0	56203#.3	25.3*
Item 5	56534.5	56584#.7	50.2*
Item 6	55660.4	55670#.2	9.8*
Item 7	55836.9	55882#.2	45.3*
Item 8	56004.3	56098#.1	93.8*
Item 9	55219.5	55245#.0	25.5*

Replication 10	G(1)	G(2)	Difference
Item 1	59069.1	59081_.3	12.2*
Item 2	59433.5	59454#.6	21.1*
Item 3	58729.1	58738#.9	9.8*
Item 4	58555.9	58579#.6	23.7*
Item 5	59020.9	59050#.8	29.9*
Item 6	58119.5	58155#.7	36.2*
Item 7	58203.5	58253#.3	49.8*
Item 8	58436.9	58514#.9	78.0*
Item 9	57488.4	57535#.8	47.4*

Replication 11	G(1)	G(2)	Difference
Item 1	54533.0	54543.0	10.0*
Item 2	54908.1	54929.5	21.4*
Item 3	54295.4	54304.9	9.5*
Item 4	54039.0	54067.1	28.1*
Item 5	54511.7	54546.0	34.3*
Item 6	53429.4	53452.4	23.0*
Item 7	53717.9	53755.2	37.3*
Item 8	53977.4	54077.1	99.7*
Item 9	53008.0	53047.1	39.1*

Table C.3 *(continued)*

Replication 12	G(1)	G(2)	Difference
Item 1	55744.4	55757.5	13.1*
Item 2	56139.7	56166.6	26.9*
Item 3	55482.3	55491.3	9.0*
Item 4	55246.1	55261.7	15.6*
Item 5	55775.5	55823.5	48.0*
Item 6	54691.4	54718.7	27.3*
Item 7	54948.4	54980.8	32.4*
Item 8	55217.2	55305.5	88.3*
Item 9	54217.3	54259.2	41.9*
Replication 13	G(1)	G(2)	Difference
Item 1	55125.5	55137.2	11.7*
Item 2	55506.8	55521.1	14.3*
Item 3	54895.3	54905.6	10.3*
Item 4	54601.1	54622.1	21.0*
Item 5	55042.9	55095.6	52.7*
Item 6	54051.2	54075.2	24.0*
Item 7	54261.0	54347.5	86.5*
Item 8	54550.4	54628.5	78.1*
Item 9	53553.7	53619.9	66.2*
Replication 14	G(1)	G(2)	Difference
Item 1	60265.8	60275.5	9.7*
Item 2	60603.1	60625.8	22.7*
Item 3	59955.5	59971.5	16.0*
Item 4	59732.2	59755.1	22.9*
Item 5	60124.6	60165.4	40.8*
Item 6	59144.9	59189.0	44.1*
Item 7	59384.2	59443.8	59.6*
Item 8	59707.0	59760.0	53.0*
Item 9	58637.8	58685.1	47.3*
Replication 15	G(1)	G(2)	Difference
Item 1	58211.3	58222.9	11.6*
Item 2	58638.1	58649.5	11.4*
Item 3	57898.8	57915.2	16.4*
Item 4	57655.7	57688.3	32.6*
Item 5	58170.4	58201.8	31.4*
Item 6	57214.7	57237.3	22.6*
Item 7	57377.1	57421.1	44.0*
Item 8	57657.7	57732.2	74.5*
Item 9	56648.7	56698.4	49.7*

Table C.3 (continued)

Replication 16	G(1)	G(2)	Difference
Item 1	51556.8	51566.4	9.6*
Item 2	51904.0	51911.5	7.5*
Item 3	51170.6	51185.2	14.6*
Item 4	51002.2	51026.2	24.0*
Item 5	51371.9	51405.9	34.0*
Item 6	50499.4	50534.8	35.4*
Item 7	50695.9	50738.9	43.0*
Item 8	50847.7	50909.5	61.8*
Item 9	49971.1	50007.8	36.7*
Replication 17	G(1)	G(2)	Difference
Item 1	56215.8	56232.1	16.3*
Item 2	56582.0	56607.6	25.6*
Item 3	55981.0	55983.3	2.3
Item 4	55731.2	55761.4	30.2*
Item 5	56191.7	56246.2	54.5*
Item 6	55231.8	55270.0	38.2*
Item 7	55395.6	55456.9	61.3*
Item 8	55674.5	55739.5	65.0*
Item 9	54713.2	54764.6	51.4*
Replication 18	G(1)	G(2)	Difference
Item 1	59510.4	59515.4	5.0
Item 2	59926.0	59934.1	8.1*
Item 3	59155.3	59179.3	24.0*
Item 4	58988.1	59007.5	19.4*
Item 5	59411.9	59460.6	48.7*
Item 6	58486.2	58513.3	27.1*
Item 7	58657.3	58682.2	24.9*
Item 8	58838.4	58926.6	88.2*
Item 9	57917.7	57980.4	62.7*
Replication 19	G(1)	G(2)	Difference
Item 1	56481.5	56494.7	13.2*
Item 2	56810.2	56821.6	11.4*
Item 3	56214.9	56221.4	6.5*
Item 4	55901.7	55935.4	33.7*
Item 5	56358.7	56400.8	42.1*
Item 6	55363.3	55391.3	28.0*
Item 7	55569.9	55618.6	48.7*
Item 8	55908.4	55955.1	46.7*
Item 9	54875.4	54915.6	40.2*

Table C.3 *(continued)*

Replication 20	G(1)	G(2)	Difference
Item 1	58430.8	58441.6	10.8*
Item 2	58835.4	58836.9	1.5
Item 3	58159.2	58170.2	11.0*
Item 4	57868.9	57899.7	30.8*
Item 5	58268.6	58298.6	30.0*
Item 6	57367.7	57393.7	26.0*
Item 7	57585.7	57628.2	42.5*
Item 8	57812.7	57879.3	66.6*
Item 9	56940.1	57004.7	64.6*
Replication 21	G(1)	G(2)	Difference
Item 1	57188.2	57201.7	13.5*
Item 2	57523.9	57544.3	20.4*
Item 3	56801.7	56829.5	27.8*
Item 4	56689.6	56713.2	23.6*
Item 5	57140.2	57179.4	39.2*
Item 6	56172.9	56208.5	35.6*
Item 7	56285.8	56314.7	28.9*
Item 8	56582.5	56671.6	89.1*
Item 9	55651.9	55688.9	37.0*
Replication 22	G(1)	G(2)	Difference
Item 1	56275.7	56294.3	18.6*
Item 2	56670.9	56681.9	11.0*
Item 3	55944.4	55958.0	13.6*
Item 4	55728.4	55730.0	1.6
Item 5	56248.5	56299.3	50.8*
Item 6	55072.6	55098.4	25.8*
Item 7	55360.9	55407.0	46.1*
Item 8	55635.9	55716.1	80.2*
Item 9	54716.2	54738.9	22.7*
Replication 23	G(1)	G(2)	Difference
Item 1	61365.3	61371.9	6.6*
Item 2	61797.7	61809.4	11.7*
Item 3	61100.5	61109.5	9.0*
Item 4	60805.2	60827.3	22.1*
Item 5	61246.6	61315.8	69.2*
Item 6	60356.0	60388.4	32.4*
Item 7	60430.5	60459.8	29.3*
Item 8	60767.5	60828.9	61.4*
Item 9	59800.1	59865.1	65.0*

Table C.3 *(continued)*

Replication 24	G(1)	G(2)	Difference
Item 1	53985.7	53994.7	9.0*
Item 2	54364.4	54389.5	25.1*
Item 3	53815.2	53823.6	8.4*
Item 4	53449.7	53470.1	20.4*
Item 5	53945.4	53994.2	48.8*
Item 6	52917.8	52944.3	26.5*
Item 7	53138.5	53177.9	39.4*
Item 8	53410.1	53512.7	102.6*
Item 9	52539.1	52581.1	42.0*
Replication 25	G(1)	G(2)	Difference
Item 1	61958.7	61972.5	13.8*
Item 2	62317.1	62344.8	27.7*
Item 3	61684.5	61700.7	16.2*
Item 4	61414.4	61424.3	9.9*
Item 5	61902.9	61938.5	35.6*
Item 6	60835.0	60884.6	49.6*
Item 7	61066.3	61112.6	46.3*
Item 8	61388.2	61457.5	69.3*
Item 9	60468.9	60522.9	54.0*

