August 2015

An Investigation of Subtest Score Equating Methods under Classical Test Theory and Item Response Theory Frameworks

MinJeong Shin
University of Massachusetts - Amherst

An Investigation of Subtest Score Equating Methods under Classical Test Theory and Item Response Theory Frameworks

A Dissertation Presented

by

MINJEONG SHIN

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF EDUCATION

May 2015

Education

Psychometric Methods, Educational Statistics, and Research Methods

An Investigation of Subtest Score Equating Methods under Classical Test Theory and Item Response Theory Frameworks

A Dissertation Presented

By

MINJEONG SHIN

Approved as to style and content by:

____

Ronald K. Hambleton, Chairperson

____

Lisa A. Keller, Member

____

Daeyoung Kim, Member

____

Christine B. McCormick, Dean
College of Education

ACKNOWLEDGEMENTS

Writing this dissertation could not have been possible without the support of many people. First and foremost, I owe my deepest gratitude to Professor Ron Hambleton, my dissertation committee chair. I have been fortunate to have him as my mentor. I also would like to thank the members of my dissertation committee, Professors Lisa Keller and Daeyoung Kim for their insightful and constructive comments despite their busy schedule. In addition to my committee members, I am grateful to all faculty and colleagues with whom I have interacted during the course of my doctoral studies at Umass. I appreciate their encouragement and support.

None of this would have been possible without the love and patience of my family. I would like to thank my parents and sister. Lastly, I offer my regards to all those who supported me in any respect during the completion of this dissertation.

ABSTRACT

AN INVESTIGATION OF SUBTEST SCORE EQUATING METHODS UNDER CLASSICAL TEST THEORY AND ITEM RESPONSE THEORY FRAMEWORKS

MAY 2015

MINJEONG SHIN, B.A., YONSEI UNIVERSITY, KOREA

M.A., YONSEI UNIVERSITY, KOREA

Ed.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Ronald K. Hambleton

Test scores are usually equated only at the total score level. If a test mainly measures a single trait, indicating that the test is essentially unidimensional, equating at the total score level could be the best choice. However, when a test is composed of subtests having negligible relationships among them, separate equating for each subtest offers the best choice. Given a moderate amount of correlations among the subtests, performing individual equating for each subtest may be misleading in that it ignores the relationship of the subtests.

This study applied and compared several possible subtest score equating methods based on classical test theory and item response theory examining some important factors including correlations among dimensions, different proficiency distributions with skewness or mean shifts, and the number of items and common items. Based on the methods from a classical test theory perspective, the results showed that when the correlations among dimensions were high, using either the total or anchor total score as the anchor could produce better equating results than using the anchor score from each subtest. Among the different input scores for equating—observed scores, weighted

averages, and augmented scores—using augmented scores yielded slightly less equating error than the other two methods.

Under the item response theory framework, concurrent calibration and separate calibration as well as unidimensional IRT equating and the unidimensional approximation method using multidimensional IRT parameters were applied. The unidimensional approximation method did not perform well compared to unidimensional IRT methods. The proficiency distribution with relatively high skewness or mean shifts yielded the largest equating errors compared to other distributions.

Further study is recommended: using more complex models, rather than a simple structure model, to simulate item responses, as well as using direct multidimensional IRT equating rather than the two steps of the unidimensional approximation method and unidimensional IRT equating.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

1.1 Background

Interest in subtest score reporting has increased substantially because of the educational reform initiative in the US and desire for diagnostic information. Under the No Child Left Behind Act of 2001 (NCLB), individual students must receive interpretive, descriptive, and diagnostic reports. Recently, states have been under considerable pressure to report subtest scores to help students, teachers, and curriculum specialists make wider use of statewide testing programs. Accordingly, there has been increasing interest in reporting subtest scores to meet these demands (e.g., Sinharay & Haberman, 2011a). Educational or psychological tests are often composed of several subsections based on content or cognitive areas. In a single subject, for instance, subtest scores in a math test can be based on the content strands such as algebra, number sense, and geometry, or based on cognitive skills including knowledge, application, and analysis. Scores from diverse subject areas such as math, language arts, social studies, and science, can provide other examples of subtest scores, also known as domain scores.

Many testing programs including large-scale testing programs, such as SAT, ACT, and LSAT, report subtest scores as well as a total score to provide diagnostic information regarding examinees' strengths and weaknesses in different content areas or cognitive skills. However, it is important to confirm that subtest scores are reliable and comparable across forms, if multiple forms are administered, when reporting subtest scores and making interpretations of those scores (American Psychological Association, American Educational Research Association, & National Council on Measurement in Education, 1999; Puhan & Liang, 2011a; 2011b). In order to report subtest scores and make a meaningful interpretation of those scores, subtest scores should be comparable across different forms of a test. For example, for teachers, they often monitor students' subtest scores over time to check their students' performance, as well as whether their instructions were effective and that they thoroughly covered important components in different content areas. To compare last year's subtest scores with this year's, equating—a statistical method used to adjust form difficulty and allowing scores on the different forms to be used interchangeably—is needed to make subtest scores comparable. Identical scores from last year and this year may not indicate the same level of proficiency. In other words, a higher mean score this year does not necessarily indicate that this year's students outperformed last year's students.

For students, they may want to measure their growth through tests administered twice a year such as in September and in May during the academic year. A student's higher scores on a reading test in May than in September is not always an indication of growth. That is, one cannot justify concluding that her reading skills have improved. The form difficulty should be adjusted before making such interpretations.

Comparability of subtest scores across forms is also important, especially when using subtest scores to make high-stakes decisions such as admission or certification. Candidates who fail the test may want to know their strengths and weaknesses in content areas covered in the test in order to plan for future studies. Moreover, institutions such as colleges may want to create a profile for their students' performance to evaluate their programs and develop remediation programs if necessary (Haladyna & Kramer, 2004; Puhan, Sinharay, Haberman, & Larkin, 2010).

1.1.1 A Brief Overview of Subtest Score Equating based on Classical equating

Under a common item nonequivalent groups design, choosing appropriate common items that represent content and statistical characteristics of the entire test is crucial. Using unrepresentative anchor sets may lead to inaccurate equating results. Especially in the context of subtest score equating, there are usually not sufficient items in each subtest; thus, only a small number of common items can be selected. For instance, if a subtest has 8 items and 25% of them are chosen as common items, then that becomes only 2 items. It is difficult to obtain desirable equating results with such a small number of common items. In this sense, using different anchor sets that utilize more items can be a good option when conducting subtest score equating. For instance, when subtests are highly related to each other and primarily measure one single trait, equated total scores can be used as the anchor score rather than the anchor score from each subtest after equating total scores of the two forms. Utilizing more items, representing the total test as the anchor, can lead to more accurate equating results. However, the downside of this method is that the content match may not be very high if each subtest measures quite a different trait. Thus, there is a trade-off between using more items as the anchor and using items with high content match. The decision to include either more items of which content match is not high or fewer items that represent content well can be made based on the relationship among subtests. If subtests are highly related to each other, which is common in many testing programs, using the total score may perform better than using the anchor score corresponding only to the subtest. On the other hand, if subtests are not so highly related to one another, it is not reasonable to use the total score as the anchor despite providing a sufficient number of items.

In addition to using different anchor sets, using different input data such as adjusted subtest score estimates rather than observed subtest scores may offer more desirable subtest score equating results. Since previous studies (Dwyer et al., 2006; Sinharay & Haberman, 2008; Sinharay, 2010) reported that adjusted scores produced better estimates of subtest scores, it is worthwhile to consider using those scores instead of observed subtest scores when equating subtest scores. Many researchers have studied methods for increasing subtest score reliability and improving subtest score estimation; however, there has been little research conducted on subtest score equating.

Weighted averages suggested by Sinharay and Haberman (2011a) represent one option to estimate subtest scores in the classical test theory framework. Weighted averages are a special case of augmented scores introduced by Wainer et al. (2001). They differ in that weighted averages place the same weight on all other subtest scores that are not of interest as compared to having different weights for them. Sinharay and Haberman (2011a) found that using weighted averages yielded more accurate subtest score estimates and smaller equating error than observed scores. To compute weighted averages, information regarding a total score, subtest scores, their reliability coefficients, correlations between true and estimated scores, etc. is used. In brief, using different subtest scores, including both observed and adjusted scores as input for equating, offers a method to perform subtest score equating.

1.1.2 A Brief Overview of Subtest Score Equating based on Item Response Theory Equating

Unlike equating based on classical test theory, IRT equating involves parameter estimation for examinees and items. Parameters are usually estimated based on the different IRT models that best fit the data. In the context of subtest score equating, both unidimensional and multidimensional IRT models can be considered.

First, within the unidimensional IRT framework, there are two possible scenarios—each subtest is equated separately or subtests are combined into one total test and equated at the total score level. For instance, if a test contains four subtests, four separate equating procedures need to be performed for the former approach whereas only one equating is required for the latter approach. However, these cases do not take into account the relationship among subtests and between the total test and individual subtests. Subtests within the total test are likely to be unidimensional, but the total test may not be close to unidimensional if each subtest measures somewhat different traits. In this situation, equating at the total score level can be problematic because it ignores the distinctiveness of each subtest. On the other hand, when separate equating for each subtest is conducted, insufficient number of common items can affect inaccurate linking constants.

Compared to unidimensional IRT models, multidimensional IRT models estimate parameters more accurately if the test measures more than one factor/trait. When each subtest measures a single trait, there exists more than one dimension at the total test level. If different subtests within the total test measure essentially the same trait and are highly related to one another, applying unidimensional models is more reasonable. Otherwise, multidimensional models provide better solutions in that these models take into account the relationship among dimensions. That is, the multidimensional IRT approach enables simultaneous estimation of item parameters in subtests and correlations among those subtests. However, multidimensional IRT models include more parameters to estimate and are computationally more intensive. Another challenge is that there is no software package for multidimensional IRT equating that provides a raw-to-scale score conversion table.

Brossman (2010) conducted MIRT equating with the unidimensional approximation method as well as multidimensional extensions of IRT observed score equating and found that those two methods performed very similarly. To conduct MIRT equating, item parameters estimated from MIRT models can be approximated to unidimensional IRT item parameters using a method proposed by Zhang (1996) and Zhang and Stout (1999), and those parameters can be used to perform unidimensional IRT equating. In this procedure, both multidimensional and unidimensional IRT linking methods can be applied when a common item nonequivalent design is used. After estimation of item parameters from MIRT models, item parameters from different test forms are placed on the same scale via MIRT linking methods before applying the unidimensional approximation method. Another possible method is to employ the unidimensional approximation method first before applying unidimensional IRT linking methods. Without the approximation method, direct MIRT equating is also possible through the extension of unidimensional IRT equating which requires a vector of ability parameters instead of a single ability parameter.

Depending on the situation such as high or low correlations among dimensions, either unidimensional IRT or multidimensional IRT equating can be applied. Considering both the advantages and disadvantages of unidimensional and multidimensional IRT equating methods, it is worthwhile investigating more parsimonious and accurate solutions for subtest score equating under both unidimensional and multidimensional IRT frameworks.

1.2 Statement of Problem

Often, subtest score reliability is a concern, and rarely are these subtest scores equated. Perhaps in the past, equating was not essential at the subtest level, but now teachers and curriculum specialists often monitor students' subtest scores over time, and examinees make interpretations of their subtest scores to obtain more comprehensive diagnostic information as well as to measure their growth, especially in the case of tests administered with time intervals. Thus, equating from form to form (or year to year) has acquired increasing importance.

Test scores have usually been equated only at the total score level. If a test mainly measures a single trait, indicating that the test is essentially unidimensional, equating at the total score level could be the best choice. Given high or even perfect correlations among the subtests, conducting equating based on the total score is reasonable. However, if each subtest shows distinctiveness but borrows information from one another or shares some information, equating at the total score level may not always afford the best choice.

A possible unintended consequence of equating at the total score level is its ignoring the distinctiveness of each subset and combining them into one test measuring a single trait or a dominant factor. That is, total score equating does not take into account the information gleaned from each subtest or a second factor/dimension. When the traits of each subtest are related to each other but vary somewhat across the test, utilizing all the information available including the relationship of the subtests would be preferable.

On the other hand, when the relationship among the subtests is negligible, separate equating for each subtest offers the best choice. Given a moderate amount of correlations among the subtests, performing individual equating for each subtest may not be appropriate in that it ignores the relationship of the subtests. In addition, one of the practical problems in subtest score equating is the presence of a few common items in one subtest, or at least this is the way current tests are designed. Meeting equating requirements at the subtest score level already poses a challenge. Few publishers have been eager to also meet requirements to properly equate subtest scores. In particular, when a nonequivalent anchor test design is applied, having only a small number of common items in a subtest makes it problematic to equate at the subtest level. This results from the limited number of common items that may not represent the subtest score from one form to the next with respect to content and/or difficulty (Livingston, 2004; Puhan & Liang, 2011a). To date, however, test publishers have yet to equate at the subtest level. If they accomplish this by placing many internal common items in the subtests, failure to release these common items could make subtest score interpretations more challenging. Common items external to the actual subtest scoring do not need to be released, but including many common items in a subtest would greatly increase test length. How to resolve this is a conundrum.

In brief, when correlations among subtests are high and equating at the total score level is performed, it may not be easy to differentiate the distinctiveness of each subtest, but it is possible to use more items as the anchor. On the other hand, when correlations are low and separate equating is conducted, it is easy to find the distinctiveness of each subtest, but there are usually a small number of anchor items available at the subtest level.

From the perspective of IRT, the concept of multidimensionality of the test can be applied. Although many current testing programs, including statewide assessments, regard their tests as unidimensional, in the future, it is highly probable that multidimensionality will be introduced to their tests via well distinguished content areas, cognitive skills, item types, etc. If the level of multidimensionality is insignificant, there is no need to complicate the situation by introducing extra information; unidimensional models can be simply adopted. A significant amount of multidimensionality, however, cannot be ignored; multidimensional models come into play. Few studies exist on the topic of equating subtest scores. Moreover, the impact of correlation among dimensions on subtest score equating have not been thoroughly examined. The consequences of performing separate equating for each subtest, which does not take into account the relationship of subtests, and equating at the total score level, which ignores distinctiveness of each subtest, have not been compared.

1.3 Purposes of the Study

As the demand for subtest score reporting has increased, reporting reliable and accurate subtest scores has become more important. Previous researchers concentrated on subtest score reliability and subtest score estimation methods. When more than two test forms are used due to test security, comparability of test forms also needs to be examined. However, little research has been conducted on the topic of subtest score equating. The purposes of this dissertation are to investigate the applicability of subtest score equating methods in both classical test theory and IRT frameworks and to compare the methods in terms of the accuracy of equating results. Several important factors that need to be considered in the context of subtest score equating are included: correlations among dimensions, different proficiency distributions, and test length with different proportions of common items. Specific research questions will be presented in Chapter 3 and Chapter 4.

1.4 Study Overview

Literature review in Chapter 2 provides the foundations of subtest score equating methods under classical test theory and IRT frameworks. First, methods to examine the added value of subtest score including proportional reduction in mean squared error and multidimensional scaling analysis are described. Classical equating methods are summarized in the following section, and specific considerations to subtest score equating are presented including using different subtest score estimates as input for equating and using different anchor sets. Both unidimensional IRT and multidimensional IRT models as well as linking and equating methods that can be applied to subtest score equating are presented.

Chapters 3 and 4 present simulation studies focusing on the comparison of subtest score equating methods under each framework—classical test theory or IRT. In Chapter 3, all the manipulation factors and methods are compared under the classical test theory framework. Preliminary studies to check whether subtest scores have added value are included in this chapter. Chapter 4 involves the same factors but different methods based on the IRT framework. These two chapters have a brief introduction, purpose of each study, description of methodology, results, and discussion. Chapter 5 includes overall discussion, limitations of this study, and suggestions for future studies.

CHAPTER 2

REVIEW OF LITERATURE

2.1 Overview of Literature Review

This chapter covers a review of the literature pertinent to subtest score equating methods based on classical test theory and item response theory. The chapter is broken down by topic into the following sections:

2.2 Examining the Added Value of Subtest Scores

2.3 Subtest Score Equating based on the Classical Test Theory Framework

2.4 Subtest Score Equating based on the Item Response Theory Framework.

Whether subtest scores have added value over the total score is first examined by applying a few selective methods: proportional reduction in mean squared error (PRMSE) and multidimensional scaling approaches. According to Lyren (2009), Puhan, Sinharay, Haberman, and Larkin (2010), and Sinharay (2010), if a subtest score is reliable and has distinctiveness compared to the other subtest scores, subtest scores provide additional information over the total score. When subtest scores that have added value over the total test score are reported, comparability across forms also needs to be examined if multiple forms are administered. In this case, form difficulty should be adjusted in order to make comparisons across test forms. Thus, subtest score equating can be performed based on either classical test theory or item response theory framework, which will be described in the following sections.

2.2 Examining the Added Value of Subtest Scores

Per-Erik and Vedman (2013) compared several methods, including DIMTEST, DETECT, confirmatory factor analysis, and PRMSE values, to examine the added value of subtest scores. They found that the index of PRMSE performed better than the other methods when assessing dimensionality. Based on PRMSE values, it is possible to determine whether it is worth reporting subtest scores along with the total score. The main idea of this index is that the correlation between the true subtest score and the observed subtest score (which is related to PRMSEs) should be greater than the correlation between the true subtest score and the observed total score (which is related to PRMSEx) in order for subtest scores to have added value (Sinharay, Haberman, & Puhan, 2007). Although there is a rule to make a decision about the added value of subtest scores, which is that PRMSEs should be greater than PRMSEx, it is not possible to investigate the distinctiveness of each subtest when the difference between PRMSEs and PRMSEx is very small. In other words, there is no rule to test the significance of the difference. For instance, if PRMSEs is greater than PRMSEx but the difference between them is negligible such as .01, can subtest scores still be considered to have added value over the total score? In this case, it is questionable whether reporting subtest scores provide useful additional information. Thus, using more than one criterion to support the decision about reporting or not reporting subtest scores is recommended. Among several popular dimensionality assessment techniques, this study applies multidimensional scaling (MDS) analysis which provides a visual inspection of the test configuration. The index of PRMSE is more related to the concept of reliability whereas MDS is closer to the validity perspective. In brief, both PRMSE values and MDS results can be used to determine whether subtest scores create added value and need to be reported, which will be described in the following sections.

2.2.1 Proportional Reduction in Mean Squared Error

Haberman (2008), Haberman, Sinharay, and Puhan (2009), and Sinharay, Haberman, and Puhan (2007) proposed a statistical index called proportional reduction in mean squared error (PRMSE) to examine the added value of subtest scores over total scores. The underlying assumption of this index is that if the true subtest score is more accurately predicted by the observed subtest score than by the observed total score, one can conclude that the subtest score provides additional information than what is provided by the total score. A subtest score may be considered worthwhile reporting only when the PRMSE of the subtest score is larger than that of the total score. That is, the true subtest score estimate predicted from the observed subtest score should be more accurate than that from the observed total score when the subtest score is reported (Puhan, Sinharay, Haberman, & Larkin, 2010). In addition to these two predictors—the true score estimate predicted by the observed subtest score and the observed total score—a combination of the observed subtest score and the observed total score can also be a predictor of the true subtest score, which was called a weighted average of subtest scores (Haberman, 2008; Sinharay & Haberman, 2011a). Furthermore, Sinharay (2010) and Sinharay and Haberman (2008; 2011a) reported some cases in which although observed subtest scores did not have added value for several tests, weighted averages, which produced more accurate estimates of true subtest scores compared to observed subtest scores, did have added value.

A weighted average is a special case of the augmented subtest score proposed by Wainer et al. (2001). The weighted average uses the same weight on all the other subtest scores that are not of interest whereas the augmented subtest score varies the weight.

Sinharay (2010) showed that weighted averages and augmented scores produced very similar true subtest score estimates.

Likewise, estimates of true subtest score s_t can be computed based on the observed subtest score, total score, and weighted average of subtest and total scores. Sinharay, Haberman, and Puhan (2007) suggested that PRMSEs of these estimates be used to examine whether subtest scores provide additional value and offered equations to estimate true subtest scores. Let s and x denote the observed subtest score and the total score of an examinee, respectively.

First, the estimate of true subtest score s_t based on the observed subtest score is computed by

$$\hat{s}_s = \bar{s} + \hat{\rho}^2(s, s_t)(s - \bar{s}), \tag{2.1}$$

where $\bar{s}$ is the mean subtest score and $\hat{\rho}^2(s, s_t)$ is the subtest score reliability estimated by the KR-20 formula (Kuder& Richardson, 1931),

$$\text{KR20}_{\text{sub}} = \left(\frac{i}{i-1}\right)\left(1 - \frac{\sum pq}{\sigma_{sub}^2}\right) \tag{2.2}$$

where i is the number of subtest items, p is the proportion of examinees who answered an item correctly, q is 1-p, and σ_{sub}^2 is the variance of the subtest. The PRMSE of $\hat{s}_s$ can be computed using

$$\text{PRMSE}_s = \frac{\hat{\sigma}^2(s_t) - \hat{\sigma}^2(s_t)[1 - \hat{\rho}^2(s, s_t)]}{\hat{\sigma}^2(s_t)} = \hat{\rho}^2(s, s_t), \tag{2.3}$$

which is equal to the reliability of the subtest score. Second, the estimate of true subtest score s_t based on the observed total score is

$$\hat{s}_x = \bar{s} + \frac{\hat{\sigma}(s_t)}{\hat{\sigma}(x)}\hat{\rho}(s_t, x)(x - \bar{x}) \tag{2.4}$$

in which $\hat{\sigma}(x)$ is the standard deviation of the total score, $\hat{\sigma}(s_t)$ is the standard deviation of the true subtest score which can be estimated by $\hat{\sigma}(s)\sqrt{\hat{\rho}^2(s, s_t)}$, and

$$\hat{\rho}(s_t, x) = \sqrt{\hat{\rho}^2(x_t, s_t)\hat{\rho}^2(x_t, x)} \quad (2.5)$$

where $\hat{\rho}^2(x, x_t)$ is the total score reliability and $\hat{\rho}(x_t, s_t)$ is the correlation between the true total score and the true subtest score estimated via the following equation:

$$\hat{\rho}(x_t, s_t) = \frac{\hat{\rho}(s,x)}{\hat{\rho}(s,s_t)\hat{\rho}(x,x_t)} - \frac{\hat{\sigma}^2(e_s)}{\hat{\sigma}(s_t)\hat{\sigma}(x_t)} \quad (2.6)$$

where $\hat{\sigma}(e_s)$ is the standard error of measurement of the observed subtest score s, $\hat{\sigma}^2(e_s) = \hat{\sigma}^2(s) - \hat{\sigma}^2(s_t)$, and $\hat{\sigma}(x_t)$ is the standard deviation of the true total score, $\hat{\sigma}(x_t) = \hat{\sigma}(x)\sqrt{\hat{\rho}^2(x, x_t)}$. The PRMSE of $\hat{s}_x$ is $\hat{\rho}^2(s_t, x)$. For subtest scores to have added value, PRMSE of $\hat{s}_s$ should be greater than PRMSE of $\hat{s}_x$; that is, $\hat{\rho}^2(s, s_t) > \hat{\rho}^2(s_t, x)$ (for details, see Sinhary, Puhan, & Haberman, 2011). In other words, the correlation between the observed subtest score and the true subtest score should be larger than the correlation between the true subtest score and the observed total score (Sinharay, Haberman, & Puhan, 2007).

Lastly, Haberman (2008) suggested weighted averages of the subtest and total score, which is a special case of the augmented score introduced by Wainer et al. (2001) in that the weighted average uses the same weight on all subtest scores whereas the augmented score can weight them differently. There are various ways to place weights on subtest scores. A simple case of augmented scores is weighted averages with the same weights across all subtest scores. Differing from the two true subtest score estimates described above, weighted averages and augmented scores utilize information not only from a particular subtest score but from other subtest scores. Given moderate or high correlations among the subtests, it is reasonable to borrow information from other subtests. For instance, an examinee's math subtest score can provide some information about his/her science subtest score. Note that the amount of information hinges on how strongly the two subtests are related. Haberman (2008) presented how weighted averages, which are based on both the subtest score and the total score, can be estimated:

$$\hat{s}_{sx} = \bar{s} + a(s - \bar{s}) + b(x - \bar{x}) \tag{2.7}$$

where

$$a = \frac{\hat{\sigma}(s_t)[\hat{\rho}(s_t,s) - \hat{\rho}(s_t,x)\hat{\rho}(s,x)]}{\hat{\sigma}(s)[1 - \hat{\rho}^2(s,x)]} \tag{2.8}$$

and

$$b = \frac{\hat{\sigma}(s_t)[\hat{\rho}(s_t,x) - \hat{\rho}(s_t,s)\hat{\rho}(s,x)]}{\hat{\sigma}(x)[1 - \hat{\rho}^2(s,x)]}. \tag{2.9}$$

The PRMSE of $\hat{s}_{sx}$ can be expressed as

$$\text{PRMSE}_{sx} = \hat{\rho}^2(s, s_t) + \frac{[\hat{\rho}(x,x_t)\hat{\rho}(s_t,x_t) - \hat{\rho}(s_t,s)\hat{\rho}(s,x)]^2}{1 - \hat{\rho}^2(s,x)} = \hat{\rho}^2(s_t, s_{sx}). \tag{2.10}$$

According to Haberman (2008) and Haberman, Sinharay, and Puhan (2009), for the weighted average to confer added value, the PRMSE of the weighted average (PRMSE_{sx}) should exceed both PRMSE_s and PRMSE_x. Using the above equations, weighted averages can be computed to examine whether they provide additional information over the total score. In addition, since the weighted average in an extreme score range—either very low or high—is pooled to mean, it has smaller variance than the observed subtest score.

2.2.2 Multidimensional Scaling

Factor analysis has been one of the popular methods to assess dimensionality. Multidimensional scaling (MDS) is a relatively newer technique than factor analysis and is useful to analyze the structure of data with a visual inspection. Previous research found that MDS provides parsimonious representation of test structure, especially for non-metric MDS which can represent the data structure with fewer dimensions than factor analysis (Davison, 1985; Schlessinger & Guttman, 1969; Shepard, 1972). Davison and Skay (1991) compared factor analysis with MDS and suggested that MDS is more task-oriented whereas factor analysis is more person-oriented. Implied is that if the focus is more on the instrument such as items and tests when assessing dimensionality, MDS could be a better choice.

MDS can be used to find the most parsimonious dimensional solutions that account for the proximity data (Davison & Sireci, 2000). Proximity refers to a numerical value indicating the similarity or dissimilarity of two objects. The first step of the MDS analysis is to select stimuli to be analyzed and to obtain proximity data (similarity or dissimilarity data)—either direct or derived data. For example, direct proximity data can be collected by asking people to report their perception of the dissimilarity or similarity between the stimuli. Derived proximity, such as Pearson correlations for interval data and tetrachoric correlations for dichotomous data, can also be used in MDS.

Likewise, MDS uses proximity (similarity or dissimilarity) data as input and produces a coordinate matrix representing the stimulus structure. MDS models enable each object to be represented by a point in a multidimensional space where dissimilar objects represented by points are far from each other. MDS solutions are obtained by minimizing proximity or distance discrepancies. In MDS, model-data fit indices are used to determine the most appropriate dimensional solution: STRESS and R-squared (RSQ). STRESS describes the mismatch between the distance estimates and the transformed proximities. STRESS can be computed by using the equation:

$$\text{STRESS} = \sqrt{\frac{\sum_{i,j>i}[\tau(\delta_{ij})-d_{ij}]^2}{\sum_{i,j>i} {d_{ij}}^2}} \tag{2.11}$$

where $\tau(\delta_{ij})$ indicates the transformed proximities and d_{ij} is the distance estimates. The value of STRESS is computed from the square root of squared discrepancies for all proximities. The larger STRESS indicates worse fit. In this sense, STRESS can be interpreted as "badness-of-fit." Another widely used fit index in MDS is R-squared (RSQ) —the squared multiple correlation between the distances and the transformed proximities. It indicates the proportion of variance in the transformed proximities accounted for by the distance estimates (Davison & Sireci, 2000). Unlike STRESS, larger values indicate better fit. Although there are rules of thumbs to determine better or worse fit, these fit indices do not provide an absolute criterion for determining the best dimensional solution.

To determine the most parsimonious and useful solution, it is recommended to examine the change in fit as the number of dimensions decreases. If there is very little improvement by adding one additional dimension, the additional dimension may not be necessary in the model. Kruskal and Wish (1978) suggested a scree-plot of STRESS as an eigenvalue plot in factor analysis. In addition to computing fit indices, visual inspection is also a common method to interpret a dimensional solution. The examination can be preceded by first checking the location of the stimuli in two-dimensional space. The stimulus coordinates are also used to locate relatively large positive or negative coordinates which indicate the difference between those two objects in the dimension.

MDS analysis is more beneficial when the real data are available because it is difficult to simulate all the possible factors and dimensions that could occur in a real situation. Moreover, using a real data set makes it possible to examine the actual items or test specifications to help interpret dimensions. When it comes to a simulation study, replicated MDS (RMDS) where more than one matrix exists can be applied to check the consistency of fit statistics over replications. STRESS in RMDS can be computed by

$$\text{STRESS} = \sqrt{\frac{1}{m}\sum_{k=1}^{m}\left[\frac{\sum_{i,j>i}[\tau(\delta_{ij})-d_{ij}]^2}{\sum_{i,j>i} {d_{ij}}^2}\right]} \tag{2.12}$$

where m is the number of matrix (replications). In RMDS, all matrices are the same except for error, which is similar to the concept of doing multiple replications in a simulation study. STRESS and RSQ values are computed for an overall solution as well as for each matrix. Average RSQ is the same with the mean of individual RSQs indicating the average proportion of variance accounted for in all of the transformed data. However, RMDS is more difficult to fit than MDS (classical MDS) because RMDS uses multiple matrices simultaneously whereas MDS uses only one matrix.

Another type of MDS is weighted MDS (WMDS) which accounts for individual differences in the cognitive process of generating responses. Using WMDS, it is possible to scrutinize structural differences across groups in their relative weights on dimensions. Two sets of matrices are obtained from WMDS: an X matrix representing configuration that fits an entire group and a W matrix representing weights on each dimension. In other words, the W matrix represents information that is unique to each individual group while the X matrix represents information that is shared in common by all the groups. Weights that are close to 0 indicate that the dimension is not so important whereas weights close to 1 have large influence on the dimension. Elements in a weight matrix show the relative emphasis that an individual or a group places on a dimension. It is possible that dimensions perform differently for different groups. In this case, WMDS can be applied to discover the structure of the data and investigate differences among groups.

Cluster analysis refers to a data reduction technique to identify meaningful item groupings. Cluster analysis can be used in addition to MDS to discover clusters—a set of objects that are more similar to each other than they are to objects outside the cluster. Stimulus coordinates obtained from MDS can be employed as input for cluster analysis. There are a number of different types of clustering methods (for details, see Aldenderfer & Blashfield (1984) and Milligan (1996)). In general, clustering methods are classified into two categories: hierarchical and partitioning methods. In the hierarchical method, each object is considered as a separate cluster at the first stage. At each step, most similar objects are congregated together, and all objects are under a single cluster in the last stage. On the other hand, the second method, partitioning (K-means) cluster analysis, starts from a specified number of clusters and reassigns cases from one cluster to another as it proceeds.

Although interpreting MDS solutions can be both objective and subjective, MDS analysis enables researchers to parse the data structure visually and help determine a more parsimonious dimensional solution. As more and more testing programs start introducing multidimensionality into their tests, such as with subtests based on content areas, MDS can be a useful technique to analyze test dimensionality.

2.3 Subtest Score Equating based on the Classical Test Theory Framework

Equating has been widely used to adjust form difficulty when multiple test forms are administered to different groups of examinees and need to be compared to each other. After conducting equating, scores of examinees taking different forms can be compared and used interchangeably. In classical equating, three equating methods can be considered: mean equating, linear equating, and equipercentile equating. These methods are briefly reviewed in Section 2.3.1.

In general, observed scores are usually used as input data in classical equating. However, previous studies reported that adjusted subtest scores are more reliable and produce more accurate true subtest score estimates than unadjusted observed scores (Dwyer et al., 2006; Sinharay, 2010; Skorupski & Carvajal, 2010; Stone, Ye, Zhu, & Lane, 2010). Many researchers have proposed different methods to estimate more accurate true subtest scores and improve subtest score reliability. The methods include Kelly's regressed score method (1927), Yen's (1987) objective performance index, Wainer et al.'s (2001) augmented score method, Haberman's (2008) weighted average, and subtest score estimates based on multidimensional IRT models described in Yao and Boughton (2007). These methods were evaluated and compared in Haberman, Sinharay, and Puhan (2009), Puhan, Sinharay, Haberman, and Larkin (2010), Sinharay (2010), and Stephens (2012). However, there is little research on the topic of comparability of those scores. Sinharay and Haberman (2011a; 2011b) used adjusted subtest scores—weighted averages in equating—and concluded that using weighted averages produced smaller error than using observed subtest scores. Applying Haberman's weighted averages and Wainer's augmented scores in subtest equating is described in Section 2.3.2.

In addition to input data used in equating, one of the data collection designs should be selected. Three commonly used designs are a single group design, a random groups design, and a common item nonequivalent groups design. In the single group design, two test forms are administered to one group of examinees. The random groups design requires two randomly equivalent groups of examinees. In this design, each examinee takes only one test form, which reduces testing time. There should be two test forms when either the single group design or the random groups design is applied. If it is not possible to administer more than one test form in one test administration and test security is a concern, the common item nonequivalent groups design can be adopted. As the name implies, in this design, two groups of examinees are not randomly equivalent, and two forms (or more) include a set of common items which plays a key role in adjusting form difficulty. For instance, although the mean score of Group 1 taking Form 1 is higher than that of Group 2 taking Form 2, one cannot conclude that examinees in Group 1 are more proficient than those in Group 2, or that Form 1 is easier than Form 2. Scores of common items should be compared before making such interpretations. In the common item nonequivalent design, choosing appropriate common items is very important. The set of common items should be a "mini version" of a total test (Angoff, 1984). According to Kolen and Brennan (2004), a common item set should represent the total test form in terms of content and statistical characteristics. Larger numbers of common items are preferable; as a rule of thumb, at least 20% of the total test length is needed (Kolen & Brennan, 2004). Three different anchor sets that can be considered in the context of subtest score equating—subtest anchor score corresponding to each subtest, total score, and anchor total score which is the sum of each subtest anchor score—are described in Section 2.3.3.

2.3.1 Classical Equating Methods

In classical equating, scores in one scale can be converted to scores in another scale by setting equal characteristics of the score distributions. Classical equating methods include three types of equating: mean equating, linear equating, and equipercentile equating. Mean equating is the simplest way to adjust mean difference by setting the means of the two forms equal. Linear equating adjusts both mean and standard deviation. That is, scores on one form are converted to have the same mean and/or standard deviation as scores on the other form.

Equipercentile equating defines a nonlinear relationship between two forms. The distribution of scores on a new form is set equal to the distribution of scores on a base form. Thus, converted scores have approximately the same mean, standard deviation, skewness, and kurtosis as scores in a base form. The process of equipercentile equating involves identifying scores on the two forms in the same percentile rank. Unlike mean equating and linear equating, equipercentile equating uses a curve to identify different form difficulty across score points. In other words, it can deal with a situation where one form can be more difficult than the other form at the extreme score range but less difficult at the middle score range.

When the common item nonequivalent groups design is used, two different equipercentile equating methods can be considered—the frequency estimation method or the chained equipercentile method. First, the frequency estimation method uses the distributions for the synthetic population. The distributions for FormX and FormY, which are *f* and *g*, respectively, can be expressed using the concept of the synthetic population,

$$f_s(x) = w_1 f_1(x) + w_2 f_2(x) \tag{2.13}$$

and

$$g_s(y) = w_1 g_1(y) + w_2 g_2(y), \tag{2.14}$$

where the subscript s denotes the synthetic population, the subscript 1 refers to Population 1, and the subscript 2 refers to Population 2. However, $f_2(x)$ and $g_1(y)$ are not available because FormX is administered only to examinees in Population 1, and FormY is administered only to examinees in Population 2. The assumption entailed in the frequency estimation method is that the conditional distributions remain the same in both populations given the common item set; that is,

$$f_1(x|v) = f_2(x|v), \text{and } g_1(y|v) = g_2(y|v). \tag{2.15}$$

The assumptions can be used to obtain $f_2(x)$ and $g_1(y)$ by applying the equations:

$$f_2(x, v) = f_1(x|v)h_2(v) \text{ and } g_2(y, v) = g_1(y|v)h_1(v), \tag{2.16}$$

where $h_1(v)$ and $h_2(v)$ are the distributions of the scores on the common item set in Populations 1 and 2, respectively. Equipercentile equating can then be applied to the synthetic population, $f_s(x)$ and $g_s(y)$.

Second, in the chained equipercentile equating method, scores on FormX are equated to scores on the common items using Population 1, which is referred to as $e_{V1}(x)$. Scores on the common items are then equated to scores on FormY using Population 2; the equipercentile equating function for this is $e_{Y1}(v)$. To convert a FormX score to a FormY score, $e_{Y(chain)} = e_{Y2}[e_{V1}(x)]$. Kolen and Brennan (2004) stated that chained equipercentile equating is less computationally intensive because it does not require the joint distribution of total and common-item scores as needed in the frequency estimation method. Livingston, Dorans, and Wright (1990) suggested that the chained equipercentile method could produce accurate and stable equating results in practice. The chained equipercentile equating method links the two forms together by equating each form to the common items, thus making choosing an anchor set and its length a more important factor than the other methods (Kolen & Brennan, 2004; Ricker & von Davier, 2007)

Previous studies (Braun & Holland, 1982; Harris & Kolen, 1990; Livingston, Dorans, & Wright, 1990; Marco, Petersen, & Stewart, 1983; Holland, von Davier, Sinharay, & Han, 2006; Wang, Lee, Brennan, & Kolen, 2008) compared these two methods and found that they produced quite different equating results. They suggested that the chained equipercentile method could work better especially when the two groups differ.

To reduce irregularities of the score distributions and produce more accurate equating, smoothing methods can be adopted. Holland and Thayer (1987, 2000) provided descriptions of the method that uses a polynomial log-linear model. Log-linear models use the ordered property of test scores to estimate test score distributions:

$$\log[N_X f(x)] = w_0 + w_1 x + w_2 x^2 + \cdots + w_C x^C. \qquad (2.17)$$

The fitted distribution preserves the first C moments that are identical to those of the sample distribution. For instance, with a polynomial degree of 2 (C=2), the mean and standard deviation of the fitted distribution are the same as the mean and standard deviation of the sample distribution. Choosing the appropriate C is crucial when this smoothing method is applied. One possible procedure involves the calculation of likelihood ratio chi-square goodness-of-fit statistics for each C. Likelihood ratio difference chi-squares can then be tested for significance.

2.3.2 <u>Using Three Different Score Estimates as Input for Subtest Score Equating</u>

When considering equating at the subtest level, the simplest way to compute scores for equating is using the observed score in each subtest. However, many researchers (Yen, 1987; Bock, Thissen, & Zimowski, 1997; Pommerich, Nicewander, & Hanson, 1999; Wainer et al., 2001; Tate, 2004; Gessaroli, 2004; Yao & Boughton, 2007; Sinharay, 2010; Skorupski & Carvajal, 2010; Stone, Ye, Zhu, & Lane, 2010) have found that adjusted subtest scores are more reliable than observed subtest scores.

Yen (1987) proposed the Objective Performance Index (OPI) which combined IRT and empirical Bayesian methods. Bock, Thissen, and Zimowski (1997) and Pommerich, Nicewander, and Hanson (1999) used IRT domain score estimation methods based either on maximum likelihood estimation (MLE) or Bayesian estimation. Tate (2004) extended previous studies to multidimensional cases with various correlations among subtests. Gessaroli (2004) and Yao and Boughton (2007) used multidimensional IRT to compute subtest scores. Likewise, many researchers computed adjusted subtest scores based on the IRT framework.

Under the classical test theory framework, Sinharay and Haberman (2008), Haberman (2008), Sinharay (2010), and Sinharay and Haberman (2011a) introduced ways to predict the true subtest score based on observed subtest scores, observed total scores, and weighted averages of these scores. The authors reported that weighted averages provided more accurate diagnostic information compared to observed subtest scores. Equations to compute weighted averages were described in Section 2.2.1.

In addition to observed subtest scores, weighted averages can be used as input for subtest score equating as in Sinharay and Haberman (2011a). In their study, weighted averages utilizing reliability information, correlations, etc. were computed prior to equating, and those scores were used as input in equating. They reported that using weighted averages produced less equating error as compared to using observed subtest scores. In addition, they found that for some tests, weighted averages could have added value even when subtest scores did not.

A weighted average is a special case of an augmented subtest score (Wainer et al., 2001) in that the former one uses the same weight on all other subtest scores that are not of interest whereas the latter one places different weights. Dwyer et al. (2006) and Sinharay and Haberman (2011) found that weighted averages and augmented scores performed very similarly. Augmented scores can be computed based on classical test theory, item response theory, or the combination of elements from both theories. The basic concept of an augmented score derives from an empirical Bayesian method, which is similar to Kelley's (1927) regressed score. According to Wainer et al. (2001), this method of using an augmented score is a multivariate version of Kelley's regressed estimates, which can be expressed as

$$\hat{\tau} = \mathbf{x.} + \mathbf{B}(\mathbf{x} - \mathbf{x.}) \tag{2.18}$$

where **x** is a vector format of subtest scores, **x.** is the mean of x, and **B** is a matrix for estimated reliability. To compute **B**, the observed covariance matrix ($\mathbf{S}^{\text{obs}}$) needs to be obtained,

$$\mathbf{B} = \mathbf{S}^{\text{true}}(\mathbf{S}^{\text{obs}})^{-1} \tag{2.19}$$

where

$$\mathbf{S}^{\text{true}} = \mathbf{S}^{\text{obs}} - \mathbf{D} \tag{2.20}$$

in which

$$\mathbf{D} = \begin{bmatrix} (1-\rho_1)s_{11}^{\text{obs}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & (1-\rho_v)s_{vv}^{\text{obs}} \end{bmatrix}. \tag{2.21}$$

The diagonal elements of $\mathbf{D}$ are error variance for each subtest—observed subtest score variance multiplied by 1 minus the reliability of the subtest (ρ_{v}).

Wainer et al. (2001) presented how to compute the conditional covariance matrix of the estimated true score by applying,

$$\mathbf{S}_{\hat{\tau}|\mathbf{x}_i} = \mathbf{S}^{\text{true}} - \mathbf{S}^{\text{true}}(\mathbf{S}^{\text{obs}})^{-1}\mathbf{S}^{\text{true}}. \tag{2.22}$$

In addition, to examine whether each subtest provides distinctive information, the rank of $\mathbf{S}^{\text{true}}$ can be used. If it has full rank, each subtest offers independent information. If it has a dominant eigenvalue, the test can be considered as unidimensional.

When classical equating methods are applied in subtest score equating, above score estimates including observed subtest scores, weighted averages, and augmented scores can be used as input for equating.

2.3.3 Using Different Anchor Sets in Subtest Score Equating

In the common item nonequivalent groups design, choosing anchor items that represent the total test is one of the most important procedures in equating. Common item sets should adequately reflect test specifications as well as form difficulty. That is, anchor sets should be content and statistically representative. In general, including larger number of common items results in less random equating error (Budescu, 1985; Wingersky, Cook, & Eignor, 1987; Kolen & Brennan, 2004; Ricker & von Davier, 2007). Because most educational tests deal with heterogeneous content, it is desirable to have larger number of common items in practice (Kolen & Brennan, 2004). As a rule of thumb, at least 20% of the total test items are required for common items (Kolen & Brennan, 2004). Having very few common items or common item sets with inadequate content representation could lead to equating problems (Petersen, Cook, & Stocking, 1983; Kolen & Brennan, 2004). To obtain accurate equating results, a sufficient number of common items that are both content and statistically representative should be selected.

When equating is considered at the subtest level, one possible problem is that there are usually not many items in each subtest. Because of having insufficient number of items in subtests, only a few items can be considered as an anchor set, which may result in large equating error. To reduce error by utilizing the anchor score based on more items, Puhan and Liang (2011a; 2011b) suggested using the total score as the anchor. They employed two approaches for equating subtest scores: using the anchor score in each subtest and using the total score as the anchor. They found that when the number of common items was very small at the subtest level and/or when the total score and the subtest scores were moderately or highly correlated, using the total score as the anchor produced less equating error.

Based on their suggestions, three different anchor scores can be applied in subtest score equating. First, the anchor scores only in the corresponding subtest are used to equate subtest scores. Second, the total score on the new form is equated to the scale of the old form. By using the total scaled score on the new form as the anchor, subtest scores on the new form are equated to the equivalent subtest scores on the old form. Third, anchor total scores are used to conduct subtest score equating. For instance, if there is a total of 80 items (20 items from each of the 4 subtests) with 6 items common to each subtest, 24 items would be used to compute anchor total scores (6 items x 4 subtests = 24 items for linking). Rather than using anchor scores for each subtest, scores of all common items (across subtests) are used in equating.

The second method of using equated total scores as the anchor could be applied in almost every current testing situation. The third method of using anchor total scores as the anchor could also be used if the number of subtests and their common items are sufficient. The first method of using subtest anchor scores, however, would require some test design work (the use of at least reasonable numbers of linking items in a subtest). Although having a sufficient amount of common items in the subtest may not always be feasible in practice, applying the first method may offer the best solution if effective equating is the goal. The second or the third method can be especially useful when the number of common items is very small and subtests share common features in terms of content. The problem with using these methods, however, is that doing so violates the rule that linking items should mirror the two subtests being linked. Even though the number of items used as the anchor is larger when the second method is used, the content match may not be as high. For instance, if each subtest is intended to measure somewhat different content areas which are not highly correlated, the total score or the anchor total score of the two tests being linked may not be as accurate as the anchor items in the two subtests. Since common items in a subtest play a key role in linking two subtests, a test should contain sufficient common items with which content match is well achieved. Thus, it is important to consider different conditions including the number of common items and correlations among subtests when applying the methods described above.

In addition, when subtests are highly related with the total test and when there are very few common items in each subtest such as 1 or 2, using the equated total score could work better than using the anchor total score because the former method utilizes more items. The method using the equated total score as the anchor, however, requires two steps of equating: at the total test level and at the subtest level. When the anchor total score is used, only one equating procedure needs to be performed. Depending on the situation, either the second or the third method could be chosen instead of the first method.

2.4 Subtest Score Equating based on the Item Response Theory Framework

Classical equating methods have been widely used; however, those methods are limited because they cannot be used with testing programs that are built based on IRT. In practice, item analysis and test equating in many testing programs and statewide assessments are performed under the IRT framework. Unlike classical test theory, IRT has invariance property—item and ability parameter values remain unchanged regardless of the proficiency distribution or characteristics of items including item difficulty or discrimination (Hambleton, Swaminathan, & Rogers, 1991; Embretson & Reise, 2000).