Note. The Critical values is 5.991 for chi-square with df = 2.
* $p < .05$

APPENDIX D

Chi-square Values of CFA Model Comparison

Table D.1

Chi-square Values of CFA Model Comparison for a DIF

Replication 1			Replication 4		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2549.7324	3.226	1	2220.1232	4.327*
2	2547.2992	0.602	2	2233.4843	2.455
3	2582.3432	5.906*	3	2224.2701	2.253
4	2564.7484	13.859*	4	2266.0047	12.246*
5	2545.7893	0.112	5	2233.6178	0.101
6	2609.1180	10.701*	6	2276.2402	31.159*
7	2587.5235	20.935*	7	2306.0119	18.623*
8	2575.5239	1.071	8	2275.6386	0.040
9	2642.5639	22.824*	9	2324.1325	24.183*
Replication 2			Replication 5		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2359.6561	0.077	1	2727.7742	0.896
2	2362.7175	1.311	2	2714.6090	0.004
3	2387.7481	9.686*	3	2742.4338	5.451*
4	2386.8969	11.675*	4	2740.1523	16.719*
5	2349.2024	0.003	5	2711.2390	3.496
6	2413.3261	20.513*	6	2773.0922	23.473*
7	2420.4279	25.052*	7	2795.7767	18.110*
8	2381.0113	0.428	8	2756.1431	1.405
9	2429.7513	33.108*	9	2801.8819	47.829*
Replication 3			Replication 6		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2028.2200	0.450	1	2435.1059	3.099
2	2027.7823	0.605	2	2436.2769	0.203
3	2042.3099	7.701	3	2460.9723	5.327*
4	2081.1116	30.478	4	2463.4860	15.160*
5	2021.8623	0.428	5	2434.8413	0.543
6	2061.7184	15.670	6	2480.2564	25.656*
7	2067.3579	15.282	7	2465.7103	5.640*
8	2046.6884	0.255	8	2446.0654	1.110
9	2093.1343	33.323	9	2507.4965	41.239*

Table D.1 *(continued)*

Replication 7			Replication 10		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2630.1757	6.724*	1	2088.9863	3.214
2	2616.8427	0.229	2	2095.3552	0.914
3	2648.8803	4.156*	3	2091.3142	0.792
4	2637.5703	10.677*	4	2112.9164	4.670*
5	2637.3163	0.014	5	2089.0782	0.018
6	2671.2152	7.902*	6	2116.6290	17.830*
7	2674.6569	14.018*	7	2149.4925	17.953*
8	2630.9853	0.814	8	2091.6242	0.013
9	2690.3472	20.624*	9	2204.2949	42.627*

Replication 8			Replication 11		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2691.8697	4.467*	1	3209.5914	3.194
2	2686.1386	0.370	2	3203.8027	1.343
3	2694.7453	2.814	3	3208.7148	2.197
4	2712.3417	14.265*	4	3223.9450	5.436*
5	2666.9239	0.644	5	3204.1438	1.519
6	2711.3945	29.301*	6	3266.5400	29.671*
7	2760.2340	21.552*	7	3250.0450	14.278*
8	2694.5068	0.635	8	3212.4557	0.726
9	2725.1765	42.620*	9	3271.5824	29.450*

Replication 9			Replication 12		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2511.1503	1.734	1	2640.9553	8.088*
2	2528.6429	0.004	2	2619.2049	1.342
3	2514.3125	5.819*	3	2644.6858	15.765*
4	2550.8042	12.166*	4	2682.8369	8.637*
5	2505.5309	0.226	5	2614.5152	1.175
6	2561.8391	16.094*	6	2648.4427	27.653*
7	2584.3767	16.441*	7	2675.2188	5.207*
8	2514.9391	0.410	8	2624.4406	0.565
9	2582.1523	10.671*	9	2715.1585	40.876*

Table D.1 *(continued)*

Replication 13			Replication 16		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	3607.9154	3.293	1	2938.4679	3.973*
2	3582.5099	0.660	2	2967.8649	0.019
3	3611.4480	5.680*	3	2968.4897	7.690*
4	3607.0156	8.451*	4	2955.9140	5.330*
5	3610.7609	2.308	5	2950.4111	1.157
6	3661.1206	32.141*	6	2981.5559	11.389*
7	3700.7885	9.120*	7	3010.0037	8.954*
8	3611.4022	0.743	8	2979.5605	0.566
9	3693.4677	51.256*	9	3009.0757	22.023*
Replication 14			Replication 17		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	1556.6577	1.681	1	2812.8467	0.815
2	1568.8828	6.301*	2	2792.7564	1.892
3	1573.7855	5.671*	3	2818.3656	1.334
4	1581.1886	8.525*	4	2822.6104	1.817
5	1560.8863	0.010	5	2809.6393	1.084
6	1620.7762	40.698*	6	2865.7605	30.473*
7	1602.6697	11.291*	7	2855.5081	8.345*
8	1552.6418	0.027	8	2819.9227	0.079
9	1610.5099	37.781*	9	2931.0857	47.663*
Replication 15			Replication 18		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2706.9591	0.431	1	2248.4826	4.064*
2	2706.0074	0.303	2	2237.9501	0.403
3	2740.1457	16.173*	3	2242.6234	2.892
4	2745.6802	10.759*	4	2277.7088	12.809*
5	2710.1972	0.194	5	2251.5297	0.001
6	2736.6505	9.284*	6	2285.4488	23.358*
7	2803.7812	16.246*	7	2270.6165	6.653*
8	2719.1671	0.101	8	2262.9237	0.193
9	2832.0896	33.275*	9	2294.1472	46.217*

Table D.1 *(continued)*

Replication 19			Replication 22		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2651.1666	6.012*	1	2617.6606	1.904
2	2657.7090	0.323	2	2615.3007	0.395
3	2666.1515	0.079	3	2617.0184	7.459*
4	2685.1412	1.074	4	2645.8628	18.403*
5	2650.9747	0.048	5	2603.3128	0.292
6	2712.9856	17.262*	6	2669.9118	32.826*
7	2748.9053	15.740*	7	2665.9999	16.646*
8	2664.5916	0.871	8	2624.4444	4.763*
9	2708.6775	27.066*	9	2714.9251	44.049*

Replication 20			Replication 23		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2394.2039	4.572*	1	2253.0191	3.517
2	2393.7339	7.473*	2	2256.9434	0.634
3	2390.8485	1.217	3	2270.2012	7.074*
4	2420.3535	15.919*	4	2303.6026	10.355*
5	2406.5028	0.924	5	2251.6031	0.035
6	2427.1928	28.816*	6	2289.5518	7.177*
7	2455.2553	10.895*	7	2310.0555	12.864*
8	2395.4582	3.938*	8	2254.6909	0.827
9	2448.2236	42.597*	9	2380.7729	49.348*

Replication 21			Replication 24		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2944.1150	3.654	1	3649.8636	6.213*
2	2934.7187	0.470	2	3658.9283	2.219
3	2949.7200	11.757*	3	3649.1682	1.594
4	2990.6519	7.834*	4	3708.8433	21.212*
5	2948.7642	2.057	5	3651.0335	0.484
6	2947.7212	25.218*	6	3685.2768	27.091*
7	3036.6619	18.976*	7	3702.3137	20.827*
8	2947.2044	0.053	8	3698.0859	0.569
9	2987.7448	39.307*	9	3761.1994	31.963*

Table D.1 *(continued)*

Replication 25		
Item	Chi-square	Chi-square difference (C1-C2)
1	2166.1009	2.069
2	2142.1306	1.125
3	2175.7681	4.194*
4	2194.6813	12.450*
5	2135.4105	0.003
6	2202.0073	30.682*
7	2192.8202	14.114*
8	2141.1551	0.189
9	2239.5649	47.703*

Note. The Critical values is 3.841 for chi-square with df = 1.
* $p<.05$

Table D.2

Chi-square Values of CFA Model Comparison for ab DIF

Replication 1			Replication 4		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2453.3481	0.083	1	2272.8482	1.080
2	2462.2240	4.564*	2	2284.9723	0.183
3	2483.0756	11.156*	3	2276.8607	9.922*
4	2476.0396	4.427*	4	2314.4128	1.822
5	2479.5739	1.918	5	2294.4282	2.456
6	2549.0329	14.593*	6	2325.4117	37.582*
7	2490.1952	0.604	7	2342.7580	3.070
8	2486.4259	3.140	8	2310.4822	12.076*
9	2619.3169	44.835*	9	2382.6468	37.206*

Replication 2			Replication 5		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2342.2539	1.088	1	2554.5540	0.003
2	2355.7944	0.033	2	2534.3502	0.038
3	2367.9619	15.418*	3	2563.5471	13.635*
4	2365.5900	3.642	4	2578.9729	3.997*
5	2338.6407	2.186	5	2559.2055	8.762*
6	2417.8169	33.783*	6	2609.2010	38.789*
7	2393.0518	3.782	7	2600.4312	4.112*
8	2361.3073	15.862*	8	2577.1681	7.566*
9	2466.3245	77.732*	9	2660.8442	65.782*

Replication 3			Replication 6		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2180.7241	0.025	1	2475.3584	0.025
2	2172.8282	1.355	2	2478.9738	0.557
3	2173.3582	10.299*	3	2512.8859	18.154*
4	2189.4603	2.222	4	2489.4769	7.469*
5	2174.7866	4.478*	5	2492.5601	11.542*
6	2225.0480	43.257*	6	2517.2984	16.057*
7	2197.3211	1.799	7	2543.0733	4.704*
8	2191.8431	4.018*	8	2491.5719	4.655*
9	2242.8497	56.451*	9	2570.6238	53.556*

Table D.2 *(continued)*

Replication 7			Replication 10		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2482.6251	0.011	1	2079.3374	1.712
2	2490.7777	0.011	2	2084.0648	8.642*
3	2499.9259	4.168*	3	2077.9117	5.259*
4	2531.5193	4.968*	4	2118.6744	0.165
5	2514.7490	11.947*	5	2077.0194	0.833
6	2543.5697	19.823*	6	2095.6389	22.938*
7	2522.4238	11.188*	7	2135.2931	0.267
8	2503.9938	6.132*	8	2096.0195	14.227*
9	2579.0925	41.580*	9	2176.1894	61.904*