The validity of IRT applications depends greatly on meeting the assumptions underlying the models. One such assumption is unidimensionality—a single latent construct determines performance on a test. If a test is essentially unidimensional and measures one single trait even though this is not perfectly realistic, reporting subtest scores along with the total score provides little diagnostic information and could lead to inappropriate interpretation (Stone, Ye, Zhu, & Lane, 2010). In other words, distinctiveness of the subtest scores as well as adequate evidence for reliability and validity should be examined; test dimensionality needs to be assessed. As mentioned in the previous section, Haberman (2008) found that subtest scores have added value when they have relatively high reliability and the subtest scores are moderately correlated among themselves and with the total score. If the correlations are nearly 1, different subtests do not provide independent information. In other words, it is not desirable to report subtest scores when a test is essentially unidimensional.

With unidimensional IRT (UIRT) models, test scores are based on a single dimension usually covering a broad domain but not specific content or cognitive domains. For instance, if the underlying trait intended to measure in a math test is general ability with overall math skills, we cannot interpret a test score for this test as scores from separate subdomains such as number sense, geometry, algebra, etc. When specific content domains are considered as separate dimensions, multidimensional IRT models can be better suited for estimating subtest scores and providing diagnostic information (Yao & Boughton, 2007; Yao, 2011). Likewise, subtest score equating can be considered based on either unidimensional IRT or multidimensional IRT depending on the structure of the data.

2.4.1 Subtest Score Equating using Unidimensional IRT

2.4.1.1 Unidimensional IRT Models

Given one dominant dimension that accounts for a relatively large portion of the covariance among the items, unidimensional IRT models can be applied. Three popular unidimensional IRT models used for dichotomous item response data include one-, two-, and three-parameter logistic models depending on the number of item parameters used in the model. The expression for the three-parameter logistic model is

$$P_i(\theta) = c_i + (1 - c_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} \tag{2.23}$$

where $P_i(\theta)$ is the probability that an examinee with ability θ answers item *i* correctly; a_i is the item discrimination parameter—proportional to the slope of the item characteristic curve (ICC) at the point b_i; b_i is the item difficulty parameter for item *i*—the point on the ability scale where the probability of answering item *i* correctly is $(1 + c_i)/2$; c_i is the pseudo-guessing parameter—the lower asymptote for the ICC and the probability of examinees with low proficiency answering item *i* correctly; and *D* is a scaling factor making the logistic function close to the normal ogive function.

Figure 2.1 presents an example of an item characteristic curve (ICC) describing the relationship between the performance on an item and the proficiency level (θ).

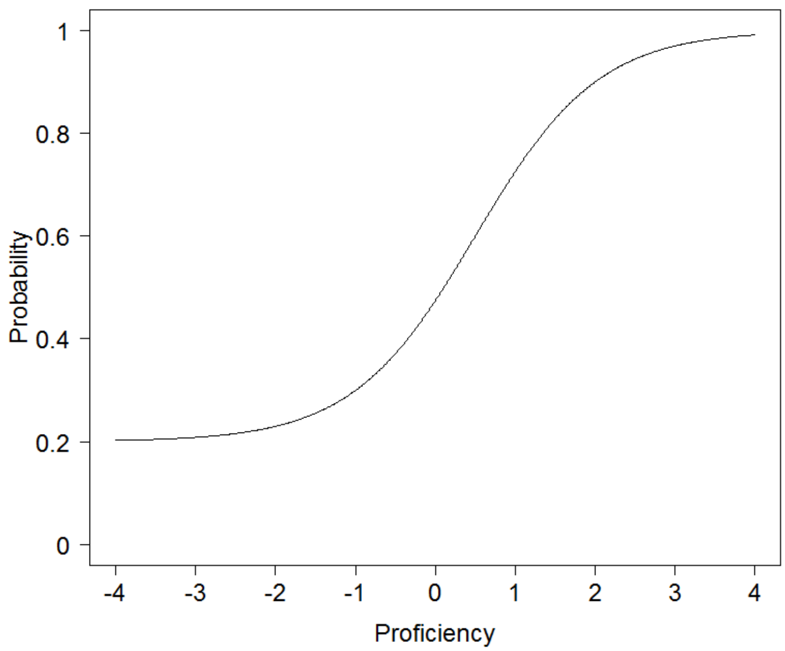


Figure 2.1. Example Item Characteristic Curve (ICC)

a=1.3, *b*=0.5, and *c*=0.2

The test characteristic curve (TCC) —the sum of the item characteristic curves— is another important concept in IRT. The TCC can be used to transform the θ scale to the true score scale and predict true score performance in a test at a given proficiency level. The true score (τ) given θ can be expressed as

$$\tau|\theta = \sum_i^n P_i(\theta) \quad (2.24)$$

where *n* is the number of items. In the context of equating, TCCs obtained from different test forms can illustrate how true scores of examinees can vary from one test form to another given the same proficiency level (θ). By comparing those TCCs, it is feasible to find a conversion table. Test scores of examinees taking different test forms are comparable based on the conversion table. Figure 2.2 shows an example of a TCC. The transformation of θ to the true score or the domain score that ranges from 0 to n or 0% to 100% makes it easier to interpret scores.

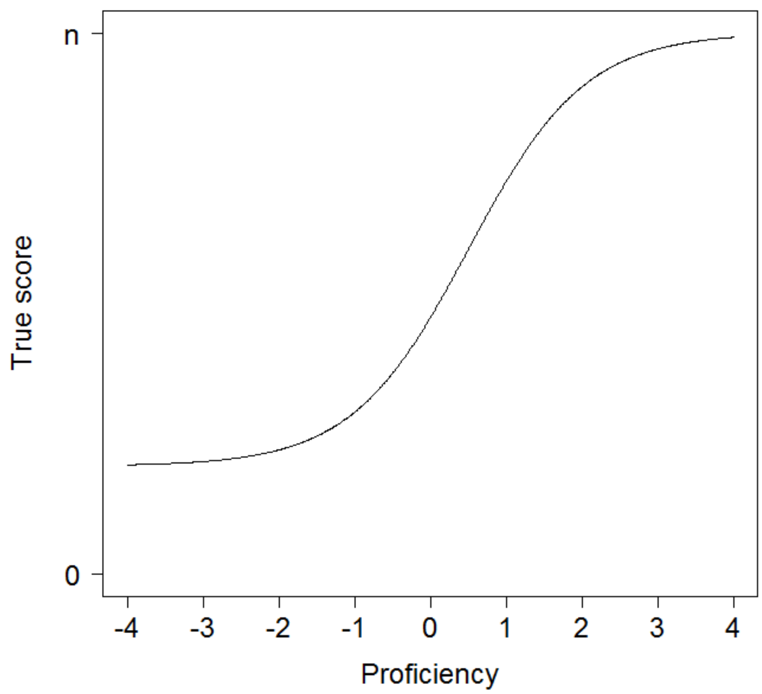


Figure 2.2. Example Test Characteristic Curve (TCC)

2.4.1.2 Unidimensional IRT Linking Methods

Prior to equating, the parameters estimated from different test forms should be placed on the same IRT scale due to the scale indeterminacy property of IRT. Parameter calibration software packages often use a scale with a mean of 0 and a standard deviation of 1 regardless of different proficiency levels depending on the group. Thus, it is necessary to consider a transformation of IRT scales. If an IRT model fits the data, linear transformation equations can be applied to convert IRT parameters estimated from separate calibrations under the common item nonequivalent groups design to the same scale using the item parameters of common items. That is, if parameters for one group are calibrated first and the parameter estimation for another group was separately run from the former calibration, these parameter estimates need to be placed on the same scale via a linear transformation. Item parameters of the common items can be used to perform the scale transformation.

The scales of two test forms, Scale I and Scale J respectively, could differ in mean and standard deviation; they are linearly related as follows:

$$\theta_{Jj} = \mathrm{A}\theta_{Ij} + \mathrm{B} \tag{2.25}$$

$$a_{Ji} = \frac{a_{Ii}}{\mathrm{A}} \tag{2.26}$$

$$b_{Ji} = \mathrm{A}b_{Ii} + \mathrm{B} \tag{2.27}$$

and

$$c_{Ji} = c_{Ii}, \tag{2.28}$$

where A and B are linking constants, θ_{Jj} and θ_{Ij} are θ values for examinee *j* on Scale J and Scale I, a_{Ji}, b_{Ji}, and c_{Ji} are item parameters for item *i* on Scale J, and a_{Ii}, b_{Ii}, and c_{Ii}

are item parameters for item i on Scale I. For any two examinees, j and j^*, or any two items i and i^*, the linking constants A and B are

$$\mathrm{A} = \frac{\theta_{Jj} - \theta_{Jj^*}}{\theta_{Ij} - \theta_{Ij^*}} = \frac{b_{Ji} - b_{Ji^*}}{b_{Ii} - b_{Ii^*}} = \frac{a_{Ii}}{a_{Ji}} \tag{2.29}$$

and

$$\mathrm{B} = b_{Ji} - \mathrm{A}b_{Ii} = \theta_{Jj} - \mathrm{A}\theta_{Ij}. \tag{2.30}$$

Several methods are available to determine the linking constants A and B: the mean/mean method, the mean/sigma method, and the characteristic curve methods including the Haebara method (Haebara, 1980) and the Stocking and Lord method (Stocking & Lord, 1983). First, the mean/mean method described by Loyd and Hoover (1980) uses the mean of *a* parameter estimates to find the A constant and the mean of *b* parameter estimates to find the B constant, which can be expressed as

$$\mathrm{A} = \frac{\mu(a_I)}{\mu(a_J)}, \tag{2.31}$$

and

$$\mathrm{B} = \mu\big(b_J\big) - \mathrm{A}\mu(b_I). \tag{2.32}$$

Second, the mean/sigma method described by Marco (1977) uses the mean and standard deviations of the parameter estimates to obtain the A and B constants. Instead of using the mean of *a* parameter estimates, the standard deviation of *b* parameter estimates are used to find the A constant as follows:

$$\mathrm{A} = \frac{\sigma(b_J)}{\sigma(b_I)}. \tag{2.33}$$

An advantage of using the mean/sigma method over the mean/mean method is that *b* parameters produce more stable estimates than *a* parameters (Kolen & Brennan, 2004). According to Baker and Al-Karni (1991), however, the mean/mean method might be more advantageous than the mean/sigma method in that means tend to be more stable than standard deviations.

Unlike the mean/mean and mean/sigma methods, the characteristic curve methods consider all item parameter estimates simultaneously when estimating linking constants (Kolen & Brennan, 2004). Due to the indeterminacy of IRT, the probability of answering an item correctly given a particular proficiency (θ) remains unchanged regardless of the scale, which can be expressed as

$$p_{ij}(\theta_{Jj}; a_{Ji}, b_{Ji}, c_{Ji}) = p_{ij}\left(A\theta_{Ij} + \mathrm{B}; \frac{a_{Ii}}{\mathrm{A}}, Ab_{Ii} + B, c_{Ii}\right). \quad (2.34)$$

The Haebara method uses the difference between item characteristic curves of Scale I and Scale J; the difference is squared and summed over the common items (V) as follows:

$$\mathrm{Hdiff}(\theta_j) = \sum_{i=1}^{\mathrm{V}}\left[p_{ij}(\theta_{Jj}; a_{Ji}, b_{Ji}, c_{Ji}) - p_{ij}\left(A\theta_{Ij} + \mathrm{B}; \frac{a_{Ii}}{\mathrm{A}}, Ab_{Ii} + B, c_{Ii}\right)\right]^2. \quad (2.35)$$

The Stocking and Lord method uses the squared difference between the test characteristic curves:

$$\mathrm{SLdiff}(\theta_j) = \left[\sum_{i=1}^{\mathrm{V}} p_{ij}(\theta_{Jj}; a_{Ji}, b_{Ji}, c_{Ji}) - \sum_{i=1}^{\mathrm{V}} p_{ij}\left(A\theta_{Ij} + \mathrm{B}; \frac{a_{Ii}}{\mathrm{A}}, Ab_{Ii} + B, c_{Ii}\right)\right]^2. \quad (2.36)$$

The linking constants A and B can be found to minimize the above functions in the Haebara method and the Stocking and Lord method.

Previous studies (Baker & Al-Karni, 1991; Huang, et al., 1991; Way & Tang, 1991; Kim & Cohen, 1992; Kaskowitz & De Ayala, 2001; Hanson & Béguin, 2002; Ogasawara, 2002; Kim & Lee, 2006) found that the characteristic curve methods produced more accurate linking results than the mean/mean and mean/sigma methods. Kim and Lee (2006) compared four IRT linking methods described above and found that the Haebara method usually yielded the least linking error among the four methods.

However, Way and Tang (1991) reported that the Stocking and Lord and the Haebara methods performed similarly.

The scale transformation methods described above are required when separate calibrations are used for the two test forms under the common item nonequivalent groups design. However, if parameters from two test forms are calibrated at the same time, which is referred to as "concurrent calibration" (Wingersky & Lord, 1984), scale transformation methods do not need to be considered because the parameter estimates are already on the same scale. In this case, items which are not administered to a group of examinees are treated as "not reached" items (Kolen & Brennan, 2004).

Kim and Cohen (1998) compared scale linking using the Stocking and Lord method to concurrent calibration and found that concurrent calibration yielded more accurate results than scale linking when the number of common items is small. With a larger number of common items, however, scale linking and concurrent calibration produced similar results. Hanson and Béguin (2002) also found that concurrent calibration yielded more accurate results than the test characteristic curve methods. In Béguin, Hanson, and Glas (2000) as well as Béguin and Hanson (2001), however, the Stocking and Lord method outperformed the concurrent calibration procedure when multidimensionality was introduced with highly correlated abilities under the nonequivalent groups design.

2.4.1.3 Unidimensional IRT Equating

After estimating ability parameters from separate calibration and placing those parameters on the same scale or estimating them from concurrent calibration, examinees' scores on the θ scale are obtained. If these θ scores are reported, scale linking or concurrent calibration could be sufficient. However, if examinees' scores using other than the θ scale need to be reported, it is necessary to conduct equating to make observed scores or true scores comparable across parallel test forms in a more easily interpretable score scale. There are two approaches under IRT equating: IRT observed score equating and IRT true score equating.

First, IRT observed score equating estimates observed score distributions on both test forms, and equipercentile equating is then applied to find the equating relationship. Conditional observed score distributions given θ_j are determined using a recursion formula (Lord & Wingersky, 1984). For instance, in the case of a three-item test, the probability that examinees at the ability level θ_j answer all three items incorrectly and obtain a score of 0 can be expressed as $f(x = 0|\theta_j) = (1 - p_{1j})(1 - p_{2j})(1 - p_{3j})$ where p_{ij} is defined from the IRT model such as a one-, two-, or three-parameter model. The same logic is applied to the probability of answering one item correctly and two items incorrectly and earning a score of 1, which is $f(x = 1|\theta_j) = p_{1j}(1 - p_{2j})(1 - p_{3j}) + (1 - p_{1j})p_{2j}(1 - p_{3j}) + (1 - p_{1j})(1 - p_{2j})p_{3j}$. The recursion formula is as follows:

$$\begin{aligned} f_r(x|\theta_j) &= f_{r-1}(x|\theta_j)(1 - p_{rj}) && x = 0 \\ &= f_{r-1}(x|\theta_j)(1 - p_{rj}) + f_{r-1}(x - 1|\theta_j)p_{rj} && 0 < x < r \\ &= f_{r-1}(x - 1|\theta_j)p_{rj} && x = r \end{aligned} \tag{2.37}$$

where $f_r(x|\theta_j)$ is the conditional distribution of observed scores over the first *r* items for examinees of ability θ_j. After obtaining the conditional observed score distributions at each θ level, these distributions are accumulated either by summing or integrating over all θ levels. When the ability distribution is continuous,

$$f(x) = \int_{\theta} f(x|\theta)\psi(\theta)d\theta, \tag{2.38}$$

and when the ability distribution is discrete,

$$f(x) = \sum_j f\left(x\middle|\theta_j\right)\psi\left(\theta_j\right) \tag{2.39}$$

where $\psi(\theta)$ is the distribution of θ.

To conduct IRT observed score equating, observed score distributions are found for both FormX and FormY, and traditional equipercentile equating is then applied as described in Section 2.3.1.

Second, IRT true score equating seeks the relationship between true scores on FormX (τ_{X}) and true scores on FormY (τ_{Y}), which can be expressed as

$$\mathrm{irt}_{\mathrm{Y}}(\tau_{\mathrm{X}}) = \tau_{\mathrm{Y}}(\tau_{\mathrm{X}}^{-1}) \tag{2.40}$$

where τ_{X}^{-1} is the θ_j corresponding to τ_{X}. This equation involves three steps: 1) specify τ_{X}, 2) find θ_j corresponding to τ_{X}, and 3) find τ_{Y} corresponding to θ_j. For the second step, the iterative procedure using the Newton-Raphson method can be applied to find θ_j by minimizing the expression $func\left(\theta_j\right) = \tau_{\mathrm{X}} - \sum_{i:\mathrm{X}} p_{ij}(\theta_j; a_i, b_i, c_i)$. It is appropriate to use true scores to find an equating relationship for IRT true score equating; however, in practice, true scores are never known. IRT true score equating can be viewed as the comparison of TCC on FormX with TCC on FormY. Figure 2.3 provides an example of a graphical representation of true score equating.

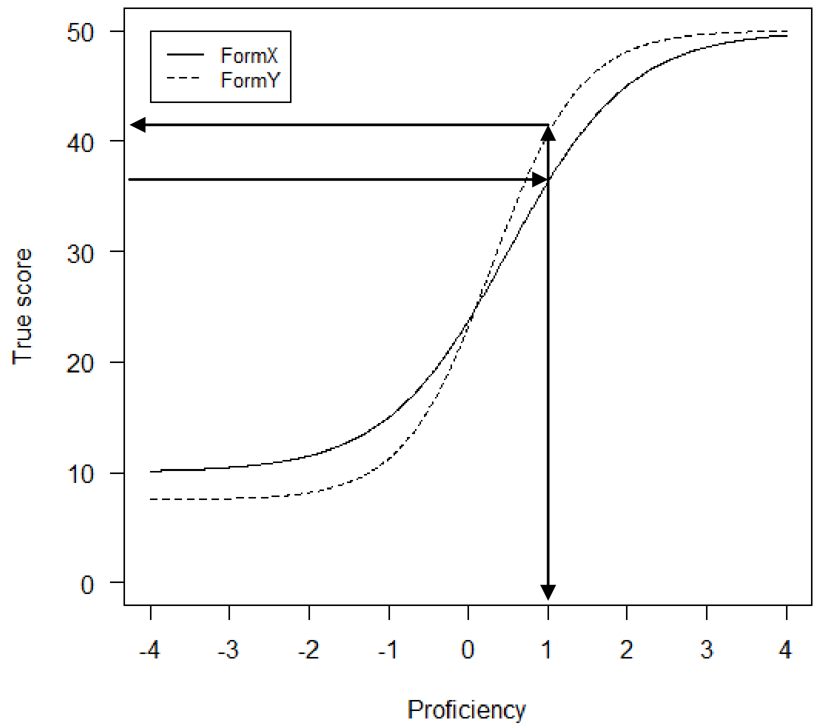


Figure 2.3. True Score Equating using TCCs

2.4.2 Subtest Score Equating using Multidimensional IRT

2.4.2.1 Multidimensional IRT Models

Multidimensional IRT has become popular as many testing programs utilize items measuring more than one latent trait such as multiple cognitive skills or content areas. By applying MIRT models, ability parameters on several dimensions can be estimated simultaneously. In IRT, the probability of answering an item correctly is computed from item parameters and a single ability parameter θ. In MIRT, however, the probability depends on the vector of abilities.

There are two types of MIRT models: compensatory models and noncompensatory models. In compensatory models, a low ability level on one dimension can be overcome by a high ability level on other dimensions. In noncompensatory models, however, a high ability level on one dimension cannot compensate for a low ability level on other dimensions. The compensatory extension of the three-parameter logistic model is as follows (Reckase, 2009):

$$P_i(\boldsymbol{\theta}_j) = c_i + (1 - c_i)\frac{1}{1+e^{-1.7(\mathbf{a}'_i\boldsymbol{\theta}_j + d_i)}}, \tag{2.41}$$

where $\mathbf{a}_i$ is a 1 x m vector of item discrimination parameters for item i, d_i denotes a parameter related to the item difficulty, c_i is a guessing parameter, $\boldsymbol{\theta}_j$ is a 1 x m vector of ability parameters, and m indicates the number of dimensions. It is of note that d_i is not identical to the item difficulty parameter (b_i) in IRT. The MIRT equivalent of the item difficulty parameter is obtained from

$$b = \frac{-d}{\sqrt{\mathbf{a}\mathbf{a}^T}}. \tag{2.42}$$

The compensatory model is based on the linear combination of θ coordinates ($a_i\theta'_j + d_i = a_{i1}\theta_{j1} + a_{i2}\theta_{j2} + \cdots + a_{im}\theta_{jm} + d_i$) whereas the noncompensatory model has nonlinear features with the product of the probabilities of the correct performance on each component of the test item. The noncompensatory model or partially compensatory model is as follows:

$$P(U_{ij} = 1|\theta_j, a_i, \mathbf{b}_i, c_i) = c_i + (1 - c_i)\left(\prod_{l=1}^{m}\frac{1}{1+e^{-1.7a_{il}(\theta_{jl} - b_{il})}}\right) \tag{2.43}$$

where l indicates the dimension of interest.

In addition to these two categories of MIRT models, three types of dimensional structure models can be considered: simple structure, approximate simple structure, and complex structure. First, for a simple structure model—the most basic type of a multidimensional structure—each item is loaded on only one dimension. In other words, all items are unidimensional within each dimension, but the test itself measures multiple traits. Second, with an approximate simple structure, each item is primarily loaded on one dimension, but every item is loaded on multiple dimensions having nonzero discrimination parameters. Third, items in a complex structure contribute to more than one dimension; there is no one primary dimension for each item.

2.4.2.2 Multidimensional IRT Linking Methods

When parameters are estimated using MIRT models from separate runs, those parameters need to be on the same scale in order to make comparisons regardless of the test forms. In this case, MIRT linking procedures are required instead of unidimensional linking methods. Previous studies proposed several MIRT linking methods (Hirsch, 1989; Davey, Oshima, & Lee, 1996; Li & Lissitz, 2000; Min, 2003; Yao & Boughton, 2009).

Ohima, Davey, and Lee (2000) evaluated four MIRT linking procedures: the direct method, the equated function method, the test characteristic function (TCF) method, and the item characteristic function (ICF) method. They found that the TCF and ICF methods performed better than the other two methods. They, however, noted that the linking methods should be chosen based on according to the purpose of linking. The authors suggested that if the focus of linking is on making examinees' true scores equivalent regardless of the test forms, the TCF method, which minimizes the true score differences, may work better than the other methods.

Simon (2008) compared concurrent and separate MIRT linking methods under several conditions including sample size, test length, group equivalence, and correlations between two dimensions. Five MIRT linking methods were compared: concurrent calibration, the TCF, the ICF, the direct method, and Min's method (Min, 2003). The author concluded that concurrent calibration generally outperformed the other methods because it benefited from using a larger sample size. Among the separate linking methods, the ICF method produced less root mean square error (RMSE) and bias under the nonequivalent groups design.

Wei (2008) conducted a simulation study using the common-item nonequivalent groups design with the compensatory multidimensional 2PL model and compared four different MIRT scale linking methods: the direct method, the equated function method, the TCF method, and the ICF method. The author reported that the direct method performed the best among the four linking methods and that the TCF and ICF methods were second best.

Yao (2011) compared the TCF method with extensions of unidimensional linking methods including mean/sigma and mean/mean methods. She conducted a simulation study using tests of five dimensions and confirmed the findings of previous studies that the TCF method outperformed the mean/sigma and mean/mean methods (Kolen & Brennan, 1995; Kim & Lee, 2006).

The probability of answering item i correctly for an examinee j in the multidimensional three-parameter logistic model was presented in (2.41). As the scale is not uniquely determined in both unidimensional and multidimensional IRT models, ability parameters (or item parameters) are determined up to a linear transformation. For MIRT models, equations for the linear relationship are expressed as:

$$\mathbf{a}_i^* = (\boldsymbol{A}^{-1})'\mathbf{a}_i \tag{2.44}$$

$$d_i^* = d_i - \mathbf{a}_i'\boldsymbol{A}^{-1} + \boldsymbol{\beta} \tag{2.45}$$

$$\boldsymbol{\theta}_j^* = \boldsymbol{A\theta}_j + \boldsymbol{\beta}, \tag{2.46}$$

where $\boldsymbol{A}$ is an m x m rotation matrix and $\boldsymbol{\beta}$ is a 1 x m location vector. By transforming item parameter estimates from one scale to another using the above equations, item parameter estimates from separate calibrations are placed on the same scale. Linking methods introduced in previous studies are mostly based on the minimization function with respect to $\boldsymbol{A}$ and $\boldsymbol{\beta}$.

Aforementioned previous studies reported that one of three methods—the direct method, the TCF method, and the ICF method—performed the best depending on the conditions such as equating designs. First, the direct method is a multivariate extension of the minimum χ^2 method (Divgi, 1985). The constants $\boldsymbol{A}$ and $\boldsymbol{\beta}$ are estimated by minimizing the sum of the squared differences between the items from two different scales. The function to find **A** and $\boldsymbol{\beta}$ for this method is

$$f_1(\mathbf{A},\boldsymbol{\beta}) = \frac{1}{n(m+1)}\left\{\sum_{i=1}^{n}\sum_{k=1}^{m}[\hat{a}_{Y_{ik}} - \hat{a}^*_{X_{ik}}]^2 + \sum_{i=1}^{n}[\hat{d}_{Y_i} - \hat{d}^*_{X_i}]^2\right\} \quad (2.47)$$

where n is the number of common items, m is the number of dimensions, i denotes an item, k indicates a dimension, $\hat{a}_{X_{ik}}$ is the estimate of discrimination parameter of item i on dimension k on Scale Y, $\hat{d}_{X_i}$ is the parameter estimate related to item difficulty, and parameters with * indicate transformed parameters to an old form scale, here FormY. Second, the characteristic curve method was proposed by Stocking and Lord's (1983) — the TCF method can be extended to a multidimensional case by

$$f_2(\mathbf{A},\boldsymbol{\beta}) = \frac{1}{q^D}\sum_{\boldsymbol{\theta}} W_{\boldsymbol{\theta}}\left(\sum_{i=1}^{n} P_{Xi}(\boldsymbol{\theta}) - \sum_{i=1}^{n} P^*_{Yi}(\boldsymbol{\theta})\right)^2 \quad (2.48)$$

where q indicates quadrature points, D is the dimension, q^D represents possible choices of $\boldsymbol{\theta}$, and $W_{\boldsymbol{\theta}}$ is the weight taken at different $\boldsymbol{\theta}$ values. Third, the multidimensional extension of the characteristic curve approach proposed by Haebara (1980) —the item characteristic curve method—can be expressed as

$$f_3(\mathbf{A}, \boldsymbol{\beta}) = \frac{1}{nq} \sum_{\boldsymbol{\theta}} W_{\boldsymbol{\theta}} \sum_{i=1}^{n} (P_{Xi}(\boldsymbol{\theta}) - P_{Yi}^{*}(\boldsymbol{\theta}))^2. \tag{2.49}$$

2.4.2.3 Unidimensional Approximation

Once parameters from the two different scales are placed on the same scale by using the above linking methods, equating is then performed. As can be seen in the previous section, there have been several studies on MIRT scale linking procedures (Hirsch, 1988, 1989; Davey, Oshima, & Lee, 1996; Li & Lissitz, 2000; Min, 2003; Yao & Boughton, 2009); however, there has been little research on MIRT equating. Brossman (2010) proposed three equating procedures: full MIRT observed score equating, unidimensional approximation of MIRT true score equating, and unidimensional approximation of MIRT observed score equating. According to Zhang (1996), Zhang and Stout (1999), and Carroll, Williams, and Levine (2007), multidimensional models can be approximated by unidimensional submodels. Brossman (2010) found that the full MIRT procedure and two unidimensional approximation of MIRT equating performed very similarly.

Unlike unidimensional IRT equating, there is no unique solution in finding a true score equivalent of a new form in an old form in MIRT equating because a particular true score can be derived from a number of possible combinations of ability parameters (Brossman, 2010). By applying the unidimensional approximation method, unidimensional IRT equating methods described in Section 2.4.1.3 can be applied. Item parameters estimated from multidimensional IRT models can be approximated by the following equations (Brossman, 2010). Approximated discrimination parameter estimates are computed by

$$\hat{a}_{wi} = (1 + \hat{\sigma}_{wi}^2)^{-\frac{1}{2}} \hat{\mathbf{a}}_i^T \hat{\Sigma} \hat{\boldsymbol{w}}, \tag{2.50}$$

where $\hat{\sigma}_{wi}^2 = \hat{\mathbf{a}}_i^{\mathrm{T}}\hat{\Sigma}\hat{\mathbf{a}}_i - (\hat{\mathbf{a}}_i^{\mathrm{T}}\hat{\Sigma}\hat{\boldsymbol{w}})^2$,

where $\hat{\mathbf{a}}_i$ is an estimate of the multidimensional discrimination vector, $\hat{\Sigma}$ is an estimate of the multidimensional ability covariance matrix, and $\hat{w}_i$ is a vector of weights,

$$\hat{w}_i = \frac{\sum_{i=1}^{n} \hat{a}_{ik}}{\sqrt{\sum_{k=1}^{m}(\sum_{i=1}^{n} \hat{a}_{ik})^2}}, \qquad (2.51)$$

where $\hat{a}_{ik}$ denotes a discrimination parameter for the item i on the dimension k, n is the total number of items, and m is the number of dimensions. Approximated difficulty parameters can be estimated as:

$$\hat{b}_{wi} = \frac{-\hat{d}_{wi}}{\hat{a}_{wi}}, \qquad (2.52)$$

where $\hat{d}_{wi} = (1 + \hat{\sigma}_{wi}^2)^{-\frac{1}{2}}\hat{d}_i$,

where $\hat{d}_i$ is an index for the multidimensional location parameter. Guessing parameters remain the same as in the unidimensional model ($\hat{c}_{wi} = \hat{c}_i$). These approximated item parameters are then used in unidimensional IRT equating.

2.4.2.4 Multidimensional IRT Equating

Brossman (2010) proposed a procedure for conducting observed score equating under the MIRT framework. Extensions of unidimensional IRT equating presented in Section 2.4.1.3 are applied for full MIRT observed score equating. Instead of using a single ability level θ_j, a vector of ability levels $\boldsymbol{\theta}_j$ was applied as follows:

$$\begin{aligned} f_r(x|\boldsymbol{\theta}_j) &= f_{r-1}(x|\boldsymbol{\theta}_j)(1 - p_{rj}) && x = 0 \\ &= f_{r-1}(x|\boldsymbol{\theta}_j)(1 - p_{rj}) + f_{r-1}(x-1|\boldsymbol{\theta}_j)p_{rj} && 0 < x < r \\ &= f_{r-1}(x-1|\boldsymbol{\theta}_j)p_{rj} && x = r \end{aligned} \qquad (2.53)$$

The same logic as in the unidimensional IRT framework is applied; the conditional observed score distributions $f_r\left(x\middle|\boldsymbol{\theta}_j\right)$ are multiplied by the multivariate ability density $\psi(\boldsymbol{\theta})$, which can be expressed as

$$f(x) = \int_1 \int_2 \ldots \int_{\mathrm{m}} f(x|\boldsymbol{\theta})\psi(\boldsymbol{\theta})d\boldsymbol{\theta} \tag{2.54}$$

or

$$f(x) = \sum_1 \sum_2 \ldots \sum_{\mathrm{m}} f(x|\boldsymbol{\theta})\,\psi(\boldsymbol{\theta}) \tag{2.55}$$

where m indicates the number of dimensions. After this procedure, traditional equipercentile equating is also applied as in unidimensional IRT observed score equating.

CHAPTER 3

COMPARING SUBTEST SCORE EQUATING METHODS UNDER THE CLASSICAL TEST THEORY FRAMEWORK

3.1 Introduction

When tests are composed of several subtests, reporting meaningful subtest scores that provide additional information to the total score depends on the distinctiveness of each subtest. If subtests are highly related to each other and the total test measures a single primary factor, reporting a total score would suffice. On the other hand, when subtests are unique, the information provided by each subtest score is not duplicated by the total score. Thus, those individual subtest scores should be reported. Likewise, the decision to report subtest scores is pertinent to the relationship among subtest scores as well as the relationship between the total score and subtest scores.

Whether to report individual subtest scores also plays a key role in subtest score equating. Equating has usually been performed at the total score level. However, if subtest scores are reported and the comparability of multiple subtest forms is a concern, equating at the subtest score level should be carefully considered. Given high or nearly perfect correlations among subtests, equating at the total score level offers the best choice. On the other hand, given low or close to zero correlations among subtests, separate equating for each subtest provides the best option. Then, if subtests are not so highly related to each other but still have a relationship between them, can we choose one of the two methods? Given a moderate amount of correlations, however, total score equating or separate equating may not afford an appropriate solution because either method does not take into account the relationship of the subtests. It is not clear what high, low, and moderate correlations indicate. Thus, this study included various correlation values to examine which method produces the most accurate equating results under various correlation values. In brief, depending on the strength of the relationship—correlations among dimensions that differentiate subtests—appropriate methods should be applied: equating using total scores, separate equating for each subtest, and subtest score equating taking into account the relationship of the subtests.

Previous studies (Yen, 1987; Pommerich, Nicewander, & Hanson, 1999; Wainer et al., 2001; Shin, 2007) considered the correlation among subtests as an important factor affecting subtest score estimation results. However, there has been little research on the impact of correlation in the context of subtest score equating. Sinharay and Haberman (2011) selected three levels of correlation—0.7, 0.8, and 0.9—among the components of $\boldsymbol{\theta}$ based on the operational data. They concluded that as correlation increases, using the equated total score as the anchor produced less error when performing subtest score equating.

In addition to correlations, other important factors including group differences in their abilities are examined in the current study. Many testing programs use a common item nonequivalent equating design for their equating studies. In this design, two different test forms as well as two different groups of examinees are included; each form is administered to each group. Commonly, ability distributions of two groups are not identical in terms of shape and location. Skewed distributions were also often found in many testing programs (Kolen, 1985).

Previous studies reported that differences in ability of the two groups affect the accuracy of equating results (Skaggs & Lissitz, 1986; Lawrence & Dorans, 1990;

Swediati, 1997; Wang, Lee, Brennan, & Kolen, 2008; Meng, 2012). Chen et al. (2010) compared classical equating methods, including the Levine observed score method and the Tucker method, varying differences in ability and form difficulty. They concluded that when differences are small, both methods produce similar results. Similarly, Kolen (1990) suggested that if two populations are similar in terms of ability and the correlation between scores on common items and total scores is high, it is likely that all equating methods yield similar results (Kolen, 1990).

The purposes of this study are to investigate the impact of the correlation levels as well as the different proficiency distributions involving mean shifts and skewness on subtest score equating methods. This chapter focuses only on the methods based on classical test theory. The following chapter will describe studies applying IRT equating methods.

3.2 Purpose of the Study

Several methods within classical test theory framework are evaluated and compared in this study. Important factors when considering subtest score equating are included: (1) correlations among subtests, (2) the number of items and common items in each subtest, and (3) different proficiency distributions between two groups of examinees. The primary purpose of this study is to examine the impacts of these three variables on subtest score equating. A simulation study was chosen so that these key variables could be manipulated.

Prior to equating, the added value of subtest scores was examined via PRMSE values and multidimensional scaling analysis. The underlying assumption is that as correlation among dimensions increases, subtest scores are less likely to provide additional information to the total score.

Under the classical test theory framework, three different scores (observed scores, weighted averages, and augmented scores) and three different anchor sets (subtest anchor score, equated total score, and anchor total score) were used in subtest score equating. Specific research questions are as follows:

1. Which approach performs better: using observed scores, weighted averages, or augmented scores under each condition?

2. Which method (using subtest anchor score, equated total score, or anchor total score as the anchor) produces the most accurate equating results under each condition?

    2-1. When subtests are highly correlated, does using total or anchor total scores as the anchor produce more accurate equating results than using subtest anchor scores?

    2-2. When the number of total items or common items is relatively small, does using total or anchor total scores as the anchor yield better equating results than using subtest anchor scores?

3. Does the difference between proficiency distributions in two groups have an impact on equating results:

    3-1. when two groups of examinees have different proficiency distributions including mean shifts or skewness, which approach performs better—using observed scores, weighted averages, or augmented scores?

3-2. when the ability distributions differ in two groups, which method produces the most accurate equating results among using the subtest anchor score, equated total score, or anchor total score as the anchor?

3.3 Methodology

3.3.1 Data Generation

Real data from an operational testing program were obtained and used to estimate item parameters. Two test forms had three content areas including reading, science, and math. The sample size for each test form was over 50,000. To create a data set that fits a multidimensional IRT model, these three different subject tests were combined into one test and treated as one multidimensional test. Item parameters were estimated based on multidimensional 3PL model. Note that item parameters in different subject tests were separately calibrated in reality. Using this procedure, however, multidimensional item parameters could be obtained and used to simulate item responses. Only dichotomous items were of interest in this study. Descriptive statistics from the real data set are presented in Table 3.1; correlations among the content areas and the total score are displayed in Table 3.2.

Table 3.1. Descriptive Statistics of Real Data

	Total		Reading		Science		Math	
	FormX	FormY	FormX	FormY	FormX	FormY	FormX	FormY
Min	14	13	0	0	0	0	0	0
Max	106	106	36	36	38	38	32	32
Mean	77.25	77.53	28.86	27.98	26.70	26.64	21.68	22.92
SD	19.02	17.80	6.39	6.17	7.02	6.62	7.33	6.86

Table 3.2. Correlations from Real Data

	FormX				FormY		
	Reading	Science	Math		Reading	Science	Math
Total	.903	.931	.917	Total	.893	.914	.911
Reading		.779	.727	Reading		.730	.713
Science			.778	Science			.750

Before estimating item parameters using the real data, as a preliminary study, confirmatory factor analysis was conducted to assess dimensionality by running Mplus (Muthén & Muthén, 2007). Table 3.3 shows fit statistics examined to determine whether three dimensions could be appropriate. Chi-square goodness-of-fit statistics were not included because the large sample size rendered it useless. The Mplus provides the Comparative Fit Index (CFI) and the Tucker Lewis Index (TLI) which have a range from 0 to 1 with higher values indicating better fit. The root mean square error of approximation (RMSEA) is another measure of model fit that Mplus offers. Although there is no single rule to evaluate these indices, Bentler and Bonett (1980) and Hu and Bentler (1999) provided a rule of thumb. They reported that RMSEA values close to .06 or below, CFI values close to .90 or above, and TLI values close to .95 or greater indicate a reasonably good fit. Based on these criteria, the model fit statistics displayed in Table 3.3 show that the three dimensional model fits the data better than the unidimensional model. Thus, item parameters were estimated using a multidimensional model and used to simulate item responses.

Table 3.3. Model Fit Statistics obtained from Mplus with Real Data

Fit Index	FormX		FormY	
	1 Factor	3 Factor	1 Factor	3 Factor
CFI	.886	.911	.876	.934
TLI	.985	.987	.982	.990
RMSEA	.022	.021	.022	.016

* Note : Comparative Fit Index (CFI), Tucker Lewis Index (TLI), and Root Mean Square Error of Approximation (RMSEA)

After estimating item parameters from the real data, three different sets of item parameters for the new and old forms were created. Item parameters are presented in Appendix A. Based on these item parameter estimates, item responses were generated using the equation presented in (2.41). Note that a simple structure was used, where each item was loaded on only one dimension. Theta values were sampled from a multivariate normal distribution with the mean vector **0** given four different correlations, 0.4, 0.6, 0.8 and 0.9. Instead of using only one value for the correlations among subtests, slightly different correlations were adopted to mimic a more realistic situation (Sinharay & Haberman, 2011a). The real correlations presented in Table 3.2 were used as a starting point. The original correlation matrices are C_X for the new form and C_Y for the old form. That is,

$$C_X = \begin{bmatrix} 1 & .779 & .727 \\ & 1 & .778 \\ & & 1 \end{bmatrix},\ C_Y = \begin{bmatrix} 1 & .730 & .713 \\ & 1 & .750 \\ & & 1 \end{bmatrix}.$$

Off-diagonals were recalculated using *C*-*m*+ ρ where ρ denotes the correlations among the dimensions (0.4, 0.6, 0.8 and 0.9) and *m* indicates the mean of the correlations in Table 3.2. The computed values were not exactly 0.4, 0.6, 0.8 or 0.9, but approximated these values very closely.

To generate theta values, an R package called 'mvtnorm' was used. First, theta values for two groups taking FormX and FormY were generated based on mean vectors of 0s and standard deviation of 1s with four different correlations. For FormX, different sets of theta values were also generated using $\mu_{\mathrm{P}} = (0.1, 0.1, 0.1)$, and $\mu_{\mathrm{P}} = (0, 0.1, -0.1)$. To simulate skewness while preserving correlations among dimensions, theta values obtained from the normal distribution with $\mu_{\mathrm{P}} = (0, 0, 0)$ were transformed using the constants presented in Fleishman (1978). Fleishman's method, also called the power method, uses a polynomial transformation via the following equation:

$$Y = -c + bX + cX^2 + dX^3. \tag{3.1}$$

Parameters of *b* and *d* were 1.11251460 and -0.05033445 when the skewness was .75, and parameter *c* was 0.17363002 for the positively skewed distribution whereas *c* was -0.17363002 for the negatively skewed distribution. When the skewness was .25, the parameters of *b*, *c*, and *d* were 1.008964263, 0.042632745 (or -0.042632745 for the negatively skewed distribution), and -0.003607528, respectively.

Data generation was performed using the computer program *R*, version 3.1.2 (R Development Core Team, 2014). Steps for the data simulation include:

Step 1: Obtain two real data sets (three subject tests administered in 2010 and 2011) from an operational testing program, combine three separate subject tests into one test to create a multidimensional data set, and conduct confirmatory factor analysis using Mplus (Muthén & Muthén, 2007) to assess dimensionality.

Step 2: Estimate item parameters of the data obtained in the previous step using flexMIRT (Cai, 2012). Item parameters were estimated from a multidimensional 3PL model.

Step 3: Sample ability parameters based on multivariate normal distribution for the sample size of 1,000 using four different correlation sets (0.4, 0.6, 0.8, 0.9). This was performed for two distinctive simulee groups taking the new form (FormX) and old form (FormY) separately. The mean vector of FormY was $\mu_{\mathrm{P}} = (0, 0, 0)$. For FormX, three different mean vectors were considered— $\mu_{\mathrm{P}} = (0, 0, 0)$ $\mu_{\mathrm{P}} = (0.1, 0.1, 0.1)$, and $\mu_{\mathrm{P}} = (0, 0.1, -0.1)$. The standard deviation was kept constant at 1. The normal distribution with the mean vectors of **0** was transformed to generate either positively or negatively skewed distributions using Fleishman's method (1978). This transformation was conducted only for FormX.

Step 4: Choose item parameters and specify test configuration in terms of the number of items and common items. Out of 106 items from the real data, a total of 96 items or 48 items were selected. Two different subtest lengths were considered—each subtest had either 32 or 16 items. For internal common items, 12.5% or 25% of the total test was selected.

Step 5: Generate item response data by applying equation (2.41) for each condition based on the multidimensional three parameter logistic model; 100 replications were performed.

3.3.2 Simulation Conditions

In this study, three factors were manipulated: 1) correlations among dimensions (0.4, 0.6, 0.8, and 0.9), 2) test length and proportions of internal common items (32 items with 4 common items, 32 with 8 common items, and 16 with 4 common items), and 3) ability distributions: normal distributions with mean shift—$\mu_{\mathrm{P}} = (0.1, 0.1, 0.1)$ and $\mu_{\mathrm{P}} = (0, 0.1, -0.1)$, and skewed distributions with skewness of -.75, -.25, .25, and .75, which account for 84 total conditions. Table 3.4 displays the summary of all simulation conditions.

Table 3.4. Summary of Simulation Conditions

Correlation among dimensions (4)	Test configuration (3)	Proficiency distribution of FormX (7)			
0.4	32 items with 8 common items	Normal distribution	μ1	$\mu_P = (0, 0, 0)$	
0.6	32 items with 4 common items		μ2	$\mu_P = (0.1, 0.1, 0.1)$	
0.8	16 items with 4 common items		μ3	$\mu_P = (0, 0.1, -0.1)$	
0.9		Skewed distribution	μ4	Positively skewed	Skewness: 0.75
			μ5	Negatively skewed	Skewness: -0.75
			μ6	Positively skewed	Skewness: 0.25
			μ7	Negatively skewed	Skewness: -0.25

Correlations among dimensions were manipulated to represent relatively low correlation 0.4; moderate correlation 0.6; relatively high correlation 0.8; and high correlation 0.9. Four correlations from low to high were chosen to investigate whether there was a systematic pattern as correlations either decreased or increased. In addition to correlations, different test lengths (relatively short and moderate) and common item sets (small and moderate proportions of common items) were considered: 32 items with 4 or 8 common items and 16 items with 4 common items for each subtest.

Different ability distributions were chosen to simulate group difference in their abilities. Starting from the normal distribution with the mean vector of 0s, mean values were shifted in the same direction, $\mu_P = (0.1, 0.1, 0.1)$, or to the opposite directions $\mu_P =$

$(0, 0.1, -0.1)$. This study also considered the condition where proficiency distributions were either negatively or positively skewed for one group of examinees.

3.3.3 Analysis

3.3.3.1 PRMSE and MDS

Proportional reduction in mean square error and MDS solutions were examined to determine whether subtest scores are worth reporting along with the total score. First, PRMSE values were computed using the equations described in Section 2.2.1. According to Haberman's (2008) criteria, PRMSEs should be greater than PRMSEx so that subtest scores could confer added value, and PRMSE of weighted averages should be larger than both PRMSEs and PRMSEx for weighted averages to have added value. Under each condition, PRMSE values were compared based on the criteria. Since this study used a simulation technique and 100 replications were performed, the number of replications out of 100 showing either subtest scores or weighted averages with added value was also calculated.

In addition to the computation of PRMSE values, MDS analyses were conducted using one of the SPSS MDS programs called ALSCAL (Young & Harris, 1993). First, the simulated item responses were transformed to dissimilarity data. This study used derived proximity data coming from binary item responses. Derived proximities could be obtained in two different formats: distances and correlations. Using ALSCAL, distances among items were computed using $\delta_{jj'} = \sqrt{\sum_{j=1}^{N}(\mathrm{u_{ij}} - \mathrm{u_{i'j}})^2}$ where $\mathrm{u_{ij}}$ indicates the score on item *i* for examinee *j*. For the correlation data, tetrachoric correlations were first computed with the R package called *psych* (Revelle, 2013). Correlations (r_{ij}) were transformed to dissimilarities using $\delta_{ij} = \sqrt{2 - 2r_{ij}}$. Using 100 sets of the derived data, classical MDS was performed.

Two fit indices—STRESS and RSQ—were used to evaluate the model-data fit. Average values for STRESS and RSQ were computed over 100 replicated data sets. As a rule of thumb, STRESS values below .10 and RSQ values above .90 indicate good fit (Kruskal & Wish, 1978). It could prove difficult to determine one absolute solution through MDS; however, fit statistics and visual inspection could provide a general pattern of the data as correlations among dimensions either increased or decreased.

3.3.3.2 <u>Traditional equating</u>

Before conducting equating, weighted averages and augmented scores were computed using equations (2.7) and (2.18) described in the previous chapter. These scores were then used as input in equating. For the equating method, chained equipercentile equating under the common item nonequivalent groups design was implemented using the *equate* package in R (Albano, 2011). A polynomial log linear smoothing method (Holland & Thayer, 1987; 2000) was applied to smooth the frequency distribution prior to equating.