Replication 8			Replication 11		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2854.7505	0.485	1	3258.1714	2.722
2	2845.8191	0.330	2	3254.4797	2.297
3	2852.2253	9.060*	3	3295.6543	16.057*
4	2857.6496	4.797*	4	3262.4452	0.045
5	2850.4953	2.360	5	3254.9232	5.596*
6	2905.6919	51.044*	6	3345.3227	54.284*
7	2901.8008	4.723*	7	3283.6474	2.824
8	2882.0815	4.240*	8	3284.3036	19.921*
9	2938.8355	70.700*	9	3355.4460	52.096*

Replication 9			Replication 12		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2419.9335	0.000	1	2502.5652	0.835
2	2419.4669	0.324	2	2513.3015	5.690*
3	2440.3488	6.053*	3	2520.2441	16.020*
4	2430.3442	0.566	4	2553.6573	2.586
5	2433.6614	7.598*	5	2520.0382	3.158
6	2480.8134	16.703*	6	2560.6812	30.326*
7	2450.1536	3.852*	7	2562.7833	0.399
8	2428.2449	1.213	8	2523.4002	12.171*
9	2549.5680	38.558*	9	2597.2803	34.591*

Table D.2 *(continued)*

Replication 13			Replication 16		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	3463.0131	1.730	1	3052.5133	0.032
2	3456.6810	0.727	2	3061.3688	1.189
3	3467.0032	9.581*	3	3069.1056	11.604*
4	3491.1178	3.110	4	3073.9824	1.522
5	3462.0150	5.689*	5	3052.5655	1.283
6	3521.6939	46.115*	6	3085.9949	38.624*
7	3514.7542	1.112	7	3079.6477	2.076
8	3494.8335	9.877*	8	3080.2963	4.893*
9	3587.1699	80.577*	9	3198.2710	46.720*

Replication 14			Replication 17		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	1677.6624	1.084	1	2738.7244	0.327
2	1677.4466	3.581	2	2737.6081	1.597
3	1697.7623	13.570*	3	2741.3124	1.834
4	1649.2130	1.373	4	2734.7964	0.126
5	1655.2398	0.404	5	2746.9887	0.895
6	1731.1371	69.162*	6	2787.4479	22.744*
7	1694.7005	1.837	7	2777.8440	0.583
8	1680.5613	6.524*	8	2775.2729	8.388*
9	1753.9996	61.389*	9	2833.4983	43.621*

Replication 15			Replication 18		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2765.7635	5.387*	1	2291.7398	2.469
2	2763.7813	1.076	2	2269.5272	1.804
3	2780.2720	4.342*	3	2279.3453	10.923*
4	2818.4588	7.724*	4	2287.7954	1.553
5	2783.7064	5.401*	5	2263.4929	0.919
6	2842.1591	21.604*	6	2302.1363	33.463*
7	2827.6840	1.243	7	2286.4336	2.252
8	2788.4358	11.469*	8	2277.4599	1.553
9	2928.2470	54.947*	9	2329.7245	70.630*

Table D.2 *(continued)*

Replication 19			Replication 22		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2773.7566	0.062	1	2562.5395	1.698
2	2766.6051	1.386	2	2582.3690	0.422
3	2796.7404	1.912	3	2583.9493	17.171*
4	2803.5071	2.241	4	2584.7213	9.594*
5	2787.2343	0.208	5	2569.8837	10.789*
6	2855.5744	39.076*	6	2660.1734	57.647*
7	2827.2648	4.227*	7	2628.2386	7.274*
8	2786.0355	1.430	8	2606.3358	5.925*
9	2917.3050	48.002*	9	2679.4352	35.694*

Replication 20			Replication 23		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2383.8879	1.842	1	2394.4344	3.714
2	2389.6049	0.148	2	2379.9464	1.196
3	2404.6497	4.447*	3	2413.6261	19.620*
4	2430.5947	0.600	4	2392.7263	3.161
5	2413.8257	6.773*	5	2380.7955	3.745
6	2474.6654	42.404*	6	2426.2856	38.959*
7	2444.0263	1.935	7	2417.0019	9.064*
8	2423.0724	4.611*	8	2389.6664	0.612
9	2449.5028	44.946*	9	2501.1538	61.920*

Replication 21			Replication 24		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2946.9847	0.007	1	3290.4016	7.400*
2	2957.1980	2.776	2	3275.4411	0.664
3	2958.9757	16.353*	3	3271.0907	4.025*
4	2976.1584	2.875	4	3298.7953	3.271
5	2952.2420	7.089*	5	3273.1274	2.325
6	2981.6403	50.829*	6	3363.4256	53.487*
7	3009.3997	5.662*	7	3335.5929	10.527*
8	2986.6896	8.438*	8	3319.0573	13.063*
9	3017.2245	52.225*	9	3360.2079	42.095*

Table D.2 *(continued)*

Replication 25		
Item	Chi-square	Chi-square difference (C1-C2)
1	2224.1824	0.008
2	2253.7175	5.103*
3	2248.8253	8.299*
4	2265.7548	3.707
5	2232.6851	4.002*
6	2327.0077	52.122*
7	2257.8267	7.515*
8	2228.8208	3.756
9	2357.3521	54.262*

Note. The Critical values is 3.841 for chi-square with df=1.
* $p<.05$

Table D.3

Chi-square Values of CFA Model Comparison for b DIF

Replication 1			Replication 4		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2522.9126	9.765*	1	2231.0075	4.433*
2	2523.1977	24.096*	2	2252.4091	14.361*
3	2537.8194	14.179*	3	2274.3482	17.430*
4	2529.7713	8.044*	4	2318.4963	12.668*
5	2555.9270	35.570*	5	2280.9038	25.016*
6	2584.8368	10.252*	6	2281.9191	21.021*
7	2552.5051	22.063*	7	2323.2292	19.147*
8	2617.2849	90.108*	8	2430.6749	95.267*
9	2715.1524	20.142*	9	2367.5725	19.213*
Replication 2			Replication 5		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2292.6649	12.343*	1	2733.6294	8.552*
2	2294.1783	6.560*	2	2741.9128	13.079*
3	2315.4591	8.456*	3	2771.5889	6.482*
4	2304.1461	12.895*	4	2766.6330	17.450*
5	2316.3244	34.377*	5	2802.1651	39.527*
6	2344.1218	24.564*	6	2808.2904	16.669*
7	2353.8301	11.067*	7	2809.2229	19.914*
8	2424.3334	109.978*	8	2893.6117	57.536*
9	2453.8869	61.325*	9	2918.2269	52.654*
Replication 3			Replication 6		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2156.4693	19.100*	1	2454.1038	17.099*
2	2160.8193	24.281*	2	2459.9650	18.328*
3	2148.4883	19.435*	3	2485.7099	28.014*
4	2169.5843	8.764*	4	2497.0353	10.469*
5	2193.7348	39.929*	5	2531.2230	68.014*
6	2237.4125	33.933*	6	2541.2131	19.926*
7	2177.5008	26.384*	7	2519.1612	30.429*
8	2299.5271	71.171*	8	2561.8310	70.931*
9	2287.5514	44.162*	9	2639.4179	47.069*

Table D.3 *(continued)*

Replication 7			Replication 10		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2699.9901	8.783*	1	2138.4159	9.852*
2	2697.1186	13.953*	2	2163.7252	21.254*
3	2717.2428	8.265*	3	2150.4572	7.813*
4	2722.2483	29.766*	4	2156.1273	19.503*
5	2736.0925	41.792*	5	2154.9188	29.506*
6	2769.2332	15.707*	6	2199.9421	34.659*
7	2718.9090	16.951*	7	2204.2114	28.161*
8	2809.0700	82.015*	8	2234.7105	60.186*
9	2779.3457	16.012*	9	2298.6076	39.065*

Replication 8			Replication 11		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2854.6234	9.120*	1	3123.2328	5.474*
2	2838.8292	23.999*	2	3129.8751	19.787*
3	2822.5082	0.518	3	3112.3855	7.181*
4	2855.4907	23.360*	4	3144.1328	26.867*
5	2857.6763	27.373*	5	3129.7046	33.572*
6	2869.8027	31.849*	6	3231.2584	20.437*
7	2865.6688	15.678*	7	3148.5659	22.608*
8	2990.0255	95.579*	8	3262.7466	93.935*
9	2950.6997	66.381*	9	3293.4311	33.188*

Replication 9			Replication 12		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2438.5697	8.012*	1	2630.4863	13.512*
2	2468.9488	19.022*	2	2670.4555	27.192*
3	2460.6809	10.311*	3	2647.6004	8.839*
4	2463.1739	17.602*	4	2676.0706	14.493*
5	2482.6602	47.960*	5	2689.5567	46.910*
6	2475.2180	9.087*	6	2694.6123	25.883*
7	2513.2397	24.593*	7	2688.0265	24.835*
8	2577.0280	91.192*	8	2783.7581	80.038*
9	2559.1200	20.030*	9	2832.0318	42.119*

Table D.3 *(continued)*

Replication 13			Replication 16		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	3346.2703	7.548*	1	2768.3813	9.485*
2	3342.6463	12.595*	2	2769.9153	7.361*
3	3370.8075	10.215*	3	2769.0384	14.716*
4	3387.6931	17.169*	4	2803.6157	14.785*
5	3375.1724	42.518*	5	2800.0684	29.790*
6	3406.8585	24.020*	6	2836.7180	34.743*
7	3434.5202	38.744*	7	2817.0846	26.545*
8	3427.5514	61.020*	8	2885.9006	52.665*
9	3516.4282	65.158*	9	2940.5603	33.280*

Replication 14			Replication 17		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	1601.1789	9.978*	1	2741.4109	14.664*
2	1602.8522	23.406*	2	2747.5270	26.192*
3	1587.0981	12.148*	3	2727.6487	2.414
4	1596.8814	21.004*	4	2746.9378	27.144*
5	1634.7986	42.388*	5	2789.9592	51.615*
6	1649.2167	39.447*	6	2822.5088	36.374*
7	1640.1464	34.321*	7	2795.7339	36.922*
8	1660.7026	53.531*	8	2829.9628	62.917*
9	1653.9079	38.126*	9	2904.2070	45.293*

Replication 15			Replication 18		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2819.6907	9.838*	1	2284.4975	4.797*
2	2800.1847	10.928*	2	2267.4747	7.959*
3	2814.1801	13.711*	3	2281.8337	21.201*
4	2835.5841	18.770*	4	2289.6748	17.565*
5	2842.9298	32.076*	5	2294.2124	48.076*
6	2890.0378	19.320*	6	2326.3549	22.232*
7	2852.0594	23.365*	7	2299.5288	20.098*
8	2943.0251	75.132*	8	2413.7919	75.103*
9	2954.4259	36.567*	9	2415.3639	58.880*

Table D.3 *(continued)*

Replication 19			Replication 22		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2726.8062	10.303*	1	2601.6877	18.378*
2	2727.4015	9.644*	2	2588.2935	7.720*
3	2769.5512	6.300*	3	2582.7217	13.469*
4	2787.9491	26.047*	4	2604.9466	1.904
5	2779.2417	41.373*	5	2615.8122	46.169*
6	2810.3751	26.932*	6	2687.1559	25.727*
7	2791.6391	27.257*	7	2630.4691	23.304*
8	2822.1048	41.204*	8	2712.1101	67.832*
9	2909.0998	33.893*	9	2736.3113	23.277*

Replication 20			Replication 23		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	2337.9432	10.592*	1	2406.4632	6.127*
2	2325.4276	0.585	2	2395.4926	10.969*
3	2351.8164	11.101*	3	2414.7568	8.800*
4	2397.1995	24.332*	4	2413.6586	17.016*
5	2401.8112	28.602*	5	2463.7617	67.817*
6	2430.1992	25.395*	6	2460.3800	30.410*
7	2395.6569	27.066*	7	2435.7753	20.855*
8	2484.6712	63.334*	8	2476.4266	57.642*
9	2454.8657	57.662*	9	2554.4009	55.619*

Replication 21			Replication 24		
Item	Chi-square	Chi-square difference (C1-C2)	Item	Chi-square	Chi-square difference (C1-C2)
1	3046.4546	11.915*	1	3475.7696	6.141*
2	3059.9759	19.857*	2	3491.8829	24.381*
3	3046.8355	26.425*	3	3466.1145	8.085*
4	3077.7630	16.677*	4	3551.0202	16.083*
5	3115.6636	35.709*	5	3533.0609	48.651*
6	3058.8130	34.193*	6	3519.1359	26.032*
7	3080.5083	20.822*	7	3516.0461	17.548*
8	3235.6530	81.284*	8	3658.2825	96.762*
9	3134.0550	31.458*	9	3671.7224	34.800*

Table D.3 *(continued)*

Replication 25		
Item	Chi-square	Chi-square difference (C1-C2)
1	2374.0318	12.207*
2	2385.5246	25.291*
3	2394.6239	14.860*
4	2403.3910	9.368*
5	2385.2962	36.238*
6	2479.5995	44.064*
7	2414.8294	24.028*
8	2439.4274	63.380*
9	2506.9123	41.113*

Note. The Critical values is 3.841 for chi-square with df = 1.
* $p<.05$

APPENDIX E

Modification Index of Total, High, and Low Groups

Table E.1

Modification Index of Total, High, and Low Groups for a DIF

Replication 1	MI-total	MI-high	MI-low
Item 1	3.288	8.796*	0.029
Item 2	0.595	1.189	5.464*
Item 3	5.925*	0.744	10.239*
Item 4	14.132*	5.942*	7.903*
Item 5	0.129	3.214	2.466
Item 6	10.821*	8.090*	4.363*
Item 7	21.617*	2.528	19.495*
Item 8	1.132	10.368*	29.380*
Item 9	22.986*	22.795*	2.702
Replication 2	MI-total	MI-high	MI-low
Item 1	0.070	3.720	3.710
Item 2	1.379	3.430	0.041
Item 3	9.697*	2.645	10.361*
Item 4	12.041*	11.449*	3.639
Item 5	0.007	9.642*	9.484*
Item 6	20.677*	11.752*	15.420*
Item 7	25.937*	2.942	24.856*
Item 8	0.410	11.257*	25.473*
Item 9	33.246*	32.639*	1.865
Replication 3	MI-total	MI-high	MI-low
Item 1	0.438	3.659	0.337
Item 2	0.590	5.029*	1.123
Item 3	7.891*	2.761	5.840*
Item 4	30.936*	11.101*	20.669*
Item 5	0.414	17.412*	9.729*
Item 6	15.992*	5.513*	18.832*
Item 7	15.454*	0.644	16.767*
Item 8	0.238	19.588*	15.212*
Item 9	33.817*	30.095*	4.033*

Table E.1 *(continued)*

Replication 4	MI-total	MI-high	MI-low
Item 1	4.384*	8.869*	0.096
Item 2	2.509	3.514	0.051
Item 3	2.273	0.550	2.747
Item 4	12.526*	3.145	8.478*
Item 5	0.111	5.563*	3.976*
Item 6	31.534*	15.316*	22.597*
Item 7	19.101*	0.160	24.037*
Item 8	0.040	7.076*	10.830*
Item 9	24.446*	18.354*	6.956*
Replication 5	MI-total	MI-high	MI-low
Item 1	0.942	0.188	1.038
Item 2	0.001	1.714	1.399
Item 3	5.415*	3.347	1.893
Item 4	17.217*	11.820*	8.857*
Item 5	3.825	7.757*	0.008
Item 6	23.533*	11.791*	16.685*
Item 7	18.683*	2.250	19.290*
Item 8	1.386	11.598*	21.614*
Item 9	48.089*	38.690*	9.038*
Replication 6	MI-total	MI-high	MI-low
Item 1	3.267	1.624	1.602
Item 2	0.254	2.044	0.536
Item 3	5.199*	3.093	1.880
Item 4	15.705*	4.532*	11.420*
Item 5	0.459	13.244*	24.738*
Item 6	25.563*	15.041*	14.383*
Item 7	5.965*	0.414	6.725*
Item 8	0.964	6.919*	17.938*
Item 9	41.119*	33.324*	9.911*
Replication 7	MI-total	MI-high	MI-low
Item 1	6.679*	0.301	7.875*
Item 2	0.209	1.560	0.279
Item 3	4.263*	0.762	4.969*
Item 4	10.731*	7.028*	5.457*
Item 5	0.026	5.297*	5.655*
Item 6	8.165*	1.491	15.839*
Item 7	14.119*	0.773	16.708*
Item 8	0.781	20.333*	10.923*
Item 9	21.063*	17.090*	7.623*

Table E.1 *(continued)*

Replication 8	MI-total	MI-high	MI-low
Item 1	4.592*	7.490*	0.401
Item 2	0.400	2.513	0.450
Item 3	2.782	0.440	11.766*
Item 4	14.675*	2.613	12.451*
Item 5	0.607	0.480	3.089
Item 6	29.353*	10.140*	28.220*
Item 7	22.176*	1.709	23.026*
Item 8	0.592	22.114*	30.583*
Item 9	42.755*	31.295*	12.937*
Replication 9	MI-total	MI-high	MI-low
Item 1	1.732	2.524	0.195
Item 2	0.006	0.203	0.221
Item 3	5.931*	0.979	7.499*
Item 4	12.388*	10.296*	4.610*
Item 5	0.254	4.331*	8.105*
Item 6	16.398*	8.351*	9.926*
Item 7	16.735*	0.868	18.536*
Item 8	0.418	18.261*	12.313*
Item 9	10.841*	5.795*	8.645*
Replication 10	MI-total	MI-high	MI-low
Item 1	3.357	1.080	2.312
Item 2	0.876	0.027	1.231
Item 3	0.760	2.378	14.175*
Item 4	4.896*	7.845*	1.290
Item 5	0.008	6.660*	5.942*
Item 6	17.838*	4.111*	25.333*
Item 7	18.606*	6.434*	15.631*
Item 8	0.002	23.857*	22.181*
Item 9	42.638*	43.662*	1.631
Replication 11	MI-total	MI-high	MI-low
Item 1	3.340	8.299*	0.001
Item 2	1.303	0.714	6.118*
Item 3	2.159	0.101	9.917*
Item 4	5.677*	6.934*	0.916
Item 5	1.715	14.827*	4.655*
Item 6	29.820*	21.584*	11.372*
Item 7	14.777*	1.785	14.239*
Item 8	0.685	17.418*	34.565*
Item 9	29.495*	24.428*	9.804*