This study adopted three approaches for subtest score equating (see Figure 3.1). For the first approach, observed scores including total scores, subtest scores, and subtest anchor scores of the two forms (FormX and FormY) were computed. In Method 1, equating was conducted using scores from each subtest and its anchor score. Note that three separate equating procedures were performed as there were three subtests. In Method 2, three subtests were treated as one test, and equating was conducted at the total test level to obtain an equated total score of FormX (new form). The equated total score was then used as the anchor, and equating at the subtest level was performed. Method 3 used anchor total scores as the anchor. Likewise, three different anchor sets were employed to equate subtest scores. For the second approach, weighted averages of three subtests for each examinee were computed by applying equation (2.7). For the third approach, augmented scores of three subtests were calculated via equation (2.18). Identical three methods having different anchor sets were also applied under these two approaches.

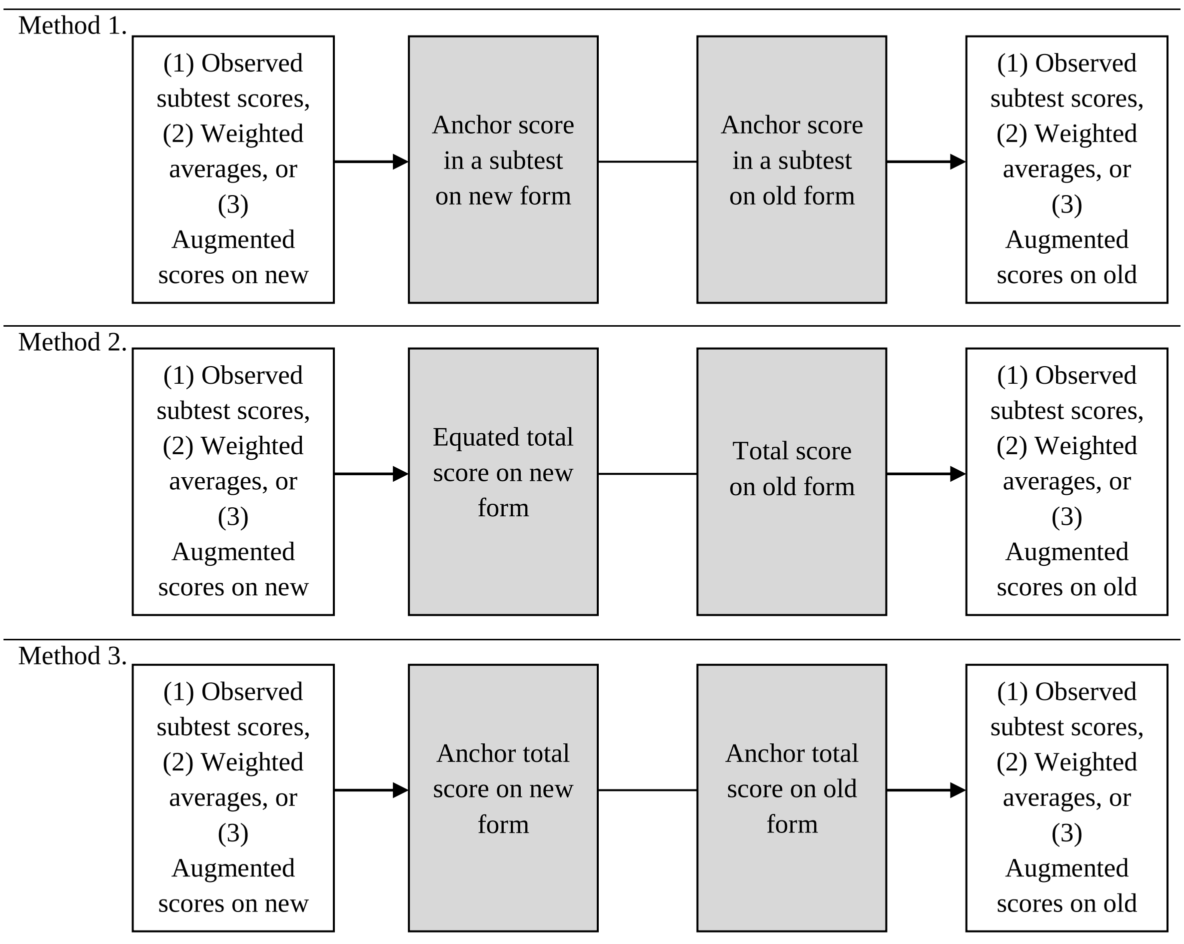


Figure 3.1. Overview of Subtest Score Equating Methods using Traditional Equating

3.3.4 Criteria

Based on the original item parameters and 2,001 quadrature points from -6.0 to 6.0, probabilities of answering each item correctly were computed using the equation in (2.41). Summed probabilities at each θ level, treated as true score estimates, for both FormX and FormY were compared to find a true equating function. Regardless of the simulation conditions, an identical true equating function was adopted as criteria. Depending on the number of items and common items, however, different true equating functions were used because true item parameters for simulating item responses were different.

Based on the true equating function, bias and root mean squared error (RMSE) were computed via the following equations:

$$\text{Bias}(x) = \frac{1}{100}\sum_{r=1}^{100}[\hat{e}_r(x) - e(x)] \tag{3.2}$$

and

$$\text{RMSE}(x) = \sqrt{\left\{\frac{1}{100}\sum_{r=1}^{100}[\hat{e}_r(x) - e(x)]^2\right\}} \tag{3.3}$$

where $\hat{e}_r(x)$ denotes the equating function in the r^{th} replication for the score x, 100 indicates the number of replications, and $e(x)$ stands for the true equating function for the corresponding score x. In addition to bias and RMSE, absolute score differences were computed. Average bias, RMSE, absolute difference in the score range where 95% of examinees fall were computed. The concept of score difference that matters (DTM; Dorans & Feigenbaum, 1994) was also applied for interpretation of equating results. Previous studies adopted the notion of DTM to evaluate the magnitude of differences of equated scores. A difference of .5 is considered as significant since it indicates a change in a reporting score. The number of score points equal in both the true and estimated conversion tables was computed to check how dissimilar or similar the two conversion tables were. First, whether each score point in the estimated conversion table—for example, 33 raw score points when the number of items was 32—was equal to the score points in the conversion table obtained from the true equating function was examined. Percentages of matching score points both in the true and estimated conversion tables were then averaged over 100 replications.

3.4 Results

3.4.1 Descriptive statistics

Table 3.5 and Table 3.6 present mean and standard deviations of simulated data sets. These values were averaged over 100 replications. In general, subtest 1 had the smallest mean difference between FormX and FormY whereas subtest 3 had the largest difference. In subtest 3, the mean of FormX was smaller than that of FormY. In the other two subtests, FormY had larger mean values in most conditions. In subtest 1, the mean difference between the two forms was the largest when the mean vector was shifted from $\mu1 = (0,0,0)$ to $\mu2 = (0.1,0.1,0.1)$. In subtest 3, the differences were the largest when the mean vector was $\mu3 = (0,0.1,-0.1)$. In subtest 2, when there were mean shifts ($\mu2$ and $\mu3$), the mean difference between the two forms was larger than in the other distributions. For the skewed distributions ($\mu4$ with skewness of .75, $\mu5$ with -.75, $\mu6$ with .25, and $\mu7$ with -.25), subtest 1 had the smallest mean score difference between the two forms. Subtest 2 had the largest difference when distributions were negatively skewed with skewness of -.75. In other conditions, subtest 3 had the largest mean difference. Subtest 3 also showed the largest standard deviation among the three subtests.

Table 3.5. Mean of Observed Scores

ρ	Form	Ability Distribution	32 Items with 4 Common Items Total	Subtest1	Subtest2	Subtest3	32 Items with 8 Common Items Total	Subtest1	Subtest2	Subtest3	16 Items with 4 Common Items Total	Subtest1	Subtest2	Subtest3
0.9	Y	μ1(0,0,0)	70.55	25.00	22.85	22.69	70.55	25.00	22.85	22.69	35.27	12.55	11.26	11.46
	X	μ1(0,0,0)	70.37	25.05	23.29	22.03	70.16	25.12	23.14	21.90	35.03	12.50	11.44	11.09
		μ2(0.1,0.1,0.1)	72.07	25.56	23.82	22.69	71.87	25.61	23.70	22.56	35.95	12.77	11.75	11.43
		μ3(0,0.1, −0.1)	70.31	25.08	23.84	21.39	70.09	25.13	23.70	21.25	35.02	12.51	11.74	10.77
		μ4 .75 skewness	69.59	24.89	23.08	21.63	69.40	24.97	22.92	21.51	34.65	12.41	11.34	10.90
		μ5 -.75 skewness	70.99	25.17	23.46	22.37	70.77	25.22	23.33	22.21	35.38	12.55	11.56	11.26
		μ6 .25 skewness	70.16	25.01	23.23	21.92	69.93	25.06	23.07	21.80	34.95	12.49	11.43	11.04
		μ7-.25 skewness	70.55	25.08	23.33	22.13	70.34	25.15	23.21	21.99	35.15	12.52	11.48	11.15
0.8	Y	μ1(0,0,0)	70.63	25.05	22.86	22.72	70.63	25.05	22.86	22.72	35.31	12.56	11.27	11.48
	X	μ1(0,0,0)	70.41	25.06	23.32	22.03	70.20	25.12	23.19	21.89	35.06	12.52	11.47	11.08
		μ2(0.1,0.1,0.1)	71.94	25.52	23.79	22.64	71.77	25.59	23.66	22.52	35.86	12.75	11.72	11.39
		μ3(0,0.1, −0.1)	70.22	25.04	23.80	21.37	70.00	25.10	23.66	21.23	34.99	12.50	11.72	10.77
		μ4 .75 skewness	69.63	24.91	23.09	21.63	69.42	24.96	22.96	21.50	34.67	12.43	11.35	10.89
		μ5 -.75 skewness	71.06	25.18	23.52	22.35	70.81	25.22	23.37	22.22	35.41	12.57	11.58	11.26
		μ6 .25 skewness	70.18	25.02	23.24	21.91	69.99	25.08	23.12	21.78	34.96	12.49	11.44	11.04
		μ7-.25 skewness	70.60	25.11	23.36	22.13	70.39	25.17	23.23	22.00	35.16	12.53	11.50	11.13
0.6	Y	μ1(0,0,0)	70.67	25.04	22.90	22.74	70.67	25.04	22.90	22.74	35.33	12.57	11.28	11.49
	X	μ1(0,0,0)	70.50	25.11	23.32	22.07	70.30	25.17	23.18	21.95	35.11	12.53	11.47	11.11
		μ2(0.1,0.1,0.1)	71.98	25.52	23.79	22.67	71.78	25.59	23.66	22.53	35.88	12.74	11.73	11.41
		μ3(0,0.1, −0.1)	70.25	25.06	23.82	21.37	70.08	25.13	23.70	21.25	35.00	12.51	11.73	10.75
		μ4 .75 skewness	69.71	24.94	23.09	21.68	69.48	24.98	22.96	21.55	34.71	12.44	11.37	10.90
		μ5 -.75 skewness	71.14	25.22	23.51	22.40	70.96	25.28	23.39	22.28	35.46	12.59	11.59	11.28
		μ6 .25 skewness	70.25	25.04	23.25	21.96	70.05	25.11	23.11	21.82	35.00	12.50	11.44	11.06
		μ7-.25 skewness	70.70	25.15	23.38	22.18	70.48	25.20	23.25	22.04	35.23	12.54	11.51	11.17
0.4	Y	μ1(0,0,0)	70.57	25.01	22.87	22.69	70.57	25.01	22.87	22.69	35.29	12.56	11.27	11.47
	X	μ1(0,0,0)	70.36	25.08	23.28	22.00	70.15	25.12	23.15	21.87	35.06	12.52	11.46	11.08
		μ2(0.1,0.1,0.1)	72.10	25.57	23.84	22.69	71.91	25.65	23.71	22.55	35.95	12.77	11.76	11.42
		μ3(0,0.1, −0.1)	70.31	25.05	23.85	21.42	70.11	25.11	23.72	21.28	35.03	12.51	11.74	10.78
		μ4 .75 skewness	69.61	24.91	23.08	21.62	69.38	24.98	22.92	21.49	34.64	12.42	11.34	10.88
		μ5 -.75 skewness	70.99	25.17	23.48	22.34	70.80	25.24	23.35	22.21	35.38	12.57	11.56	11.25
		μ6 .25 skewness	70.16	25.02	23.23	21.91	69.94	25.08	23.08	21.78	34.94	12.49	11.42	11.03
		μ7-.25 skewness	70.57	25.12	23.35	22.10	70.35	25.17	23.21	21.97	35.16	12.53	11.49	11.14

Table 3.6. Standard Deviation of Observed Scores

ρ	Form	Ability Distribution	32 Items with 4 Common Items				32 Items with 8 Common Items				16 Items with 4 Common Items			
			Total	Subtest1	Subtest2	Subtest3	Total	Subtest1	Subtest2	Subtest3	Total	Subtest1	Subtest2	Subtest3
0.9	Y	μ1 (0,0,0)	16.71	5.53	5.90	6.71	16.71	5.53	5.90	6.71	8.60	2.93	3.33	3.41
	X	μ1 (0,0,0)	17.32	5.91	5.86	6.99	17.45	5.93	5.97	7.02	9.02	3.22	3.21	3.67
		μ2 (0.1,0.1,0.1)	16.85	5.69	5.74	6.85	16.99	5.71	5.83	6.88	8.77	3.10	3.15	3.58
		μ3 (0,0.1, −0.1)	17.25	5.90	5.71	7.11	17.40	5.92	5.83	7.13	8.98	3.21	3.14	3.72
		μ4 .75 skewness	16.14	5.35	5.55	6.77	16.24	5.33	5.65	6.81	8.46	2.96	3.08	3.57
		μ5 -.75 skewness	18.43	6.39	6.21	7.30	18.58	6.41	6.32	7.35	9.55	3.45	3.36	3.81
		μ6 .25 skewness	17.05	5.78	5.78	6.94	17.20	5.80	5.90	6.97	8.89	3.16	3.18	3.64
		μ7-.25 skewness	17.62	6.06	5.94	7.07	17.76	6.06	6.05	7.11	9.16	3.28	3.26	3.70
0.8	Y	μ1 (0,0,0)	16.03	5.47	5.88	6.69	16.03	5.47	5.88	6.69	8.26	2.91	3.32	3.39
	X	μ1 (0,0,0)	16.67	5.92	5.85	6.99	16.81	5.94	5.95	7.02	8.69	3.22	3.21	3.66
		μ2 (0.1,0.1,0.1)	16.31	5.72	5.75	6.87	16.43	5.73	5.85	6.91	8.50	3.11	3.15	3.60
		μ3 (0,0.1, −0.1)	16.71	5.93	5.75	7.12	16.85	5.94	5.85	7.15	8.72	3.23	3.16	3.72
		μ4 .75 skewness	15.52	5.34	5.55	6.78	15.63	5.34	5.65	6.81	8.15	2.95	3.08	3.57
		μ5 -.75 skewness	17.67	6.39	6.17	7.30	17.85	6.43	6.30	7.34	9.17	3.44	3.35	3.79
		μ6 .25 skewness	16.42	5.78	5.79	6.93	16.55	5.80	5.86	6.97	8.56	3.15	3.17	3.63
		μ7-.25 skewness	16.95	6.04	5.95	7.05	17.09	6.06	6.05	7.09	8.82	3.28	3.24	3.70
0.6	Y	μ1 (0,0,0)	14.79	5.49	5.87	6.68	14.79	5.49	5.87	6.68	7.65	2.91	3.32	3.40
	X	μ1 (0,0,0)	15.36	5.89	5.86	6.98	15.48	5.89	5.97	7.01	8.04	3.21	3.21	3.65
		μ2 (0.1,0.1,0.1)	14.99	5.74	5.73	6.85	15.12	5.74	5.84	6.89	7.84	3.12	3.15	3.59
		μ3 (0,0.1, −0.1)	15.37	5.92	5.75	7.10	15.47	5.91	5.84	7.14	8.03	3.21	3.15	3.72
		μ4 .75 skewness	14.31	5.33	5.55	6.77	14.42	5.34	5.65	6.81	7.54	2.95	3.08	3.57
		μ5 -.75 skewness	16.21	6.36	6.18	7.28	16.32	6.37	6.28	7.31	8.43	3.42	3.35	3.79
		μ6 .25 skewness	15.14	5.77	5.77	6.93	15.25	5.76	5.87	6.97	7.93	3.14	3.17	3.64
		μ7-.25 skewness	15.57	6.02	5.93	7.05	15.73	6.03	6.04	7.10	8.13	3.26	3.23	3.68
0.4	Y	μ1 (0,0,0)	13.50	5.51	5.89	6.70	13.50	5.51	5.89	6.70	7.03	2.92	3.33	3.40
	X	μ1 (0,0,0)	13.96	5.90	5.85	6.99	14.10	5.94	5.95	7.01	7.36	3.21	3.21	3.66
		μ2 (0.1,0.1,0.1)	13.63	5.71	5.71	6.85	13.74	5.70	5.82	6.89	7.18	3.11	3.14	3.58
		μ3 (0,0.1, −0.1)	13.97	5.92	5.74	7.11	14.09	5.92	5.83	7.14	7.36	3.22	3.14	3.72
		μ4 .75 skewness	13.02	5.34	5.55	6.76	13.12	5.34	5.65	6.81	6.92	2.95	3.08	3.57
		μ5 -.75 skewness	14.71	6.40	6.17	7.29	14.83	6.41	6.29	7.32	7.70	3.44	3.34	3.80
		μ6 .25 skewness	13.76	5.76	5.77	6.94	13.89	5.79	5.87	6.97	7.26	3.15	3.18	3.63
		μ7-.25 skewness	14.18	6.04	5.93	7.05	14.29	6.05	6.04	7.09	7.45	3.27	3.25	3.69

Table 3.7 shows correlations of the total score and each subtest score as well as correlations among the three subtest scores. Correlations between the total score and each subtest score fell between 0.934 and 0.682, and correlations among subtest scores were ranged from 0.805 to 0.253. Correlations were lower when the number of items in each subtest was 16 compared to a larger number of items (32 items).

Table 3.8 presents disattenuated correlations among subtest scores. Compared to Table 3.**7**, correlations were closer to the original values which were used in simulating item responses.

Table 3.7. Correlations of Total and Subtest Scores

ρ		Ability Distribution	32 Items with 4 Common Items						32 Items with 8 Common Items						16 Items with 4 Common Items					
			Total			Sub1		Sub2	Total			Sub1		Sub2	Total			Sub1		Sub2
			Sub1	Sub2	Sub3	Sub1	Sub2	Sub3	Sub1	Sub2	Sub3	Sub2	Sub3	Sub3	Sub1	Sub2	Sub3	Sub2	Sub3	Sub3
0.9	Y	μ1 (0,0,0)	0.902	0.926	0.934	0.761	0.755	0.800	0.902	0.926	0.934	0.761	0.755	0.800	0.863	0.900	0.901	0.668	0.663	0.717
	X	μ1 (0,0,0)	0.910	0.928	0.930	0.785	0.750	0.797	0.907	0.929	0.929	0.782	0.746	0.798	0.881	0.896	0.903	0.698	0.678	0.714
		μ2 (0.1,0.1,0.1)	0.907	0.928	0.930	0.781	0.747	0.796	0.906	0.928	0.930	0.780	0.747	0.797	0.878	0.896	0.902	0.695	0.674	0.713
		μ3 (0,0.1, −0.1)	0.908	0.926	0.929	0.784	0.744	0.792	0.907	0.928	0.928	0.786	0.741	0.795	0.879	0.892	0.902	0.698	0.672	0.709
		μ4 .75 skewness	0.893	0.918	0.926	0.751	0.723	0.775	0.891	0.920	0.926	0.751	0.722	0.777	0.859	0.882	0.895	0.655	0.640	0.684
		μ5 -.75 skewness	0.915	0.932	0.930	0.797	0.755	0.805	0.913	0.934	0.929	0.797	0.751	0.806	0.891	0.903	0.905	0.722	0.693	0.729
		μ6 .25 skewness	0.907	0.926	0.929	0.778	0.746	0.793	0.906	0.928	0.929	0.778	0.745	0.796	0.878	0.892	0.902	0.689	0.675	0.707
		μ7-.25 skewness	0.912	0.929	0.929	0.789	0.752	0.799	0.910	0.931	0.929	0.788	0.750	0.801	0.883	0.899	0.903	0.705	0.681	0.721
0.8	Y	μ1 (0,0,0)	0.865	0.892	0.903	0.673	0.662	0.707	0.865	0.892	0.903	0.673	0.662	0.707	0.829	0.871	0.871	0.589	0.582	0.634
	X	μ1 (0,0,0)	0.873	0.892	0.899	0.693	0.657	0.704	0.872	0.895	0.899	0.695	0.655	0.708	0.848	0.863	0.874	0.617	0.596	0.632
		μ2 (0.1,0.1,0.1)	0.871	0.893	0.900	0.693	0.655	0.706	0.870	0.894	0.900	0.692	0.653	0.706	0.844	0.863	0.875	0.611	0.592	0.633
		μ3 (0,0.1, −0.1)	0.873	0.890	0.901	0.693	0.657	0.705	0.872	0.893	0.901	0.696	0.656	0.707	0.849	0.862	0.875	0.621	0.594	0.632
		μ4 .75 skewness	0.854	0.881	0.896	0.658	0.629	0.681	0.852	0.882	0.896	0.656	0.626	0.681	0.824	0.850	0.868	0.575	0.559	0.603
		μ5 -.75 skewness	0.877	0.894	0.897	0.699	0.656	0.707	0.876	0.896	0.897	0.700	0.654	0.709	0.855	0.868	0.874	0.631	0.601	0.640
		μ6 .25 skewness	0.870	0.891	0.900	0.687	0.654	0.702	0.870	0.892	0.900	0.689	0.653	0.704	0.843	0.860	0.874	0.607	0.590	0.627
		μ7-.25 skewness	0.875	0.894	0.899	0.697	0.658	0.708	0.873	0.896	0.900	0.696	0.657	0.712	0.850	0.864	0.874	0.621	0.595	0.634
0.6	Y	μ1 (0,0,0)	0.790	0.820	0.842	0.498	0.488	0.527	0.790	0.820	0.842	0.498	0.488	0.527	0.759	0.808	0.810	0.436	0.427	0.469
	X	μ1 (0,0,0)	0.798	0.820	0.839	0.518	0.477	0.527	0.796	0.821	0.839	0.515	0.479	0.528	0.779	0.798	0.816	0.460	0.432	0.474
		μ2 (0.1,0.1,0.1)	0.795	0.816	0.838	0.511	0.474	0.521	0.794	0.820	0.838	0.513	0.473	0.524	0.776	0.795	0.813	0.456	0.426	0.464
		μ3 (0,0.1, −0.1)	0.798	0.813	0.841	0.515	0.477	0.522	0.796	0.815	0.840	0.514	0.475	0.522	0.778	0.791	0.817	0.459	0.429	0.466
		μ4 .75 skewness	0.777	0.808	0.838	0.486	0.454	0.503	0.774	0.811	0.837	0.488	0.451	0.504	0.754	0.783	0.813	0.421	0.404	0.442
		μ5 -.75 skewness	0.800	0.817	0.833	0.514	0.471	0.522	0.799	0.818	0.832	0.513	0.470	0.519	0.785	0.797	0.811	0.467	0.432	0.469
		μ6 .25 skewness	0.797	0.818	0.840	0.517	0.478	0.525	0.795	0.819	0.839	0.515	0.477	0.523	0.776	0.796	0.816	0.458	0.429	0.467
		μ7-.25 skewness	0.800	0.819	0.837	0.519	0.477	0.526	0.798	0.822	0.837	0.519	0.477	0.529	0.782	0.797	0.815	0.463	0.436	0.473
0.4	Y	μ1 (0,0,0)	0.707	0.744	0.779	0.326	0.316	0.353	0.707	0.744	0.779	0.326	0.316	0.353	0.683	0.747	0.747	0.286	0.272	0.317
	X	μ1 (0,0,0)	0.715	0.737	0.775	0.337	0.302	0.350	0.717	0.742	0.774	0.344	0.304	0.352	0.705	0.726	0.754	0.304	0.273	0.315
		μ2 (0.1,0.1,0.1)	0.714	0.739	0.779	0.339	0.305	0.354	0.712	0.743	0.778	0.342	0.305	0.355	0.703	0.726	0.757	0.302	0.275	0.316
		μ3 (0,0.1, −0.1)	0.715	0.730	0.780	0.335	0.303	0.349	0.715	0.734	0.780	0.340	0.305	0.350	0.704	0.719	0.761	0.300	0.275	0.316
		μ4 .75 skewness	0.693	0.730	0.778	0.318	0.284	0.335	0.690	0.733	0.778	0.317	0.282	0.336	0.682	0.716	0.755	0.280	0.253	0.293
		μ5 -.75 skewness	0.719	0.733	0.766	0.333	0.292	0.341	0.718	0.736	0.764	0.334	0.293	0.340	0.710	0.721	0.748	0.301	0.269	0.308
		μ6 .25 skewness	0.714	0.738	0.777	0.342	0.302	0.350	0.712	0.741	0.777	0.340	0.303	0.353	0.701	0.724	0.756	0.297	0.272	0.313
		μ7-.25 skewness	0.719	0.739	0.773	0.341	0.302	0.353	0.717	0.740	0.772	0.340	0.302	0.350	0.707	0.726	0.754	0.305	0.275	0.318

Table 3.8. Disattenuated Correlations among Subtest Scores

ρ		Ability Distribution	32 Items with 4 Common Items			32 Items with 8 Common Items			16 Items with 4 Common Items		
			Sub1		Sub2	Sub1		Sub2	Sub1		Sub2
			Sub1	Sub2	Sub3	Sub2	Sub3	Sub3	Sub2	Sub3	Sub3
0.9	Y	μ1 (0,0,0)	0.887	0.880	0.933	0.887	0.880	0.933	0.884	0.877	0.949
	X	μ1 (0,0,0)	0.908	0.868	0.922	0.902	0.860	0.921	0.908	0.880	0.928
		μ2 (0.1,0.1,0.1)	0.906	0.866	0.923	0.902	0.864	0.922	0.906	0.879	0.930
		μ3 (0,0.1, −0.1)	0.909	0.862	0.918	0.908	0.856	0.918	0.908	0.874	0.922
		μ4 .75 skewness	0.899	0.866	0.928	0.898	0.863	0.930	0.898	0.879	0.939
		μ5 -.75 skewness	0.900	0.852	0.908	0.897	0.845	0.907	0.899	0.863	0.908
		μ6 .25 skewness	0.907	0.869	0.925	0.904	0.865	0.925	0.906	0.887	0.930
		μ7-.25 skewness	0.906	0.863	0.918	0.903	0.859	0.918	0.904	0.873	0.924
0.8	Y	μ1 (0,0,0)	0.786	0.774	0.826	0.786	0.774	0.826	0.782	0.771	0.841
	X	μ1 (0,0,0)	0.801	0.760	0.814	0.802	0.755	0.817	0.801	0.775	0.821
		μ2 (0.1,0.1,0.1)	0.803	0.759	0.818	0.800	0.755	0.816	0.795	0.771	0.824
		μ3 (0,0.1, −0.1)	0.801	0.760	0.815	0.803	0.757	0.816	0.806	0.770	0.819
		μ4 .75 skewness	0.788	0.753	0.815	0.784	0.748	0.814	0.791	0.767	0.828
		μ5 -.75 skewness	0.790	0.740	0.798	0.787	0.736	0.798	0.785	0.748	0.798
		μ6 .25 skewness	0.800	0.762	0.818	0.802	0.759	0.819	0.799	0.776	0.825
		μ7-.25 skewness	0.800	0.755	0.813	0.797	0.753	0.815	0.796	0.763	0.813
0.6	Y	μ1 (0,0,0)	0.581	0.570	0.615	0.581	0.570	0.615	0.578	0.566	0.622
	X	μ1 (0,0,0)	0.599	0.552	0.610	0.594	0.553	0.610	0.597	0.560	0.615
		μ2 (0.1,0.1,0.1)	0.592	0.549	0.604	0.593	0.547	0.606	0.594	0.554	0.604
		μ3 (0,0.1, −0.1)	0.596	0.552	0.604	0.593	0.549	0.603	0.596	0.557	0.606
		μ4 .75 skewness	0.582	0.544	0.603	0.583	0.538	0.602	0.579	0.555	0.608
		μ5 -.75 skewness	0.580	0.532	0.589	0.578	0.530	0.585	0.582	0.538	0.585
		μ6 .25 skewness	0.602	0.557	0.612	0.599	0.556	0.609	0.603	0.565	0.615
		μ7-.25 skewness	0.596	0.548	0.604	0.595	0.546	0.606	0.595	0.560	0.607
0.4	Y	μ1 (0,0,0)	0.380	0.369	0.411	0.380	0.369	0.411	0.379	0.360	0.420
	X	μ1 (0,0,0)	0.390	0.350	0.406	0.396	0.351	0.407	0.395	0.355	0.409
		μ2 (0.1,0.1,0.1)	0.393	0.354	0.411	0.396	0.352	0.410	0.393	0.358	0.412
		μ3 (0,0.1, −0.1)	0.388	0.350	0.404	0.392	0.352	0.404	0.390	0.358	0.411
		μ4 .75 skewness	0.381	0.340	0.401	0.378	0.337	0.402	0.385	0.347	0.404
		μ5 -.75 skewness	0.376	0.329	0.385	0.376	0.330	0.382	0.376	0.335	0.384
		μ6 .25 skewness	0.399	0.352	0.408	0.395	0.352	0.411	0.391	0.357	0.411
		μ7-.25 skewness	0.391	0.347	0.405	0.389	0.346	0.401	0.391	0.353	0.408

Table 3.9, Table 3.10, Table 3.11, and Table 3.12 display mean and standard deviations of weighted averages and augmented scores. The mean from the weighted averages was close to the mean from the observed scores. The standard deviation of the weighted averages was slightly larger than that of the observed scores. Similarly, mean and standard deviations of the augmented scores were very close to those of the observed scores. When the proficiency distributions were shifted to $\mu 2 = (0.1, 0.1, 0.1)$, the mean differences between FormX and FormY were the largest in subtest 1 whereas the differences were the smallest in subtest 3 in both weighted average and augmented score data sets.

Table 3.9. Mean of Weighted Averages

ρ	Form	Ability Distribution	32 items with 4 common items			32 items with 8 common items			16 items with 4 common items		
			Subtest1	Subtest2	Subtest3	Subtest1	Subtest2	Subtest3	Subtest1	Subtest2	Subtest3
0.9	Y	μ1 (0,0,0)	24.86	22.75	22.48	24.86	22.75	22.48	12.46	11.18	11.34
	X	μ1 (0,0,0)	24.86	23.14	21.90	24.94	23.00	21.78	12.35	11.37	10.99
		μ2 (0.1,0.1,0.1)	25.37	23.67	22.55	25.41	23.56	22.43	12.63	11.67	11.33
		μ3 (0,0.1, −0.1)	24.89	23.67	21.29	24.94	23.52	21.16	12.36	11.65	10.70
		μ4 .75 skewness	24.68	22.89	21.40	24.76	22.72	21.29	12.27	11.25	10.76
		μ5 -.75 skewness	25.05	23.40	22.33	25.11	23.27	22.19	12.42	11.52	11.22
		μ6 .25 skewness	24.81	23.06	21.76	24.86	22.90	21.65	12.34	11.35	10.93
		μ7-.25 skewness	24.91	23.21	22.02	24.98	23.08	21.89	12.38	11.41	11.07
0.8	Y	μ1 (0,0,0)	24.99	22.83	22.63	24.99	22.83	22.63	12.52	11.21	11.42
	X	μ1 (0,0,0)	24.97	23.27	21.97	25.04	23.14	21.84	12.43	11.43	11.03
		μ2 (0.1,0.1,0.1)	25.43	23.73	22.57	25.50	23.61	22.46	12.68	11.68	11.33
		μ3 (0,0.1, −0.1)	24.96	23.74	21.32	25.01	23.60	21.19	12.41	11.67	10.73
		μ4 .75 skewness	24.81	23.02	21.52	24.87	22.88	21.40	12.35	11.31	10.81
		μ5 -.75 skewness	25.14	23.51	22.34	25.18	23.36	22.21	12.51	11.56	11.24
		μ6 .25 skewness	24.93	23.19	21.84	24.99	23.07	21.71	12.41	11.40	10.98
		μ7-.25 skewness	25.03	23.32	22.08	25.09	23.18	21.96	12.45	11.47	11.09
0.6	Y	μ1 (0,0,0)	25.02	22.89	22.71	25.02	22.89	22.71	12.56	11.27	11.48
	X	μ1 (0,0,0)	25.09	23.31	22.06	25.14	23.18	21.94	12.51	11.47	11.10
		μ2 (0.1,0.1,0.1)	25.50	23.79	22.65	25.57	23.66	22.52	12.72	11.73	11.39
		μ3 (0,0.1, −0.1)	25.04	23.81	21.36	25.11	23.69	21.24	12.49	11.72	10.74
		μ4 .75 skewness	24.91	23.08	21.65	24.96	22.94	21.52	12.42	11.36	10.88
		μ5 -.75 skewness	25.22	23.51	22.40	25.28	23.39	22.28	12.57	11.59	11.27
		μ6 .25 skewness	25.02	23.24	21.94	25.09	23.10	21.81	12.48	11.43	11.05
		μ7-.25 skewness	25.13	23.38	22.16	25.18	23.24	22.03	12.52	11.51	11.16
0.4	Y	μ1 (0,0,0)	25.01	22.86	22.68	25.01	22.86	22.68	12.55	11.26	11.46
	X	μ1 (0,0,0)	25.08	23.28	21.99	25.11	23.15	21.87	12.51	11.46	11.08
		μ2 (0.1,0.1,0.1)	25.56	23.84	22.68	25.64	23.70	22.55	12.77	11.76	11.41
		μ3 (0,0.1, −0.1)	25.04	23.84	21.41	25.10	23.71	21.28	12.51	11.74	10.77
		μ4 .75 skewness	24.91	23.08	21.62	24.97	22.91	21.48	12.41	11.35	10.88
		μ5 -.75 skewness	25.16	23.48	22.32	25.22	23.34	22.20	12.56	11.56	11.24
		μ6 .25 skewness	25.02	23.22	21.91	25.07	23.08	21.78	12.48	11.42	11.02
		μ7-.25 skewness	25.11	23.35	22.09	25.16	23.21	21.97	12.53	11.49	11.13

Table 3.10. Standard Deviation of Weighted Averages

ρ	Form	Ability Distribution	32 items with 4 common items			32 items with 8 common items			16 items with 4 common items		
			Subtest1	Subtest2	Subtest3	Subtest1	Subtest2	Subtest3	Subtest1	Subtest2	Subtest3
0.9	Y	μ1 (0,0,0)	6.38	7.45	8.26	6.38	7.45	8.26	3.22	4.07	4.10
	X	μ1 (0,0,0)	6.77	7.56	8.42	6.75	7.68	8.44	3.59	3.97	4.36
		μ2 (0.1,0.1,0.1)	6.50	7.39	8.24	6.49	7.49	8.29	3.43	3.88	4.26
		μ3 (0,0.1, −0.1)	6.73	7.27	8.55	6.73	7.44	8.57	3.56	3.82	4.43
		μ4 .75 skewness	5.97	6.92	8.06	5.93	7.05	8.12	3.15	3.64	4.14
		μ5 -.75 skewness	7.32	8.03	8.71	7.31	8.17	8.74	3.90	4.22	4.53
		μ6 .25 skewness	6.59	7.42	8.37	6.60	7.57	8.41	3.50	3.88	4.32
		μ7-.25 skewness	6.94	7.68	8.48	6.93	7.81	8.53	3.67	4.05	4.39
0.8	Y	μ1 (0,0,0)	5.85	6.66	7.52	5.85	6.66	7.52	2.98	3.71	3.75
	X	μ1 (0,0,0)	6.31	6.66	7.74	6.31	6.80	7.79	3.35	3.57	4.03
		μ2 (0.1,0.1,0.1)	6.08	6.57	7.63	6.07	6.67	7.67	3.22	3.51	3.97
		μ3 (0,0.1, −0.1)	6.31	6.50	7.94	6.31	6.64	7.98	3.36	3.50	4.12
		μ4 .75 skewness	5.56	6.19	7.46	5.54	6.30	7.50	2.94	3.31	3.87
		μ5 -.75 skewness	6.80	7.00	8.02	6.83	7.14	8.05	3.62	3.76	4.17
		μ6 .25 skewness	6.13	6.58	7.70	6.15	6.69	7.74	3.26	3.51	4.00
		μ7-.25 skewness	6.43	6.79	7.81	6.44	6.91	7.87	3.43	3.61	4.07
0.6	Y	μ1 (0,0,0)	5.37	5.92	6.83	5.37	5.92	6.83	2.68	3.28	3.35
	X	μ1 (0,0,0)	5.80	5.91	7.11	5.80	6.02	7.15	3.05	3.13	3.64
		μ2 (0.1,0.1,0.1)	5.64	5.76	6.97	5.63	5.89	7.02	2.96	3.06	3.56
		μ3 (0,0.1, −0.1)	5.83	5.75	7.25	5.81	5.85	7.29	3.05	3.04	3.72
		μ4 .75 skewness	5.13	5.50	6.88	5.12	5.62	6.92	2.68	2.91	3.52
		μ5 -.75 skewness	6.29	6.21	7.39	6.31	6.32	7.41	3.31	3.28	3.77
		μ6 .25 skewness	5.67	5.80	7.06	5.65	5.91	7.11	2.97	3.08	3.62
		μ7-.25 skewness	5.94	5.97	7.17	5.95	6.10	7.23	3.12	3.16	3.67
0.4	Y	μ1 (0,0,0)	5.12	5.59	6.52	5.12	5.59	6.52	2.50	3.05	3.13
	X	μ1 (0,0,0)	5.56	5.51	6.80	5.59	5.63	6.82	2.88	2.86	3.41
		μ2 (0.1,0.1,0.1)	5.37	5.38	6.68	5.35	5.51	6.71	2.78	2.80	3.35
		μ3 (0,0.1, −0.1)	5.58	5.38	6.94	5.58	5.49	6.96	2.88	2.79	3.49
		μ4 .75 skewness	4.89	5.15	6.57	4.87	5.26	6.61	2.52	2.68	3.31
		μ5 -.75 skewness	6.10	5.84	7.08	6.10	5.98	7.11	3.15	3.01	3.57
		μ6 .25 skewness	5.40	5.42	6.75	5.42	5.54	6.78	2.80	2.82	3.39
		μ7-.25 skewness	5.71	5.61	6.86	5.72	5.72	6.90	2.95	2.91	3.45

Table 3.11. Mean of Augmented Scores

ρ	Form	Ability Distribution	32 items with 4 common items Subtest1	Subtest2	Subtest3	32 items with 8 common items Subtest1	Subtest2	Subtest3	16 items with 4 common items Subtest1	Subtest2	Subtest3
0.9	Y	µ1 (0,0,0)	25.00	22.85	22.69	25.00	22.85	22.69	12.55	11.26	11.45
	X	µ1 (0,0,0)	25.05	23.29	22.03	25.12	23.14	21.90	12.50	11.43	11.09
		µ2 (0.1,0.1,0.1)	25.56	23.81	22.69	25.60	23.70	22.57	12.78	11.73	11.42
		µ3 (0,0.1, −0.1)	25.08	23.84	21.40	25.13	23.70	21.26	12.51	11.74	10.77
		µ4 .75 skewness	24.88	23.08	21.63	24.97	22.91	21.52	12.42	11.33	10.89
		µ5 -.75 skewness	25.17	23.46	22.37	25.22	23.33	22.21	12.56	11.56	11.26
		µ6 .25 skewness	25.01	23.23	21.92	25.06	23.07	21.80	12.49	11.42	11.04
		µ7-.25 skewness	25.08	23.33	22.13	25.15	23.20	21.99	12.52	11.47	11.14
0.8	Y	µ1 (0,0,0)	25.05	22.86	22.72	25.05	22.86	22.72	12.57	11.26	11.47
	X	µ1 (0,0,0)	25.06	23.32	22.03	25.12	23.19	21.89	12.52	11.46	11.08
		µ2 (0.1,0.1,0.1)	25.52	23.79	22.63	25.59	23.66	22.51	12.76	11.71	11.38
		µ3 (0,0.1, −0.1)	25.05	23.80	21.37	25.10	23.67	21.23	12.50	11.72	10.77
		µ4 .75 skewness	24.91	23.09	21.63	24.96	22.95	21.50	12.43	11.35	10.88
		µ5 -.75 skewness	25.19	23.52	22.35	25.23	23.37	22.22	12.58	11.58	11.26
		µ6 .25 skewness	25.02	23.24	21.91	25.08	23.12	21.78	12.49	11.43	11.03
		µ7-.25 skewness	25.11	23.36	22.12	25.17	23.23	22.00	12.54	11.49	11.12
0.6	Y	µ1 (0,0,0)	25.04	22.89	22.73	25.04	22.89	22.73	12.57	11.27	11.48
	X	µ1 (0,0,0)	25.12	23.32	22.07	25.17	23.18	21.94	12.54	11.47	11.10
		µ2 (0.1,0.1,0.1)	25.53	23.79	22.65	25.60	23.66	22.52	12.76	11.73	11.39
		µ3 (0,0.1, −0.1)	25.07	23.82	21.36	25.14	23.70	21.24	12.52	11.73	10.74
		µ4 .75 skewness	24.94	23.09	21.68	24.98	22.96	21.54	12.45	11.36	10.90
		µ5 -.75 skewness	25.25	23.51	22.40	25.31	23.39	22.27	12.61	11.59	11.27
		µ6 .25 skewness	25.05	23.24	21.95	25.12	23.11	21.82	12.51	11.44	11.05
		µ7-.25 skewness	25.16	23.38	22.16	25.21	23.24	22.03	12.56	11.51	11.16
0.4	Y	µ1 (0,0,0)	25.01	22.86	22.68	25.01	22.86	22.68	12.55	11.26	11.45
	X	µ1 (0,0,0)	25.08	23.28	21.99	25.12	23.15	21.86	12.52	11.45	11.07
		µ2 (0.1,0.1,0.1)	25.58	23.83	22.67	25.65	23.70	22.54	12.78	11.75	11.40
		µ3 (0,0.1, −0.1)	25.05	23.84	21.40	25.11	23.71	21.27	12.52	11.74	10.76
		µ4 .75 skewness	24.91	23.08	21.62	24.97	22.92	21.48	12.41	11.34	10.87
		µ5 -.75 skewness	25.20	23.48	22.32	25.27	23.34	22.20	12.60	11.56	11.24
		µ6 .25 skewness	25.02	23.22	21.90	25.08	23.07	21.77	12.49	11.42	11.02
		µ7-.25 skewness	25.13	23.35	22.08	25.18	23.20	21.96	12.54	11.49	11.12

Table 3.12. Standard Deviation of Augmented Scores

ρ	Form	Ability Distribution	32 items with 4 common items Subtest1	Subtest2	Subtest3	32 items with 8 common items Subtest1	Subtest2	Subtest3	16 items with 4 common items Subtest1	Subtest2	Subtest3
0.9	Y	μ1 (0,0,0)	5.08	5.36	5.80	5.08	5.36	5.80	2.49	2.69	2.74
	X	μ1 (0,0,0)	5.39	5.45	6.09	5.42	5.53	6.13	2.74	2.72	2.96
		μ2 (0.1,0.1,0.1)	5.21	5.32	5.95	5.23	5.38	5.99	2.64	2.64	2.86
		μ3 (0,0.1, −0.1)	5.36	5.39	6.15	5.39	5.48	6.18	2.72	2.70	2.97
		μ4 .75 skewness	4.84	5.06	5.78	4.84	5.12	5.82	2.46	2.52	2.79
		μ5 -.75 skewness	5.89	5.84	6.49	5.91	5.92	6.54	2.99	2.92	3.15
		μ6 .25 skewness	5.26	5.36	6.02	5.29	5.44	6.05	2.67	2.66	2.91
		μ7-.25 skewness	5.54	5.56	6.19	5.56	5.63	6.23	2.80	2.78	3.00
0.8	Y	μ1 (0,0,0)	4.90	5.22	5.84	4.90	5.22	5.84	2.38	2.64	2.69
	X	μ1 (0,0,0)	5.33	5.26	6.15	5.35	5.34	6.18	2.68	2.61	2.94
		μ2 (0.1,0.1,0.1)	5.16	5.15	6.04	5.16	5.23	6.08	2.60	2.56	2.89
		μ3 (0,0.1, −0.1)	5.34	5.21	6.25	5.35	5.30	6.29	2.70	2.61	2.99
		μ4 .75 skewness	4.72	4.87	5.87	4.72	4.95	5.90	2.38	2.42	2.80
		μ5 -.75 skewness	5.84	5.62	6.55	5.89	5.73	6.58	2.95	2.81	3.14
		μ6 .25 skewness	5.18	5.17	6.07	5.21	5.24	6.12	2.61	2.56	2.91
		μ7-.25 skewness	5.46	5.37	6.24	5.48	5.45	6.28	2.76	2.66	3.00
0.6	Y	μ1 (0,0,0)	4.80	5.12	5.88	4.80	5.12	5.88	2.27	2.60	2.68
	X	μ1 (0,0,0)	5.23	5.11	6.20	5.24	5.21	6.23	2.62	2.51	2.95
		μ2 (0.1,0.1,0.1)	5.10	4.98	6.08	5.10	5.09	6.12	2.55	2.45	2.89
		μ3 (0,0.1, −0.1)	5.27	5.03	6.30	5.26	5.11	6.34	2.63	2.47	2.99
		μ4 .75 skewness	4.59	4.72	5.94	4.59	4.82	5.98	2.28	2.32	2.82
		μ5 -.75 skewness	5.78	5.48	6.57	5.80	5.59	6.60	2.90	2.70	3.15
		μ6 .25 skewness	5.10	5.00	6.13	5.09	5.10	6.18	2.54	2.46	2.92
		μ7-.25 skewness	5.38	5.19	6.28	5.40	5.31	6.33	2.69	2.55	2.99
0.4	Y	μ1 (0,0,0)	4.75	5.10	5.92	4.75	5.10	5.92	2.21	2.59	2.67
	X	μ1 (0,0,0)	5.20	5.03	6.22	5.24	5.14	6.25	2.58	2.45	2.95
		μ2 (0.1,0.1,0.1)	5.03	4.90	6.09	5.02	5.02	6.13	2.50	2.39	2.89
		μ3 (0,0.1, −0.1)	5.22	4.94	6.34	5.22	5.04	6.36	2.58	2.41	3.00
		μ4 .75 skewness	4.53	4.65	5.95	4.52	4.76	6.00	2.21	2.27	2.82
		μ5 -.75 skewness	5.81	5.41	6.59	5.81	5.54	6.62	2.91	2.64	3.16
		μ6 .25 skewness	5.03	4.93	6.16	5.06	5.04	6.19	2.50	2.40	2.91
		μ7-.25 skewness	5.37	5.13	6.30	5.39	5.24	6.34	2.67	2.50	2.99

3.4.2 PRMSE and MDS

3.4.2.1 PRMSE

Table 3.13 presents PRMSE values across all conditions. PRMSE values are used to examine whether subtest scores or weighted averages provide added value. If the values of $PRMSE_s$ are greater than those of $PRMSE_x$, subtest scores are considered to provide additional information over the total score. As can be seen in Table 3.13, regardless of the proficiency distributions (μ1=(0,0,0), μ2=(0.1,0.1,0.1), μ3=(0,0.1,-0.1), μ4 with skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25), the values of $PRMSE_x$ in subtest 2 with the correlation of 0.9 and with 16 items and 0.8 correlation were larger than those of $PRMSE_s$ indicating that subtest scores did not confer added value. The same pattern was found in the condition of 16 items and 0.9 correlation in all three subtests. In general, $PRMSE_s$ remained almost unchanged whereas $PRMSE_x$ and $PRMSE_{sx}$ decreased as the correlations became lower. The values of $PRMSE_x$ diminished more significantly from 0.9 to 0.4 because correlations between total score and true subtest score estimates directly influence the true subtest score estimate $\hat{s}_x$. The correlations between total and estimated true subtest scores affect $PRMSE_{sx}$ as well; however, the influence was not as significant as on $PRMSE_x$. With the smaller number of items, the PRMSE values were smaller than those from a larger number of items.