Table E.1 *(continued)*

Replication 12	MI-total	MI-high	MI-low
Item 1	8.592*	1.063	5.806*
Item 2	1.167	0.060	4.608*
Item 3	15.403	5.637*	11.580*
Item 4	9.462*	5.523*	8.380*
Item 5	0.931	2.425	8.752*
Item 6	27.255*	12.540*	11.763*
Item 7	5.769*	2.646	6.656*
Item 8	0.330	1.523	10.282*
Item 9	40.340*	27.788*	9.266*
Replication 13	MI-total	MI-high	MI-low
Item 1	3.492	8.481*	0.034
Item 2	0.603	0.615	3.337
Item 3	5.578*	1.096	6.647*
Item 4	8.854*	11.458*	1.984
Item 5	2.182	2.776	13.791*
Item 6	32.146*	16.280*	23.787*
Item 7	9.676*	2.349	8.193*
Item 8	1.041	26.818*	15.688*
Item 9	51.136*	40.937*	11.937*
Replication 14	MI-total	MI-high	MI-low
Item 1	1.811	2.431	0.249
Item 2	6.159*	0.296	8.915*
Item 3	5.522*	1.217	7.071*
Item 4	8.916*	8.683*	3.113
Item 5	0.000	11.251*	10.842*
Item 6	40.569*	25.448*	20.487*
Item 7	11.833*	1.312	13.116*
Item 8	0.086	14.455*	12.598*
Item 9	37.636*	34.715*	5.710*
Replication 15	MI-total	MI-high	MI-low
Item 1	6.679*	0.476	0.103
Item 2	0.209	0.272	4.101*
Item 3	4.263*	16.125*	11.481*
Item 4	10.731*	11.230*	4.169*
Item 5	0.026	0.151*	8.160*
Item 6	8.165*	9.210*	9.450*
Item 7	14.119*	16.885*	10.828*
Item 8	0.781	0.058	26.822*
Item 9	21.063*	33.155*	10.221*

Table E.1 *(continued)*

Replication 16	MI-total	MI-high	MI-low
Item 1	4.592*	4.067*	0.005
Item 2	0.400	0.014	2.249
Item 3	2.782	7.698*	5.195*
Item 4	14.675*	5.553*	0.586
Item 5	0.607	1.150	24.605*
Item 6	29.353*	11.442*	12.633*
Item 7	22.176*	9.248	6.767*
Item 8	0.592	0.691	19.052*
Item 9	42.755*	22.078*	6.556*
Replication 17	MI-total	MI-high	MI-low
Item 1	1.732	0.931	0.003
Item 2	0.006	1.720	3.633
Item 3	5.931*	1.215	5.256*
Item 4	12.388*	2.049	1.407
Item 5	0.254	0.899	13.542*
Item 6	16.398*	30.187*	15.636*
Item 7	16.735*	8.945*	8.937*
Item 8	0.418	0.209	13.497*
Item 9	10.841*	47.344*	10.283*
Replication 18	MI-total	MI-high	MI-low
Item 1	4.204*	6.329*	0.428
Item 2	0.449	0.483	0.097
Item 3	2.834	0.930	2.524
Item 4	13.312*	6.564*	7.569*
Item 5	0.009	3.407	3.275
Item 6	23.406*	11.317*	18.451*
Item 7	7.008*	0.705	6.995*
Item 8	0.144	23.243*	30.424*
Item 9	46.208*	39.678*	6.850*
Replication 19	MI-total	MI-high	MI-low
Item 1	6.017*	3.151	2.569
Item 2	0.308	1.459	0.158
Item 3	0.090	0.005	0.287
Item 4	1.054	1.014	0.308
Item 5	0.037	9.454*	8.183*
Item 6	17.665*	5.531*	21.366*
Item 7	16.044*	0.814	18.188*
Item 8	0.877	8.543*	3.640
Item 9	27.620*	18.052*	12.437*

Table E.1 (continued)

Replication 20	MI-total	MI-high	MI-low
Item 1	4.480*	4.332*	1.277
Item 2	7.360*	10.344*	0.580
Item 3	1.313	0.006	3.786
Item 4	15.900*	13.874*	6.505*
Item 5	0.831	14.239*	5.882*
Item 6	29.775*	16.462*	17.446*
Item 7	10.866*	0.430	13.200*
Item 8	3.902*	29.769*	6.470*
Item 9	43.560*	39.039*	6.018*
Replication 21	MI-total	MI-high	MI-low
Item 1	3.866*	2.316	1.813
Item 2	0.394	0.290	1.797
Item 3	11.592*	2.741	11.544*
Item 4	8.327*	8.037*	3.192
Item 5	1.887	3.242	12.186*
Item 6	25.038*	13.301*	18.052*
Item 7	19.857*	1.312	22.401*
Item 8	0.151	14.874*	9.316*
Item 9	39.066*	27.980*	15.369*
Replication 22	MI-total	MI-high	MI-low
Item 1	1.904	1.691	0.532
Item 2	0.387	1.336	0.069
Item 3	7.602*	2.400	7.364*
Item 4	18.657*	5.638*	13.706*
Item 5	0.332	6.811*	12.520*
Item 6	33.293*	18.082*	19.038*
Item 7	16.966*	0.814	18.730*
Item 8	5.121*	22.825*	4.436*
Item 9	44.743*	47.142*	1.175
Replication 23	MI-total	MI-high	MI-low
Item 1	3.556	2.980	0.761
Item 2	0.644	0.229	0.572
Item 3	7.147*	3.949*	3.261
Item 4	10.597*	9.411*	3.489
Item 5	0.036	3.382	5.868*
Item 6	7.297*	1.845	11.815*
Item 7	13.219*	5.417*	9.053*
Item 8	0.933	17.804*	10.489*
Item 9	49.779*	44.099*	6.081*

Table E.1 *(continued)*

Replication 24	MI-total	MI-high	MI-low
Item 1	6.318*	3.681	2.740
Item 2	2.224	0.150	3.346
Item 3	1.586	0.927	1.095
Item 4	21.769*	13.246*	10.672*
Item 5	0.542	6.198*	2.841
Item 6	27.287*	14.524*	18.676*
Item 7	21.366*	3.992*	19.511*
Item 8	0.581	13.704*	27.042*
Item 9	32.282*	32.395*	1.743
Replication 25	MI-total	MI-high	MI-low
Item 1	2.148	9.023*	0.283
Item 2	1.216	1.178	7.262*
Item 3	4.160*	0.084	9.957*
Item 4	12.924*	10.206*	5.132*
Item 5	0.012	6.022*	5.717*
Item 6	30.739*	16.824*	16.659*
Item 7	14.671*	5.767*	9.262*
Item 8	0.274	18.415*	15.938*
Item 9	47.935*	43.380*	4.931*

Note. The Critical values is 3.841 for chi-square with df = 1.
* $p<.05$

Table E.2

Modification Index of Total, High, and Low Groups of ab DIF

Replication 1	MI-total	MI-high	MI-low
Item 1	0.029	1.599	1.600
Item 2	4.059*	1.173	3.636
Item 3	10.509*	3.187	10.693*
Item 4	5.032*	3.529	1.987
Item 5	1.430	0.975	8.287*
Item 6	13.819*	9.756*	5.378*
Item 7	0.854	1.598	0.113
Item 8	2.271	2.989	21.238*
Item 9	43.754*	42.748*	4.336*
Replication 2	MI-total	MI-high	MI-low
Item 1	1.428	4.537*	0.067
Item 2	0.142	1.573	0.554
Item 3	14.324*	6.565*	9.476*
Item 4	4.502*	4.634*	1.166
Item 5	1.458	7.991*	21.830*
Item 6	32.203*	15.646*	25.553*
Item 7	4.623*	5.166*	1.255
Item 8	13.767*	3.243	55.713*
Item 9	75.667*	74.863*	3.644
Replication 3	MI-total	MI-high	MI-low
Item 1	0.074	3.800	1.142
Item 2	1.049	0.292	3.299
Item 3	9.599*	4.079*	5.743*
Item 4	2.777	1.765	1.366
Item 5	3.649	8.512*	27.916*
Item 6	41.985*	22.299*	26.808*
Item 7	2.183	0.121	2.176
Item 8	2.977	10.051*	30.239*
Item 9	55.190*	44.187*	12.882*