Table 3.13. Proportional Reduction in Mean Squared Error (PRMSE)

Form		ρ	PRMSE	Subtest1 Items & Common items			Subtest2 Items & Common items			Subtest3 Items & Common items		
				32 & 4	32 & 8	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
Y	μ1	0.9	st	0.878	0.878	0.789	0.887	0.887	0.811	0.899	0.899	0.815
			s	0.855	0.855	0.740	0.861	0.861	0.772	0.887	0.887	0.785
			x	0.854	0.854	0.810	0.893	0.893	0.853	0.889	0.889	0.847
		0.8	st	0.866	0.866	0.768	0.874	0.874	0.793	0.893	0.893	0.799
			s	0.852	0.852	0.738	0.860	0.860	0.771	0.887	0.887	0.782
			x	0.778	0.778	0.735	0.821	0.821	0.786	0.826	0.826	0.781
		0.6	st	0.857	0.857	0.747	0.864	0.864	0.776	0.888	0.888	0.788
			s	0.853	0.853	0.737	0.860	0.860	0.771	0.887	0.887	0.784
			x	0.634	0.634	0.590	0.680	0.680	0.652	0.706	0.706	0.651
		0.4	st	0.855	0.855	0.740	0.861	0.861	0.773	0.887	0.887	0.784
			s	0.854	0.854	0.739	0.861	0.861	0.773	0.887	0.887	0.784
			x	0.491	0.491	0.447	0.543	0.543	0.529	0.590	0.590	0.527
X	μ1	0.9	st	0.892	0.892	0.820	0.888	0.890	0.814	0.904	0.904	0.831
			s	0.877	0.877	0.791	0.852	0.857	0.749	0.895	0.895	0.811
			x	0.859	0.854	0.822	0.905	0.904	0.868	0.879	0.878	0.841
		0.8	st	0.885	0.886	0.808	0.871	0.874	0.786	0.899	0.899	0.820
			s	0.877	0.878	0.792	0.852	0.856	0.749	0.895	0.894	0.808
			x	0.785	0.784	0.752	0.828	0.832	0.792	0.817	0.817	0.779
		0.6	st	0.879	0.878	0.795	0.858	0.862	0.762	0.896	0.896	0.811
			s	0.876	0.876	0.791	0.853	0.857	0.750	0.895	0.895	0.809
			x	0.643	0.641	0.611	0.684	0.685	0.650	0.699	0.700	0.657
		0.4	st	0.877	0.878	0.792	0.853	0.857	0.751	0.895	0.894	0.808
			s	0.877	0.877	0.791	0.852	0.856	0.750	0.895	0.894	0.808
			x	0.502	0.505	0.477	0.535	0.542	0.507	0.584	0.582	0.537
X	μ2	0.9	st	0.890	0.890	0.817	0.887	0.889	0.812	0.903	0.903	0.829
			s	0.874	0.874	0.786	0.851	0.856	0.749	0.894	0.894	0.808
			x	0.855	0.854	0.819	0.904	0.903	0.866	0.879	0.880	0.841
		0.8	st	0.883	0.883	0.805	0.871	0.873	0.786	0.899	0.899	0.821
			s	0.874	0.874	0.788	0.852	0.856	0.749	0.895	0.895	0.810
			x	0.783	0.780	0.745	0.830	0.830	0.791	0.819	0.818	0.780
		0.6	st	0.877	0.877	0.793	0.856	0.861	0.760	0.895	0.895	0.810
			s	0.875	0.875	0.788	0.851	0.856	0.749	0.894	0.895	0.809
			x	0.638	0.636	0.607	0.678	0.683	0.644	0.697	0.698	0.651
		0.4	st	0.875	0.874	0.789	0.851	0.856	0.749	0.895	0.894	0.809
			s	0.874	0.874	0.788	0.850	0.855	0.748	0.895	0.894	0.808
			x	0.501	0.499	0.474	0.538	0.544	0.507	0.589	0.587	0.541
X	μ3	0.9	st	0.892	0.892	0.819	0.887	0.890	0.813	0.903	0.903	0.829
			s	0.876	0.877	0.791	0.850	0.855	0.747	0.895	0.895	0.809
			x	0.856	0.854	0.819	0.903	0.905	0.865	0.876	0.876	0.837
		0.8	st	0.886	0.885	0.809	0.871	0.874	0.788	0.899	0.900	0.820
			s	0.878	0.877	0.793	0.852	0.856	0.750	0.895	0.895	0.809
			x	0.784	0.784	0.752	0.826	0.830	0.793	0.820	0.820	0.779
		0.6	st	0.880	0.879	0.795	0.858	0.861	0.760	0.895	0.896	0.811
			s	0.877	0.877	0.791	0.852	0.856	0.749	0.895	0.895	0.809
			x	0.643	0.639	0.610	0.675	0.676	0.641	0.702	0.701	0.657
		0.4	st	0.877	0.877	0.792	0.853	0.857	0.750	0.895	0.895	0.809
			s	0.877	0.877	0.791	0.852	0.856	0.748	0.895	0.895	0.809
			x	0.502	0.502	0.475	0.526	0.532	0.499	0.591	0.590	0.546

Form		ρ	PRMSE	Subtest1 Items & Common items 32 & 4	32 & 8	16 & 4	Subtest2 Items & Common items 32 & 8	32 & 4	16 & 4	Subtest3 Items & Common items 32 & 8	32 & 4	16 & 4
X	μ4	0.9	st	0.865	0.864	0.780	0.869	0.871	0.785	0.892	0.892	0.809
			s	0.840	0.838	0.737	0.830	0.834	0.720	0.883	0.883	0.790
			x	0.840	0.838	0.798	0.890	0.891	0.846	0.870	0.871	0.824
		0.8	st	0.854	0.853	0.762	0.851	0.854	0.758	0.887	0.887	0.801
			s	0.840	0.839	0.736	0.830	0.835	0.720	0.883	0.883	0.791
			x	0.760	0.756	0.721	0.811	0.811	0.770	0.809	0.808	0.762
		0.6	st	0.844	0.843	0.743	0.836	0.841	0.731	0.884	0.884	0.793
			s	0.840	0.839	0.735	0.830	0.835	0.720	0.884	0.884	0.792
			x	0.612	0.609	0.576	0.663	0.667	0.620	0.694	0.692	0.644
		0.4	st	0.841	0.839	0.736	0.831	0.835	0.721	0.883	0.883	0.791
			s	0.840	0.839	0.735	0.830	0.834	0.719	0.883	0.883	0.791
			x	0.469	0.464	0.441	0.522	0.526	0.486	0.583	0.583	0.529
X	μ5	0.9	st	0.910	0.910	0.848	0.901	0.903	0.834	0.916	0.916	0.850
			s	0.900	0.900	0.828	0.873	0.877	0.779	0.908	0.908	0.833
			x	0.861	0.858	0.830	0.907	0.907	0.874	0.878	0.877	0.844
		0.8	st	0.905	0.906	0.839	0.886	0.889	0.809	0.912	0.912	0.840
			s	0.900	0.901	0.828	0.871	0.876	0.778	0.908	0.908	0.831
			x	0.785	0.783	0.754	0.828	0.829	0.795	0.813	0.812	0.778
		0.6	st	0.901	0.901	0.829	0.876	0.879	0.787	0.908	0.908	0.833
			s	0.900	0.900	0.827	0.872	0.876	0.778	0.908	0.908	0.831
			x	0.643	0.641	0.618	0.677	0.678	0.647	0.691	0.688	0.651
		0.4	st	0.901	0.901	0.829	0.872	0.876	0.778	0.908	0.907	0.832
			s	0.901	0.901	0.828	0.871	0.876	0.777	0.908	0.907	0.832
			x	0.507	0.507	0.485	0.529	0.534	0.502	0.571	0.569	0.531
X	μ6	0.9	st	0.887	0.887	0.812	0.884	0.886	0.807	0.901	0.902	0.826
			s	0.869	0.869	0.780	0.847	0.852	0.742	0.892	0.892	0.805
			x	0.856	0.854	0.820	0.902	0.903	0.862	0.878	0.879	0.840
		0.8	st	0.878	0.879	0.798	0.867	0.869	0.780	0.896	0.897	0.815
			s	0.869	0.870	0.779	0.848	0.850	0.742	0.892	0.892	0.804
			x	0.781	0.781	0.745	0.826	0.829	0.788	0.818	0.818	0.778
		0.6	st	0.872	0.871	0.784	0.852	0.856	0.754	0.893	0.893	0.807
			s	0.869	0.868	0.778	0.846	0.851	0.742	0.892	0.892	0.805
			x	0.642	0.639	0.609	0.682	0.682	0.647	0.701	0.700	0.655
		0.4	st	0.868	0.869	0.780	0.847	0.851	0.744	0.892	0.892	0.804
			s	0.868	0.869	0.779	0.846	0.851	0.742	0.892	0.892	0.804
			x	0.500	0.498	0.470	0.537	0.541	0.503	0.586	0.586	0.538
X	μ7	0.9	st	0.898	0.898	0.828	0.892	0.894	0.820	0.907	0.907	0.836
			s	0.885	0.884	0.802	0.857	0.862	0.759	0.899	0.899	0.815
			x	0.860	0.857	0.823	0.906	0.906	0.871	0.879	0.879	0.843
		0.8	st	0.891	0.891	0.817	0.876	0.878	0.793	0.902	0.902	0.826
			s	0.884	0.884	0.803	0.858	0.862	0.757	0.898	0.898	0.815
			x	0.786	0.783	0.751	0.831	0.832	0.793	0.817	0.819	0.777
		0.6	st	0.885	0.885	0.804	0.862	0.867	0.767	0.899	0.899	0.816
			s	0.883	0.884	0.801	0.857	0.862	0.756	0.898	0.899	0.814
			x	0.645	0.643	0.616	0.683	0.686	0.649	0.696	0.697	0.657
		0.4	st	0.884	0.884	0.802	0.858	0.862	0.759	0.898	0.898	0.814
			s	0.884	0.884	0.802	0.857	0.862	0.757	0.898	0.898	0.814
			x	0.507	0.505	0.480	0.538	0.540	0.509	0.582	0.580	0.538

* Note: μ1 is μ =(0,0,0); μ2 is μ =(0.1,0.1,0.1); μ3 is μ =(0,0.1,-0.1); μ4 is a skewed distribution with skewness of .75; μ5 with -.75; μ6 with .25, and μ7 with -.25 skewness.

Table 3.14 provides the number of replications out of 100 showing subtest scores with added value ($PRMSE_s$>$PRMSE_x$). In most cases, subtest scores proved to have additional value over the total score except for the conditions in 0.9 correlation and in 0.8 correlation with 16 items. In subtests 1 and 3 of FormX, except for μ4 in subtest 1, more than 90 % of the cases showed that subtest scores provided added value whereas about half of the replications demonstrated that subtest scores had added value in FormY when the correlation was 0.9 and the number of items was 32. However, with 0.9 correlation for all conditions in subtest 2 and 16 items with 0.9 and 0.8 correlations in all three subtests, subtest scores are less likely to have added value. The reason why subtest 2 showed a smaller number of cases of subtest scores having added value was partially due to diminished subtest score reliability resulting in slightly higher $PRMSE_x$ values. When the correlation was lower than 0.6, all cases showed that subtest scores conferred additional value over the total score. This pattern persisted regardless of the proficiency distributions.

Table 3.14. Number of Replications out of 100 for Subtest Scores to have Added Value ($PRMSE_s > PRMSE_x$)

	ρ	s*	Subtest1 Items & Common items			Subtest2 Items & Common items			Subtest3 Items & Common items		
			32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
FormY	0.9	μ1	55	55	0	0	0	0	41	41	0
FormX		μ1	99	99	1	0	0	0	99	99	1
		μ2	91	96	2	0	0	0	97	99	1
		μ3	99	99	3	0	0	0	100	100	1
		μ4	51	47	0	0	0	0	97	93	1
		μ5	100	100	48	0	0	0	100	100	17
		μ6	93	93	0	0	0	0	97	100	0
		μ7	98	100	6	0	0	0	99	100	2
FormY	0.8	μ1	100	100	53	100	100	12	100	100	53
FormX		μ1	100	100	99	99	99	0	100	100	98
		μ2	100	100	100	98	100	0	100	100	100
		μ3	100	100	99	99	98	0	100	100	100
		μ4	100	100	78	92	99	2	100	100	97
		μ5	100	100	100	100	100	15	100	100	100
		μ6	100	100	97	97	99	1	100	100	96
		μ7	100	100	100	100	99	2	100	100	99
FormY	0.6	μ1	100	100	100	100	100	100	100	100	100
FormX		μ1	100	100	100	100	100	100	100	100	100
		μ2	100	100	100	100	100	100	100	100	100
		μ3	100	100	100	100	100	100	100	100	100
		μ4	100	100	100	100	100	100	100	100	100
		μ5	100	100	100	100	100	100	100	100	100
		μ6	100	100	100	100	100	100	100	100	100
		μ7	100	100	100	100	100	100	100	100	100
FormY	0.4	μ1	100	100	100	100	100	100	100	100	100
FormX		μ1	100	100	100	100	100	100	100	100	100
		μ2	100	100	100	100	100	100	100	100	100
		μ3	100	100	100	100	100	100	100	100	100
		μ4	100	100	100	100	100	100	100	100	100
		μ5	100	100	100	100	100	100	100	100	100
		μ6	100	100	100	100	100	100	100	100	100
		μ7	100	100	100	100	100	100	100	100	100

* Note: μ1 is μ =(0,0,0); μ2 is μ =(0.1,0.1,0.1); μ3 is μ =(0,0.1,-0.1); μ4 is a skewed distribution with skewness of .75; μ5 with -.75; μ6 with .25, and μ7 with -.25 skewness.

The number of replications for weighted averages to allow added value was also computed under each condition; that is, $PRMSE_{st} > PRMSE_x$ and $PRMSE_{st} > PRMSE_s$. All conditions had added values for weighted averages even with a high correlation. Thus, tables for weighted averages are not included in this paper.

3.4.2.2 MDS

Table 3.15 shows average R-squared values over 100 replications. As the number of dimensions increased, RSQ values also increased. With high correlations, RSQ values were higher compared to the cases with low correlations. When three dimensions were applied, RSQ values were above .9, indicating over 90% of the total variance that were accounted for by the solution. This pattern was consistent regardless of study conditions.

Table 3.15. Average R-squared (RSQ) Values

μ	Dimension	0.9			0.8			0.6			0.4		
		32&4	32&8	16&4	32&4	32&8	16&4	32&4	32&8	16&4	32&4	32&8	16&4
μ1	1	0.847	0.849	0.831	0.831	0.841	0.839	0.809	0.812	0.813	0.762	0.769	0.746
	2	0.899	0.898	0.887	0.888	0.897	0.891	0.894	0.899	0.900	0.897	0.902	0.908
	3	0.920	0.919	0.915	0.916	0.919	0.921	0.920	0.922	0.928	0.916	0.919	0.928
	4	0.932	0.933	0.933	0.929	0.932	0.938	0.932	0.934	0.943	0.930	0.932	0.942
	5	0.941	0.941	0.946	0.938	0.940	0.949	0.942	0.943	0.953	0.940	0.941	0.952
μ2	1	0.856	0.861	0.861	0.833	0.847	0.856	0.811	0.817	0.828	0.768	0.768	0.749
	2	0.902	0.905	0.908	0.894	0.902	0.906	0.899	0.905	0.913	0.902	0.904	0.911
	3	0.922	0.925	0.928	0.919	0.924	0.929	0.924	0.927	0.934	0.921	0.924	0.930
	4	0.935	0.936	0.943	0.932	0.936	0.944	0.936	0.939	0.949	0.934	0.935	0.944
	5	0.943	0.945	0.952	0.942	0.945	0.955	0.945	0.946	0.957	0.944	0.944	0.954
μ3	1	0.859	0.861	0.863	0.854	0.857	0.841	0.819	0.822	0.809	0.771	0.773	0.759
	2	0.902	0.905	0.903	0.904	0.909	0.901	0.899	0.906	0.904	0.899	0.906	0.906
	3	0.923	0.924	0.926	0.925	0.927	0.926	0.925	0.925	0.926	0.915	0.919	0.925
	4	0.934	0.935	0.941	0.936	0.938	0.941	0.936	0.936	0.941	0.931	0.932	0.940
	5	0.942	0.943	0.950	0.944	0.946	0.951	0.944	0.945	0.951	0.942	0.942	0.951
μ4	1	0.823	0.831	0.834	0.818	0.822	0.829	0.785	0.796	0.789	0.743	0.754	0.743
	2	0.880	0.883	0.888	0.879	0.883	0.886	0.878	0.888	0.895	0.881	0.891	0.894
	3	0.903	0.905	0.912	0.905	0.906	0.914	0.909	0.910	0.918	0.905	0.911	0.915
	4	0.918	0.919	0.929	0.920	0.920	0.929	0.923	0.924	0.936	0.922	0.925	0.931
	5	0.928	0.928	0.940	0.929	0.929	0.941	0.934	0.934	0.946	0.935	0.936	0.943
μ5	1	0.851	0.851	0.847	0.839	0.838	0.837	0.806	0.799	0.779	0.739	0.758	0.685
	2	0.896	0.899	0.902	0.895	0.897	0.899	0.900	0.904	0.901	0.897	0.903	0.899
	3	0.920	0.921	0.926	0.921	0.921	0.926	0.924	0.926	0.927	0.916	0.921	0.925
	4	0.933	0.934	0.940	0.934	0.934	0.941	0.936	0.938	0.943	0.931	0.933	0.940
	5	0.941	0.943	0.951	0.943	0.943	0.952	0.945	0.946	0.954	0.942	0.943	0.952
μ6	1	0.838	0.856	0.838	0.829	0.836	0.825	0.804	0.809	0.814	0.754	0.752	0.739
	2	0.891	0.901	0.892	0.886	0.894	0.889	0.894	0.899	0.899	0.892	0.898	0.903
	3	0.915	0.921	0.918	0.913	0.917	0.917	0.920	0.922	0.924	0.912	0.919	0.924
	4	0.927	0.933	0.935	0.926	0.930	0.934	0.932	0.934	0.940	0.927	0.931	0.938
	5	0.937	0.941	0.947	0.936	0.938	0.945	0.941	0.943	0.950	0.938	0.940	0.948
μ7	1	0.849	0.851	0.850	0.832	0.841	0.841	0.806	0.812	0.813	0.757	0.743	0.747
	2	0.898	0.900	0.901	0.890	0.897	0.898	0.895	0.902	0.906	0.897	0.901	0.907
	3	0.920	0.920	0.922	0.916	0.920	0.925	0.921	0.925	0.931	0.917	0.920	0.927
	4	0.933	0.933	0.939	0.930	0.932	0.941	0.933	0.936	0.945	0.931	0.932	0.942
	5	0.943	0.942	0.951	0.940	0.941	0.952	0.943	0.944	0.955	0.942	0.942	0.952

* Note: μ1 is μ =(0,0,0); μ2 is μ =(0.1,0.1,0.1); μ3 is μ =(0,0.1,-0.1); μ4 is a skewed distribution with skewness of .75; μ5 with -.75; μ6 with .25, and μ7 with -.25 skewness.

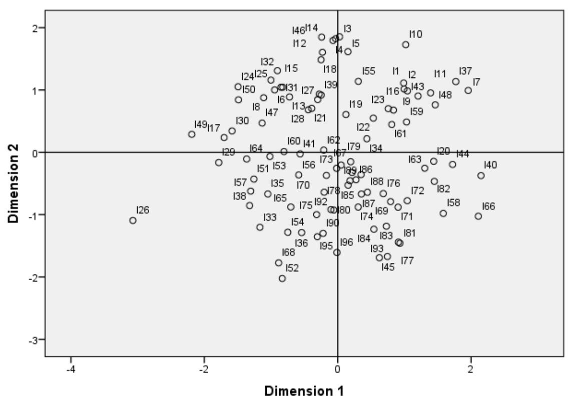
Table 3.16 presents average stress values. Under the one dimension condition, stress values were very large ranging from 0.281 to 0.402. As the correlation among dimensions decreased, stress values obtained from the one dimensional solution increased. On the other hand, when more than two dimensional solutions were applied, stress values were smaller in the case of lower correlations (0.4 and 0.6) compared to higher correlations (0.9 and 0.8). Stress values also dropped as more dimensions were used in the analysis. Considering the ratio of stress value changes as more dimensions were added, it is reasonable to use more than 3 dimensions. Based on RSQ and stress values, using the three dimensional solution provided acceptable fit.