Table E.2 *(continued)*

Replication 4	MI-total	MI-high	MI-low
Item 1	1.346	1.629	0.132
Item 2	0.338	0.791	0.033
Item 3	9.208*	2.006	10.941*
Item 4	2.294	1.811	0.936
Item 5	1.811	1.388	10.557*
Item 6	36.318*	21.287*	21.042*
Item 7	3.570	0.426	4.038*
Item 8	10.658*	0.666	30.127*
Item 9	35.963*	26.953*	12.196*
Replication 5	MI-total	MI-high	MI-low
Item 1	0.004	0.354	0.052
Item 2	0.000	3.175	2.306
Item 3	12.868*	5.761*	7.022*
Item 4	4.753*	7.433*	1.309
Item 5	7.525*	0.074	13.578*
Item 6	37.294*	15.622*	32.014*
Item 7	4.895*	0.760	4.982*
Item 8	6.033*	3.334	20.063*
Item 9	63.899*	60.390*	6.468*
Replication 6	MI-total	MI-high	MI-low
Item 1	0.005	1.227	1.093
Item 2	0.377	0.001	0.791
Item 3	17.458*	12.753*	4.850*
Item 4	8.270*	3.232	5.490*
Item 5	10.604*	0.196	27.332*
Item 6	15.331*	7.313*	13.104*
Item 7	5.374*	0.739	6.288*
Item 8	3.789	5.410*	29.757*
Item 9	52.582*	40.308*	14.536*
Replication 7	MI-total	MI-high	MI-low
Item 1	0.000	0.071	0.046
Item 2	0.001	0.902	0.735
Item 3	3.765	3.529	0.533
Item 4	5.709*	2.405	3.615
Item 5	11.003*	0.157	23.466*
Item 6	18.899*	11.915*	9.399*
Item 7	12.149*	3.810	9.391*
Item 8	5.112*	17.484*	50.313*
Item 9	40.544*	32.303*	11.895*

Table E.2 *(continued)*

Replication 8	MI-total	MI-high	MI-low
Item 1	0.684	0.502	0.206
Item 2	0.186	0.309	0.013
Item 3	8.365*	2.425	7.123*
Item 4	5.494*	1.401	4.103*
Item 5	1.742	0.040	4.253*
Item 6	49.807*	27.101*	27.161*
Item 7	5.332*	2.389	3.113
Item 8	3.310	16.208*	39.481*
Item 9	69.293*	57.556*	12.391*
Replication 9	MI-total	MI-high	MI-low
Item 1	0.009	3.242	1.728
Item 2	0.205	0.820	2.313
Item 3	5.603*	1.332	5.855*
Item 4	0.780	2.919	0.011
Item 5	6.848*	2.347	26.287*
Item 6	16.030*	12.217*	3.682
Item 7	4.367*	0.975	3.121
Item 8	0.723	14.117*	25.483*
Item 9	37.708*	26.330*	14.905*
Replication 10	MI-total	MI-high	MI-low
Item 1	2.167	4.404*	0.080
Item 2	7.739*	0.938	8.245*
Item 3	4.550*	0.119	17.889*
Item 4	0.365	2.724	0.008
Item 5	0.359	0.954	3.311
Item 6	21.526*	8.720*	20.724*
Item 7	0.491	0.451	0.497
Item 8	12.167*	16.283*	66.860*
Item 9	59.998*	54.335*	9.103*
Replication 11	MI-total	MI-high	MI-low
Item 1	3.422	10.066*	0.013
Item 2	1.731	0.814	8.218*
Item 3	14.659*	6.745*	9.538*
Item 4	0.201	0.120	0.162
Item 5	4.174*	1.444	16.546*
Item 6	52.012*	40.567*	15.102*
Item 7	3.612	0.490	3.900*
Item 8	17.184*	1.335	50.670*
Item 9	49.841*	47.037*	8.130*

Table E.2 *(continued)*

Replication 12	MI-total	MI-high	MI-low
Item 1	1.121	1.063	0.299
Item 2	4.952*	0.060	10.937*
Item 3	15.020*	5.637*	10.863*
Item 4	3.237	5.523*	0.407
Item 5	2.255	2.425	13.664*
Item 6	29.067*	12.540*	22.206*
Item 7	0.638	2.646	0.020
Item 8	10.321*	1.523	30.756*
Item 9	32.965*	27.788*	7.271*
Replication 13	**MI-total**	**MI-high**	**MI-low**
Item 1	2.229	7.109*	0.026
Item 2	0.449	0.558	2.824
Item 3	8.747*	1.969	10.016*
Item 4	3.880*	4.692*	1.051
Item 5	4.475*	8.220*	33.015*
Item 6	44.102*	26.044*	24.373*
Item 7	1.597	3.891*	0.196
Item 8	7.802*	3.055	32.056*
Item 9	78.359*	72.897*	8.769*
Replication 14	**MI-total**	**MI-high**	**MI-low**
Item 1	1.360	1.629	0.031
Item 2	3.096	0.791	5.026*
Item 3	12.689*	2.006	9.662*
Item 4	1.715	1.811	0.033
Item 5	0.150	1.388	8.770*
Item 6	67.900*	21.287*	37.373*
Item 7	2.283	0.426	3.686
Item 8	5.231*	0.666	27.324*
Item 9	59.878*	26.953*	5.588*
Replication 15	**MI-total**	**MI-high**	**MI-low**
Item 1	6.047*	2.085	4.026*
Item 2	0.803	0.155	0.706
Item 3	3.903*	2.204	1.778
Item 4	8.826*	4.294*	5.375*
Item 5	4.526*	0.495	12.446*
Item 6	20.549*	6.924*	25.329*
Item 7	1.639	0.000	2.391
Item 8	10.087*	2.941	31.834*
Item 9	53.567*	49.087*	9.236*

Table E.2 *(continued)*

Replication 16	MI-total	MI-high	MI-low
Item 1	0.007	6.430*	3.486
Item 2	0.908	0.893	4.299*
Item 3	10.931*	2.444	12.000*
Item 4	2.005	13.540*	0.184
Item 5	0.782	10.983*	17.203*
Item 6	37.438*	20.399*	22.177*
Item 7	2.600	0.875	2.173
Item 8	3.788	16.181*	36.917*
Item 9	45.321*	41.403*	10.624*
Replication 17	MI-total	MI-high	MI-low
Item 1	0.492	0.943	0.080
Item 2	1.263	0.038	2.593
Item 3	1.527	0.000	3.899*
Item 4	0.245	0.537	0.046
Item 5	0.524	4.683*	8.123*
Item 6	21.795*	13.210*	9.955*
Item 7	0.837	0.009	1.209
Item 8	7.197*	3.566	25.781*
Item 9	42.525*	39.485*	5.916*
Replication 18	MI-total	MI-high	MI-low
Item 1	2.843	1.777	1.027
Item 2	1.529	0.454	1.266
Item 3	10.380*	0.727	17.732*
Item 4	1.882	2.444	0.351
Item 5	0.602	4.859*	12.774*
Item 6	32.556*	18.030*	21.145*
Item 7	2.682	3.960*	1.016
Item 8	0.947	9.119*	21.926*
Item 9	69.730*	60.354*	10.883*
Replication 19	MI-total	MI-high	MI-low
Item 1	0.124	3.006	0.758
Item 2	1.129	0.920	0.323
Item 3	1.651	0.135	2.679
Item 4	2.716	5.625*	0.188
Item 5	0.066	4.069*	5.683*
Item 6	38.152*	19.821*	25.171*
Item 7	4.891*	0.927	4.142*
Item 8	0.900	8.807*	18.113*
Item 9	46.998*	43.273*	7.117*

Table E.2 *(continued)*

Replication 20	MI-total	MI-high	MI-low
Item 1	2.175	2.799	0.365
Item 2	0.051	1.309	1.690
Item 3	3.934*	0.278	6.765*
Item 4	0.895	4.633*	0.000
Item 5	5.760*	2.985	23.236*
Item 6	41.051*	29.673*	14.693*
Item 7	2.411	0.175	3.323
Item 8	3.427	7.506*	24.315*
Item 9	43.586*	35.969*	10.924*
Replication 21	MI-total	MI-high	MI-low
Item 1	0.005	0.485	0.170
Item 2	2.062	0.672	1.560
Item 3	15.270*	8.564*	7.201*
Item 4	3.661	2.782	1.812
Item 5	5.574*	0.386	14.357*
Item 6	48.812*	22.558*	37.091*
Item 7	6.735*	0.451	8.332*
Item 8	6.498*	12.281*	41.999*
Item 9	49.849*	46.012*	8.885*
Replication 22	MI-total	MI-high	MI-low
Item 1	2.035	5.727*	0.002
Item 2	0.255	0.227	1.255
Item 3	16.439*	2.407	22.410*
Item 4	10.610*	5.636*	6.141*
Item 5	9.750*	2.876	36.370*
Item 6	56.489*	38.181*	21.556*
Item 7	8.383*	3.578	6.458*
Item 8	4.777*	10.265*	36.523*
Item 9	34.535*	31.169*	6.378*
Replication 23	MI-total	MI-high	MI-low
Item 1	4.133*	6.275*	0.159
Item 2	0.948	0.389	1.080
Item 3	18.849*	13.709*	6.761*
Item 4	3.873*	3.741	0.991
Item 5	3.089	0.528	15.373*
Item 6	38.120*	26.434*	17.226*
Item 7	10.208*	2.698	7.948*
Item 8	0.229	5.003*	14.516*
Item 9	60.749*	55.385*	6.947*

Table E.2 *(continued)*

Replication 24	MI-total	MI-high	MI-low
Item 1	8.107*	5.177*	3.759
Item 2	0.443	0.198	1.710
Item 3	3.647	1.145	2.946
Item 4	3.818	12.815*	0.043
Item 5	1.734	2.742	10.986*
Item 6	52.305*	37.854*	18.766*
Item 7	11.651*	2.079	11.240*
Item 8	11.838*	6.522*	52.445*
Item 9	40.996*	35.119*	8.842*
Replication 25	MI-total	MI-high	MI-low
Item 1	0.003	4.074*	2.558
Item 2	4.393*	0.169	11.536*
Item 3	7.494*	2.980	5.638*
Item 4	4.474*	7.154*	0.611
Item 5	3.060	1.751	14.850*
Item 6	50.543*	37.358*	13.558*
Item 7	8.667*	6.040*	3.820
Item 8	2.522	6.944*	27.807*
Item 9	52.574*	52.674*	1.395

Note. The Critical values is 3.841 for chi-square with df = 1.
* $p<.05$

Table E.3

Modification Index of Total, High, and Low Groups of b DIF

Replication 1	MI-total	MI-high	MI-low
Item 1	7.597*	1.141	7.766*
Item 2	19.680*	11.518*	8.890*
Item 3	11.319*	6.234*	6.242*
Item 4	6.022*	4.178*	2.996
Item 5	27.990*	15.680*	14.183*
Item 6	7.570*	8.191*	0.173
Item 7	19.306*	0.639	35.825*
Item 8	77.700*	17.105*	74.125*
Item 9	16.591*	19.928*	1.028
Replication 2	MI-total	MI-high	MI-low
Item 1	9.372*	1.545	8.298*
Item 2	4.055*	0.837	3.656
Item 3	6.051*	1.377	7.793*
Item 4	9.646*	0.709	10.018*
Item 5	24.772*	12.227*	11.504*
Item 6	19.397*	17.096*	5.084*
Item 7	8.447*	1.114	18.674*
Item 8	93.609*	18.695*	81.846*
Item 9	53.908*	66.796*	0.906
Replication 3	MI-total	MI-high	MI-low ability gr
Item 1	15.786*	7.224*	9.198*
Item 2	19.933*	6.024*	15.158*
Item 3	15.709*	7.113*	9.702*
Item 4	6.409*	0.293	10.373*
Item 5	30.932*	2.758	35.941*
Item 6	27.779*	26.724*	3.183
Item 7	23.606*	1.223	25.687*
Item 8	57.146*	13.827*	49.918*
Item 9	38.543*	41.919*	.

Table E.3 *(continued)*

Replication 4	MI-total	MI-high	MI-low
Item 1	3.183	0.669	2.672
Item 2	11.332*	7.650*	4.387*
Item 3	14.255*	9.781*	4.757*
Item 4	10.192*	0.720	11.350*
Item 5	18.821*	3.019	19.130*
Item 6	17.192*	16.918*	1.160
Item 7	16.913*	2.764	16.225*
Item 8	83.613*	27.107*	61.978*
Item 9	15.706*	18.220*	.
Replication 5	MI-total	MI-high	MI-low
Item 1	7.018*	2.058	4.807*
Item 2	10.237*	0.656	12.110*
Item 3	4.814*	4.110*	0.632
Item 4	14.874*	0.283	16.555*
Item 5	32.547*	7.938*	24.825*
Item 6	13.206*	12.460*	0.452
Item 7	17.899*	1.880	36.376*
Item 8	46.911*	8.090*	41.083*
Item 9	46.712*	50.818*	1.025
Replication 6	MI-total	MI-high	MI-low
Item 1	14.627*	3.341	11.902*
Item 2	14.439*	6.053*	7.929*
Item 3	24.113*	21.411*	3.486
Item 4	8.187*	1.537	6.001*
Item 5	57.771*	22.965*	34.741*
Item 6	15.994*	13.657*	3.015
Item 7	27.742*	0.142	37.172*
Item 8	58.266*	19.297*	40.950*
Item 9	41.794*	43.968*	1.895
Replication 7	MI-total	MI-high	MI-low
Item 1	7.013*	3.398	3.765
Item 2	10.961*	7.125*	4.494*
Item 3	6.219*	7.788*	0.123
Item 4	26.505*	0.057	34.641*
Item 5	33.798*	22.587*	13.353*
Item 6	11.975*	10.939*	1.387
Item 7	15.019*	0.031	20.284*
Item 8	69.799*	10.392*	68.715*
Item 9	12.855*	13.464*	.

Table E.3 *(continued)*

Replication 8	MI-total	MI-high	MI-low
Item 1	7.097*	0.457	7.807*
Item 2	19.914*	12.518*	8.058*
Item 3	0.078	0.153	0.886
Item 4	20.452*	1.023	23.022*
Item 5	20.107*	8.070*	12.071*
Item 6	26.746*	22.879*	3.595
Item 7	13.478*	0.795	14.946*
Item 8	81.630*	16.125*	76.472*
Item 9	59.100*	62.654*	.
Replication 9	MI-total	MI-high	MI-low
Item 1	6.018*	0.599	6.280*
Item 2	14.865*	7.868*	7.275*
Item 3	7.603*	6.098*	1.748
Item 4	14.341*	0.005	19.656*
Item 5	38.164*	14.624*	24.487*
Item 6	6.641*	6.855*	0.313
Item 7	21.663*	0.730	27.343*
Item 8	76.724*	13.971*	74.119*
Item 9	16.018*	19.781*	.
Replication 10	MI-total	MI-high	MI-low
Item 1	7.667*	1.017	0.013
Item 2	16.942*	10.005*	8.218*
Item 3	5.155*	1.018	9.538*
Item 4	15.984*	1.820	0.162
Item 5	21.581*	12.800*	16.546*
Item 6	28.548*	24.654*	15.102*
Item 7	25.310*	0.154	3.900*
Item 8	46.768*	2.560	50.670*
Item 9	33.576*	37.989*	8.130*
Replication 11	MI-total	MI-high	MI-low
Item 1	3.688	0.000	5.648*
Item 2	15.602*	5.393*	10.524*
Item 3	4.991*	2.537	2.749
Item 4	22.771*	6.054*	17.816*
Item 5	25.425*	9.150*	15.397*
Item 6	16.078*	18.152*	1.546
Item 7	19.957*	1.162	22.950*
Item 8	80.081*	17.803*	67.459*
Item 9	28.339*	35.171*	.

Table E.3 *(continued)*

Replication 12	MI-total	MI-high	MI-low
Item 1	10.843*	4.331*	6.897*
Item 2	21.742*	8.184*	14.662*
Item 3	6.471*	2.657	5.226*
Item 4	11.377*	0.724	11.973*
Item 5	36.687*	17.152*	21.417*
Item 6	20.521*	21.467*	0.781
Item 7	21.369*	4.227*	17.916*
Item 8	64.570*	29.946*	41.568*
Item 9	34.468*	41.904*	.
Replication 13	**MI-total**	**MI-high**	**MI-low**
Item 1	5.501*	0.042	8.548*
Item 2	9.280*	10.603*	1.758
Item 3	7.751*	5.185*	2.844
Item 4	14.037*	1.021	15.173*
Item 5	33.473*	6.852*	31.487*
Item 6	19.204*	10.958*	11.408*
Item 7	35.111*	1.139	61.193*
Item 8	47.872*	9.902*	49.285*
Item 9	57.364*	59.715*	.
Replication 14	**MI-total**	**MI-high**	**MI-low**
Item 1	8.155*	0.009	10.744*
Item 2	19.832*	8.472*	10.839*
Item 3	9.827*	3.691	8.456*
Item 4	18.311*	0.237	19.944*
Item 5	34.671*	19.086*	14.883*
Item 6	34.759*	31.778*	5.720*
Item 7	31.824*	0.279	37.819*
Item 8	43.298*	6.988*	38.777*
Item 9	33.994*	39.121*	.
Replication 15	**MI-total**	**MI-high**	**MI-low**
Item 1	7.631*	2.647	4.681*
Item 2	8.336*	2.364	5.871*
Item 3	10.847*	14.063*	0.093
Item 4	15.997*	0.041	21.508*
Item 5	24.534*	11.783*	11.351*
Item 6	14.952*	11.883*	6.192*
Item 7	20.523*	0.027	27.540*
Item 8	62.808*	9.952*	56.662*
Item 9	31.147*	36.614*	2.011

Table E.3 *(continued)*

Replication 16	MI-total	MI-high	MI-low
Item 1	7.534*	0.563	8.330*
Item 2	5.067*	0.965	5.004*
Item 3	12.042*	5.588*	7.727*
Item 4	11.672*	0.434	19.001*
Item 5	22.424*	3.650	21.932*
Item 6	29.518*	31.177*	0.096
Item 7	23.590*	0.023	32.937*
Item 8	41.392*	5.455*	45.987*
Item 9	28.362*	31.392*	.

Replication 17	MI-total	MI-high	MI-low
Item 1	11.878*	0.752	12.790*
Item 2	21.017*	11.312*	9.445*
Item 3	1.161	0.305	0.989
Item 4	23.225*	3.055	21.037*
Item 5	41.584*	11.885*	29.885*
Item 6	30.036*	31.309*	0.604
Item 7	33.277*	0.168	45.507*
Item 8	48.715*	8.322*	44.465*
Item 9	39.280*	43.761*	0.000

Replication 18	MI-total	MI-high	MI-low
Item 1	3.244	0.256	3.623
Item 2	5.541*	5.608*	1.114
Item 3	17.364*	6.757*	11.397*
Item 4	14.418*	0.976	14.410*
Item 5	38.225*	17.998*	21.814*
Item 6	17.487*	11.242*	8.049*
Item 7	16.988*	0.067	19.841*
Item 8	60.469*	12.884*	57.781*
Item 9	51.602*	53.003*	.

Replication 19	MI-total	MI-high	MI-low
Item 1	8.330*	0.012	13.102*
Item 2	6.942*	6.628*	1.752
Item 3	4.424*	5.124*	0.434
Item 4	22.221*	0.831	25.366*
Item 5	32.948*	11.573*	23.988*
Item 6	21.809*	18.838*	4.699*
Item 7	24.245*	0.172	32.117*
Item 8	30.521*	12.313*	23.175*
Item 9	28.273*	32.054*	.

Table E.3 *(continued)*

Replication 20	MI-total	MI-high	MI-low
Item 1	8.453*	5.632*	3.071
Item 2	0.082	0.817	1.134
Item 3	8.284*	7.856*	1.095
Item 4	20.849*	0.578	20.848*
Item 5	20.861*	10.862*	7.995*
Item 6	20.570*	17.440*	5.727*
Item 7	24.361*	0.223	29.499*
Item 8	51.177*	8.997*	43.426*
Item 9	50.897*	57.699*	.
Replication 21	MI-total	MI-high	MI-low
Item 1	9.488*	5.104*	4.851*
Item 2	15.102*	6.612*	8.955*
Item 3	21.721*	25.095*	1.289
Item 4	13.540*	0.050	17.057*
Item 5	26.595*	7.122*	21.221*
Item 6	27.987*	24.054*	7.635*
Item 7	18.227*	1.539	18.276*
Item 8	66.535*	15.939*	59.115*
Item 9	25.506*	29.229*	3.016
Replication 22	MI-total	MI-high	MI-low
Item 1	15.420*	3.671	11.783*
Item 2	5.490*	1.599	4.089*
Item 3	10.724*	7.237*	4.187*
Item 4	0.960	0.970	0.159
Item 5	37.845*	6.068*	38.995*
Item 6	21.381*	24.033*	0.000
Item 7	20.325*	0.246	29.085*
Item 8	55.525*	12.044*	52.671*
Item 9	19.214*	22.190*	2.062
Replication 23	MI-total	MI-high	MI-low
Item 1	4.305*	0.184	5.568*
Item 2	7.893*	2.767	5.701*
Item 3	6.126*	4.270*	2.073
Item 4	12.923*	0.095	19.093*
Item 5	54.958*	29.327*	26.657*
Item 6	24.141*	23.674*	2.438
Item 7	17.165*	0.487	19.341*
Item 8	42.107*	13.992*	32.068*
Item 9	48.281*	57.073*	0.865

Table E.3 *(continued)*

Replication 24	MI-total	MI-high	MI-low
Item 1	4.505*	0.209	5.338*
Item 2	19.696*	8.643*	11.154*
Item 3	5.982*	4.769*	1.381
Item 4	12.827*	0.337	14.292*
Item 5	39.027*	12.876*	26.406*
Item 6	21.647*	19.363*	3.869*
Item 7	15.214*	0.382	26.944*
Item 8	82.368*	20.969*	65.106*
Item 9	29.213*	35.326*	0.334
Replication 25	MI-total	MI-high	MI-low
Item 1	9.332*	1.360	8.961*
Item 2	20.096*	3.339	20.810*
Item 3	11.333*	3.860*	11.526*
Item 4	6.730*	0.115	7.944*
Item 5	26.691*	12.058*	15.884*
Item 6	37.279*	42.768*	0.030
Item 7	20.503*	0.021	29.607*
Item 8	48.410*	13.276*	40.728*
Item 9	34.925*	42.465*	1.691

Note. The Critical values is 3.841 for chi-square with df = 1.
* $p<.05$

APPENDIX F

The McNemar Test

The McNemar Test

In the McNemar test, the data are usually summarized in a 2x2 contingency table. The data are observations on n independent bivariate random variables. The measurement scale for the X_i, and the Y_i is nominal with two categories. The 2x2 contingency table is shown as follows:

		Classification of the Y_i	
		$Y_i=0$	$Y_i=1$
Classification of X_i	$X_i=0$	a(the number of pairs where $X_i=0$ and $Y_i=0$)	b (the number of pairs where $X_i=0$ and $Y_i=1$)
	$X_i=1$	c(the number of pairs where $X_i=1$ and $Y_i=0$)	d (the number of pairs where $X_i=1$ and $Y_i=1$)

Assumptions

1. The pairs (X_i, Y_i) are mutually independent.

2. The measurement scale is nominal with two categories for all X_i and Y_i.

3. The difference $P(X_i=0, Y_i=1)-P(X_i=1, Y_i=0)$ is negative for all i, or zero for all i, or positive for all i.

Hypotheses

H_0: $P(X_i=0, Y_i=1)=P(X_i=1 Y_i=0)$ for all i

H_1: $P(X_i=0, Y_i=1)^\wedge= P(X_i=1, Y_i=0)$ for all i

Test statistic

The test statistic for the McNemar test is usually written as:

$$T_1 = \frac{(b-c)^2}{b+c}$$

Decision rule

Let n equal $b+c$. If n exceeds 20, reject H_0 at a level of significance alpha if T_1 exceeds the (1-alpha) quantile of a chi-square random variable with 1 degree of freedom. Otherwise accept H_0.

APPENDIX G

Summary of the Data for the McNemar Test

Table G.1

Summary of the Data for the McNemar Test for a DIF

			MC		
			0	1	Total
IRT	0		37	3	40
	1		61	124	185
		Total	98	127	225

			MI		
			0	1	Total
IRT	0		37	3	40
	1		60	125	185
		Total	97	128	225

			MI-Divided		
			0	1	Total
IRT	0		30	10	40
	1		8	177	185
		Total	38	187	225

			MI		
			0	1	Total
MC	0		97	1	98
	1		0	127	127
		Total	97	128	225

			MI-Divided		
			0	1	Total
MC	0		38	60	98
	1		0	127	127
		Total	38	187	225

			MI-Divided		
			0	1	Total
MI	0		38	59	97
	1		0	128	128
		Total	38	187	225

Table G.2

Summary of the Data for the McNemar Test for ab DIF

		MC		
		0	1	Total
IRT	0	52	9	61
	1	45	119	164
	Total	97	128	225

		MI		
		0	1	Total
IRT	0	52	9	61
	1	48	116	164
	Total	100	125	225

		MI-Divided		
		0	1	Total
IRT	0	46	15	61
	1	9	155	164
	Total	55	170	225

		MI		
		0	1	Total
MC	0	91	6	97
	1	9	119	128
	Total	100	125	225

		MI-Divided		
		0	1	Total
MC	0	53	44	97
	1	2	126	128
	Total	55	170	225

		MI-Divided		
		0	1	Total
MI	0	55	45	100
	1	0	125	125
	Total	55	170	225

Table G.3

Summary of the Data for the McNemar Test for b DIF

		MC		
		0	1	Total
IRT	0	4	2	6
	1	0	219	219
	Total	4	221	225

		MI		
		0	1	Total
IRT	0	6	0	6
	1	1	218	219
	Total	7	218	225

		MI-Divided		
		0	1	Total
IRT	0	6	0	6
	1	0	219	219
	Total	6	219	225

		MI		
		0	1	Total
MC	0	4	0	4
	1	3	218	221
	Total	7	218	225

		MI-Divided		
		0	1	Total
MC	0	4	0	4
	1	2	219	221
	Total	6	219	225

		MI-Divided		
		0	1	Total
MI	0	6	1	7
	1	0	218	218
	Total	6	219	225

VITA
Ratchaneewan Wanichtanom
Recruitment and Selection Center
Civil Service Commission
Pitsanulok Rd. Dusit, Bangkok, Thailand 10300
rxw100g@hotmail.com

Education
1982 Thammasat University
        Department of Psychology
        2 Phrachan Road, Phranakorn
        Bangkok 10200 Thailand
        B.S. Psychology
1985 Mahidol University
        Psychiatry Department Siriraj Hospital
        Bangkok Noi, Bangkok, Thailand
        M.S. (Clinical Psychology)
1997 Old Dominion University
        Department of Psychology, MGB.
        Norfolk, Virginia 23529, USA
        M.S. Psychology
2000 Old Dominion University
        Department of Psychology, MGB.
        Norfolk, Virginia 23529, USA
        Ph.D. Industrial/Organizational Psychology

Professional Experience
Personnel Researcher
Sommerville and Company, Inc.
727 East 16th Avenue
Denver, Colorado 80203, USA
October 1998 – April 1999

Clinical Psychologist
Department of Psychiatry
Rajavithi Hospital   (Bangkok Thailand)
June 1988 - July 1994

Script writer
Media Impact Company (Bangkok Thailand)
March 1986 - June 1988

School Psychologist
Bangkok Metropolitan (Bangkok Thailand).
January 1985 - March 1986