Table 3.16. Average Stress Values

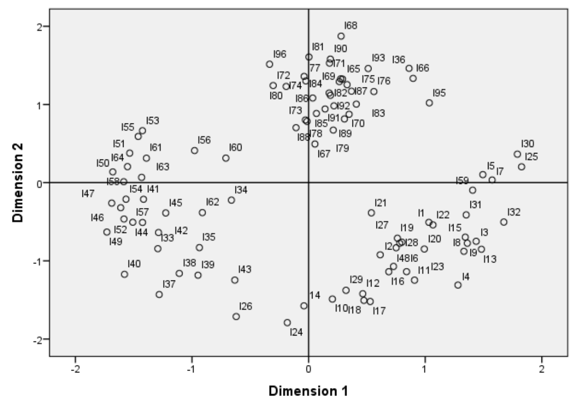
μ	Dimension	0.9			0.8			0.6			0.4		
		32&4	32&8	16&4	32&4	32&8	16&4	32&4	32&8	16&4	32&4	32&8	16&4
μ1	1	0.300	0.299	0.308	0.312	0.304	0.303	0.334	0.327	0.321	0.358	0.352	0.361
	2	0.215	0.216	0.214	0.219	0.212	0.210	0.209	0.201	0.196	0.192	0.189	0.178
	3	0.173	0.174	0.167	0.172	0.170	0.159	0.160	0.157	0.147	0.154	0.153	0.139
	4	0.148	0.148	0.137	0.145	0.144	0.130	0.136	0.134	0.120	0.131	0.131	0.115
	5	0.130	0.130	0.116	0.127	0.126	0.110	0.118	0.117	0.101	0.114	0.114	0.098
μ2	1	0.296	0.292	0.289	0.317	0.304	0.293	0.336	0.326	0.314	0.358	0.352	0.361
	2	0.215	0.214	0.205	0.219	0.211	0.204	0.205	0.199	0.189	0.191	0.188	0.178
	3	0.174	0.172	0.161	0.171	0.168	0.158	0.160	0.158	0.147	0.154	0.151	0.141
	4	0.149	0.147	0.134	0.145	0.142	0.130	0.136	0.134	0.119	0.131	0.130	0.116
	5	0.130	0.129	0.115	0.126	0.125	0.109	0.119	0.118	0.102	0.114	0.114	0.099
μ3	1	0.290	0.286	0.281	0.294	0.290	0.305	0.321	0.315	0.327	0.352	0.350	0.362
	2	0.214	0.210	0.203	0.208	0.203	0.205	0.204	0.193	0.190	0.190	0.185	0.181
	3	0.171	0.169	0.160	0.165	0.164	0.156	0.155	0.154	0.146	0.155	0.153	0.141
	4	0.147	0.145	0.132	0.140	0.140	0.129	0.133	0.132	0.120	0.130	0.130	0.117
	5	0.129	0.128	0.113	0.124	0.123	0.109	0.116	0.115	0.101	0.113	0.113	0.098
μ4	1	0.308	0.305	0.297	0.315	0.311	0.302	0.340	0.330	0.334	0.365	0.355	0.358
	2	0.220	0.219	0.207	0.220	0.217	0.208	0.214	0.206	0.197	0.203	0.195	0.190
	3	0.178	0.178	0.163	0.173	0.174	0.160	0.162	0.162	0.152	0.159	0.156	0.148
	4	0.151	0.152	0.135	0.146	0.147	0.132	0.139	0.138	0.123	0.134	0.132	0.121
	5	0.133	0.133	0.116	0.129	0.129	0.112	0.121	0.120	0.105	0.116	0.116	0.103
μ5	1	0.304	0.304	0.302	0.312	0.312	0.310	0.338	0.337	0.349	0.377	0.358	0.402
	2	0.223	0.220	0.206	0.217	0.214	0.207	0.201	0.196	0.196	0.192	0.185	0.183
	3	0.177	0.176	0.162	0.171	0.169	0.159	0.157	0.156	0.148	0.154	0.150	0.140
	4	0.150	0.150	0.134	0.144	0.143	0.130	0.133	0.132	0.120	0.130	0.128	0.116
	5	0.132	0.131	0.114	0.125	0.125	0.110	0.116	0.116	0.102	0.112	0.111	0.097
μ6	1	0.306	0.293	0.300	0.311	0.308	0.311	0.334	0.328	0.321	0.362	0.359	0.360
	2	0.219	0.214	0.209	0.220	0.214	0.209	0.206	0.201	0.196	0.196	0.190	0.181
	3	0.176	0.174	0.163	0.172	0.170	0.161	0.159	0.158	0.149	0.156	0.152	0.142
	4	0.150	0.147	0.134	0.146	0.145	0.133	0.136	0.134	0.121	0.132	0.130	0.118
	5	0.132	0.130	0.114	0.127	0.127	0.113	0.118	0.118	0.103	0.115	0.113	0.101
μ7	1	0.301	0.299	0.299	0.315	0.306	0.302	0.335	0.327	0.328	0.364	0.366	0.358
	2	0.218	0.217	0.209	0.220	0.214	0.205	0.207	0.198	0.192	0.192	0.187	0.179
	3	0.176	0.175	0.166	0.173	0.169	0.157	0.159	0.157	0.146	0.154	0.151	0.140
	4	0.148	0.148	0.135	0.144	0.144	0.128	0.136	0.134	0.120	0.131	0.129	0.115
	5	0.130	0.130	0.113	0.126	0.126	0.108	0.118	0.117	0.101	0.113	0.112	0.097

* Note: μ1 is μ =(0,0,0); μ2 is μ =(0.1,0.1,0.1); μ3 is μ =(0,0.1,-0.1); μ4 is a skewed distribution with skewness of .75; μ5 with -.75; μ6 with .25, and μ7 with -.25 skewness.

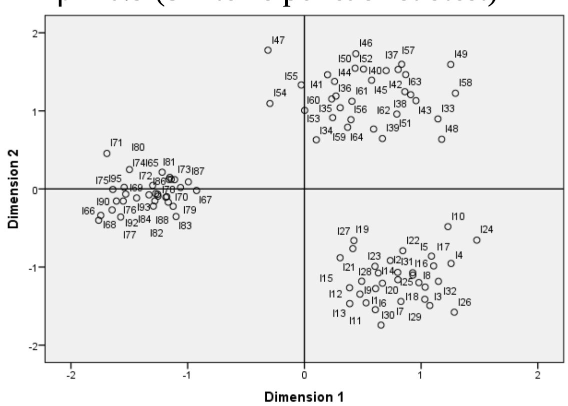
Figure 3.2 and Figure 3.3 show a two-dimensional subspace describing 96 items (32 items for each subtest) and 48 items (16 items for each subtest), respectively, when the correlations among dimensions were 0.9, 0.8, 0.6, and 0.4. Stimulus coordinates are presented in Appendix B. It is clear to see that items corresponding to each subtest (items 1-32 are subtest 1; items 33-64 belong to subtest 2; and items 65-96 are subtest 3 items) congregated together in a two-dimensional subspace as the correlation decreased from 0.9 to 0.4. When the correlation was high, dimensions were hardly distinguishable. On the other hand, when the correlation dropped, the stimulus coordinates showed that the first dimension separated subtest 1 and subtest 3 which had either all positive or all negative stimulus coordinates. In addition, the second dimension separated subtest 1 and subtest 2. This pattern was consistent regardless of the proficiency distributions. Likewise, an MDS analysis makes it possible to evaluate the configuration through a visual inspection although this could be rather subjective.



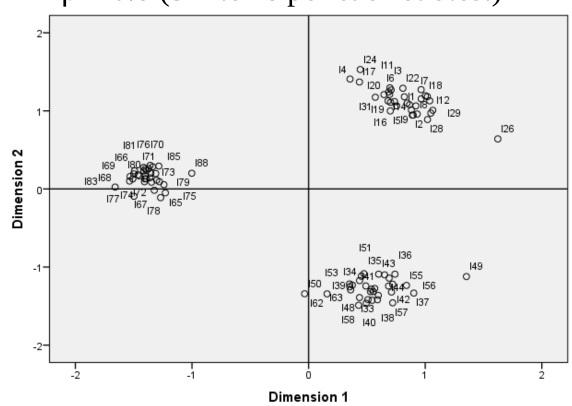
ρ =0.9 (32 items per each subtest)



ρ =0.8 (32 items per each subtest)



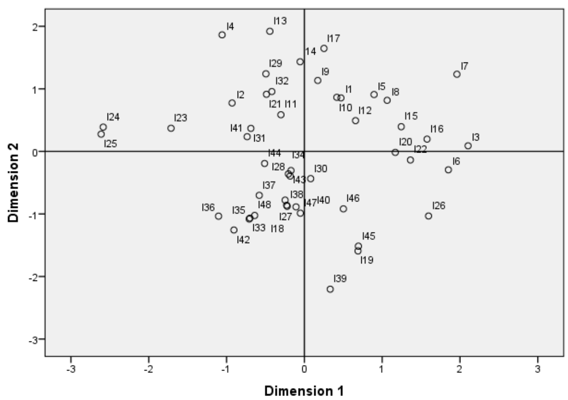
ρ =0.6 (32 items per each subtest)



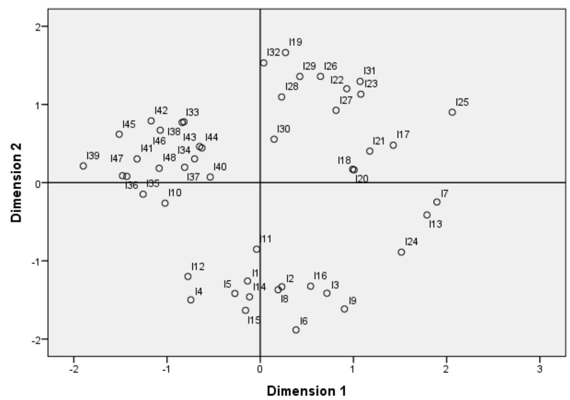
ρ =0.4 (32 items per each subtest)

Figure 3.2. Stimulus Space of 96 items

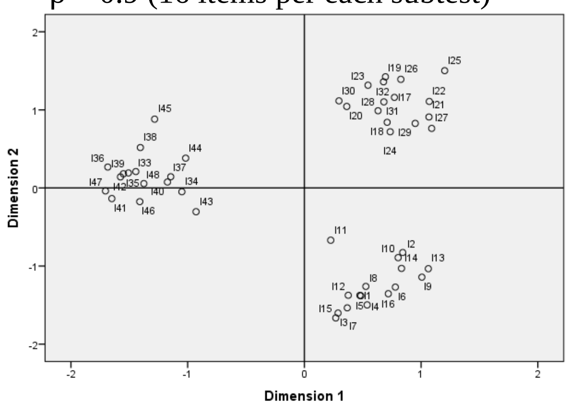
Note: I1-I32 are items in subtest 1; I33-I64 are items in subtest 2; and I65-I96 are items subtest 3.



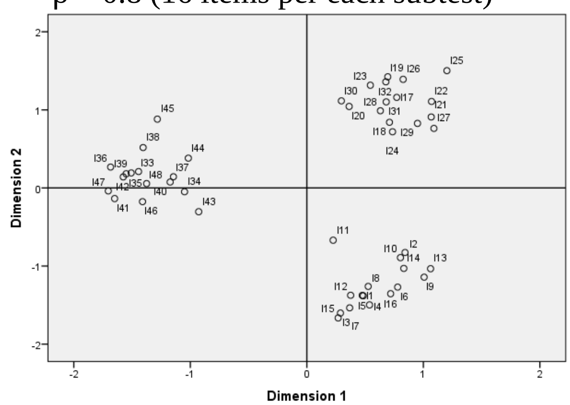
ρ =0.9 (16 items per each subtest)



ρ =0.8 (16 items per each subtest)



ρ =0.6 (16 items per each subtest)



ρ =0.4 (16 items per each subtest)

Figure 3.3. Stimulus Space of 48 items

Note: I1-I16 are items in subtest 1; I17-I32 are items in subtest 2; and I33-I48 are items in subtest 3.

3.4.3 Bias, absolute difference, and RMSE

3.4.3.1 Bias

Bias values were computed based on the true and estimated conversion tables under each condition using three different subtest scores. Table 3.17, Table 3.18, and Table 3.19 present bias from observed scores, weighted averages, and augmented scores, respectively.

When the proficiency distributions of two groups were normal and both mean vectors were equal to $\mu 1 = (0{,}0{,}0)$, bias values were relatively smaller compared to shifted or skewed distributions. In subtest 1, when the number of items was 32, the positively skewed distribution with a skewness of .75 (μ4) produced the largest bias across all correlations. Under this distribution (μ4), using the equated total score yielded relatively less bias while the other two anchor score methods (using the anchor in each subtest or the anchor total score) produced large bias. Under this skewed distribution (μ4), using the anchor score from each subtest showed the largest bias when the correlations were greater than 0.6. The difference among the three anchor approaches became smaller as correlation decreased. Applying one of the anchor score methods produced smaller average bias than the values computed from no equating, except for the positively skewed condition with .75 skewness (μ4).

In subtests 2 and 3, bias values were the largest when the ability distribution of each dimension was shifted to different directions $\mu 3 = (0{,}0.1, -0.1)$ and the equated total or anchor total score was used as the anchor. Under this distribution, no significant impact was found in subtest 1 where the mean of this dimension remained 0 while the mean values of subtests 2 and 3 were shifted. When the distribution was positively skewed (μ4), average bias values were large in subtest 2. Under the distributions with relatively small skewness (.25), either positive (μ6) or negative (μ7), bias values stayed within a comparable range with $\mu 1 = (0,0,0)$.

A similar pattern was observed in the results from weighted average and augmented score approaches (see Table 3.18 and Table 3.19). In subtest 1, compared to observed scores, using weighted averages produced smaller bias when the correlation was low (0.4) and the number of item was 16, except for the case with skewness of .75 (μ4). In subtest 2, when the proficiency distribution was negatively skewed with high skewness (μ5), weighted averages showed smaller bias. On the other hand, when the distribution was shifted to $\mu 3 = (0,0.1,-0.1)$, observed scores had smaller bias than weighted averages. In subtest 3 where the mean score difference between FormX and FormY was the largest, observed scores produced slightly smaller bias in many cases. As can be seen in Table 3.19, when the augmented score approach was applied, all cases showed smaller average bias than using observed scores as equating input. In general, observed scores tend to yield smaller bias than weighted averages, and augmented scores produced smaller bias than observed scores.

Table 3.17. Bias compared with True Equating Function using Observed Scores

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
S1	no	0.341	0.367	0.386	0.341	0.367	0.386	0.341	0.367	0.386	0.341	0.367	0.386
μ1	m1	0.101	0.101	0.007	0.018	0.031	-0.050	0.018	0.028	-0.098	-0.004	-0.027	-0.139
	m2	0.115	0.098	-0.043	0.004	0.002	-0.102	0.045	0.042	-0.118	-0.004	0.003	-0.144
	m3	0.110	0.101	0.003	0.002	0.003	-0.059	0.038	0.038	-0.084	-0.016	-0.005	-0.125
μ2	m1	0.255	0.299	0.120	0.108	0.199	0.044	0.151	0.190	-0.015	0.148	0.235	-0.039
	m2	0.141	0.154	-0.023	0.034	0.061	-0.066	0.013	0.039	-0.129	-0.015	-0.006	-0.167
	m3	0.141	0.163	0.028	0.035	0.064	-0.024	0.006	0.034	-0.087	-0.034	-0.023	-0.144
μ3	m1	0.114	0.091	0.044	-0.005	0.003	-0.067	0.018	0.003	-0.101	-0.014	-0.008	-0.137
	m2	0.134	0.124	-0.014	0.045	-0.001	-0.120	0.033	0.051	-0.128	-0.026	-0.040	-0.151
	m3	0.131	0.124	0.023	0.042	0.000	-0.079	0.030	0.049	-0.094	-0.038	-0.048	-0.129
μ4	m1	0.936	1.253	0.517	0.844	1.167	0.483	0.822	1.162	0.420	0.844	1.215	0.387
	m2	0.390	0.692	0.072	0.344	0.645	0.068	0.520	0.812	0.144	0.780	1.069	0.283
	m3	0.580	0.997	0.291	0.550	0.916	0.255	0.676	1.012	0.292	0.903	1.193	0.369
μ5	m1	-0.021	-0.116	-0.203	-0.147	-0.188	-0.270	-0.141	-0.217	-0.330	-0.151	-0.245	-0.354
	m2	-0.029	-0.092	-0.205	-0.188	-0.226	-0.283	-0.214	-0.278	-0.332	-0.333	-0.406	-0.411
	m3	-0.036	-0.110	-0.172	-0.206	-0.254	-0.252	-0.244	-0.317	-0.318	-0.366	-0.440	-0.414
μ6	m1	0.181	0.213	0.111	0.097	0.167	0.059	0.074	0.107	-0.003	0.094	0.130	-0.021
	m2	0.170	0.175	0.028	0.089	0.135	-0.030	0.140	0.150	-0.063	0.123	0.138	-0.047
	m3	0.178	0.202	0.073	0.094	0.157	0.025	0.149	0.160	-0.009	0.119	0.141	-0.022
μ7	m1	0.046	0.035	-0.035	-0.044	-0.041	-0.143	-0.033	-0.066	-0.157	-0.043	-0.090	-0.209
	m2	0.074	0.051	-0.072	-0.015	-0.043	-0.166	-0.018	-0.048	-0.187	-0.097	-0.109	-0.221
	m3	0.072	0.045	-0.038	-0.021	-0.050	-0.129	-0.031	-0.066	-0.157	-0.115	-0.131	-0.201
S2	no	-0.242	-0.444	-0.212	-0.242	-0.444	-0.212	-0.242	-0.444	-0.212	-0.242	-0.444	-0.212
μ1	m1	-0.058	-0.082	-0.085	-0.029	-0.090	-0.065	-0.028	-0.041	-0.010	-0.018	-0.028	0.013
	m2	-0.059	-0.096	-0.115	0.009	-0.014	-0.078	-0.022	-0.030	-0.028	0.031	-0.004	-0.005
	m3	-0.067	-0.105	-0.093	0.007	-0.016	-0.053	-0.025	-0.032	-0.007	0.028	-0.003	0.005
μ2	m1	0.084	0.065	0.022	0.084	0.104	0.022	0.117	0.177	0.099	0.120	0.193	0.125
	m2	0.009	-0.038	-0.082	0.008	0.034	-0.043	0.016	0.075	0.011	0.002	0.072	0.044
	m3	0.002	-0.046	-0.050	0.008	0.031	-0.018	0.015	0.074	0.041	-0.005	0.069	0.064
μ3	m1	0.040	0.076	-0.016	0.071	0.092	-0.020	0.087	0.176	0.071	0.123	0.194	0.151
	m2	0.491	0.451	0.177	0.523	0.476	0.145	0.539	0.540	0.221	0.570	0.567	0.285
	m3	0.489	0.444	0.163	0.518	0.469	0.143	0.535	0.535	0.233	0.565	0.567	0.297
μ4	m1	0.486	0.648	0.322	0.511	0.652	0.304	0.451	0.583	0.296	0.428	0.637	0.268
	m2	0.093	0.246	-0.038	0.160	0.288	-0.020	0.207	0.321	0.041	0.320	0.504	0.138
	m3	0.172	0.413	0.096	0.240	0.425	0.096	0.260	0.417	0.131	0.363	0.563	0.197
μ5	m1	-0.144	-0.215	-0.237	-0.131	-0.214	-0.234	-0.113	-0.199	-0.178	-0.121	-0.199	-0.138
	m2	-0.100	-0.159	-0.172	-0.066	-0.101	-0.159	-0.118	-0.165	-0.114	-0.158	-0.172	-0.106
	m3	-0.109	-0.174	-0.163	-0.080	-0.121	-0.149	-0.137	-0.185	-0.107	-0.173	-0.192	-0.104
μ6	m1	-0.022	-0.012	0.001	0.019	0.013	0.017	0.047	0.062	0.042	0.032	0.089	0.069
	m2	-0.035	-0.053	-0.062	0.048	0.027	-0.042	0.044	0.052	-0.001	0.077	0.085	0.046
	m3	-0.039	-0.057	-0.041	0.046	0.031	-0.009	0.048	0.056	0.032	0.076	0.090	0.055
μ7	m1	-0.070	-0.139	-0.147	-0.060	-0.127	-0.120	-0.062	-0.077	-0.074	-0.071	-0.077	-0.031
	m2	-0.068	-0.107	-0.124	-0.021	-0.068	-0.106	-0.038	-0.049	-0.045	-0.039	-0.049	-0.033
	m3	-0.068	-0.116	-0.112	-0.023	-0.069	-0.086	-0.043	-0.056	-0.028	-0.047	-0.059	-0.021
S3	no	0.864	0.738	0.485	0.864	0.738	0.485	0.864	0.738	0.485	0.864	0.738	0.485
μ1	m1	0.017	-0.015	-0.014	0.009	0.022	-0.038	0.017	0.022	-0.051	0.015	0.018	-0.069
	m2	0.001	-0.003	-0.030	-0.021	-0.016	-0.075	-0.006	0.001	-0.047	0.009	-0.024	-0.068
	m3	0.002	-0.008	-0.013	-0.019	-0.016	-0.061	-0.005	-0.001	-0.035	0.010	-0.019	-0.061
μ2	m1	0.182	0.254	0.123	0.146	0.242	0.076	0.143	0.251	0.089	0.165	0.249	0.062
	m2	0.115	0.129	0.017	0.038	0.101	-0.021	0.025	0.115	-0.012	-0.035	0.046	-0.049

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
	m3	0.116	0.126	0.048	0.041	0.097	-0.003	0.027	0.112	0.009	-0.037	0.052	-0.031
μ3	m1	-0.112	-0.205	-0.109	-0.117	-0.232	-0.161	-0.137	-0.245	-0.172	-0.145	-0.203	-0.160
	m2	-0.604	-0.633	-0.301	-0.609	-0.634	-0.342	-0.634	-0.624	-0.359	-0.609	-0.604	-0.339
	m3	-0.593	-0.632	-0.297	-0.606	-0.636	-0.333	-0.634	-0.624	-0.354	-0.609	-0.603	-0.338
μ4	m1	0.064	0.068	0.245	0.029	-0.019	0.179	0.015	0.005	0.132	0.032	0.017	0.109
	m2	-0.069	-0.059	-0.015	-0.080	-0.096	-0.054	-0.009	-0.024	-0.048	0.025	0.028	0.015
	m3	-0.060	-0.029	0.069	-0.067	-0.069	0.002	-0.002	-0.003	0.002	0.031	0.043	0.051
μ5	m1	0.039	0.064	-0.082	0.031	0.041	-0.113	0.054	0.053	-0.115	0.037	0.078	-0.121
	m2	0.053	0.090	-0.040	0.014	0.021	-0.088	0.000	0.018	-0.106	-0.021	0.011	-0.137
	m3	0.052	0.085	-0.028	0.014	0.015	-0.081	0.001	0.019	-0.104	-0.011	0.019	-0.138
μ6	m1	0.018	-0.010	0.045	0.002	-0.003	0.003	0.011	0.002	-0.017	-0.006	-0.044	-0.030
	m2	-0.001	-0.025	-0.022	-0.041	-0.030	-0.057	-0.010	-0.002	-0.060	0.009	-0.012	-0.043
	m3	0.002	-0.029	-0.005	-0.039	-0.030	-0.033	-0.006	-0.002	-0.036	0.011	-0.008	-0.034
μ7	m1	0.022	0.043	-0.029	0.034	0.023	-0.064	0.045	0.048	-0.065	0.034	0.014	-0.071
	m2	0.008	0.039	-0.011	0.003	0.011	-0.083	0.001	0.030	-0.064	-0.006	-0.015	-0.076
	m3	0.013	0.034	0.004	0.007	0.013	-0.070	0.003	0.026	-0.054	-0.006	-0.016	-0.067

* Note: S1, S2, and S3 indicate subtests 1, 2, and 3, respectively. 'no' refers to no equating. $\mu1 = (0,0,0)$, $\mu2 = (0.1,0.1,0.1)$, $\mu3 = (0,0.1,-0.1)$, μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. m1 is the first method of using subtest anchor score as the anchor; m2 is the second method of using equated total score; and m3 is the third method of using anchor total score.

Table 3.18. Bias compared with True Equating Function using Weighted Averages

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
S1	no	0.341	0.367	0.386	0.341	0.367	0.386	0.341	0.367	0.386	0.341	0.367	0.386
μ1	m1	0.164	0.089	-0.055	0.082	0.063	-0.095	0.069	0.062	-0.097	0.037	0.004	-0.122
	m2	0.202	0.105	-0.068	0.084	0.050	-0.114	0.103	0.080	-0.102	0.037	0.037	-0.124
	m3	0.195	0.088	-0.050	0.077	0.038	-0.093	0.096	0.073	-0.073	0.026	0.029	-0.100
μ2	m1	0.228	0.271	0.033	0.125	0.205	0.009	0.183	0.222	-0.019	0.167	0.256	-0.026
	m2	0.099	0.085	-0.101	0.047	0.054	-0.092	0.037	0.060	-0.132	-0.016	-0.002	-0.170
	m3	0.095	0.078	-0.073	0.045	0.048	-0.066	0.031	0.054	-0.098	-0.031	-0.014	-0.141
μ3	m1	0.158	0.084	-0.020	0.051	0.041	-0.100	0.069	0.044	-0.098	0.022	0.025	-0.112
	m2	0.211	0.133	-0.035	0.132	0.057	-0.122	0.101	0.096	-0.107	0.009	-0.005	-0.125
	m3	0.209	0.113	-0.027	0.123	0.046	-0.099	0.095	0.088	-0.078	-0.001	-0.012	-0.098
μ4	m1	0.814	0.927	0.657	0.819	1.034	0.652	0.821	1.141	0.563	0.821	1.156	0.493
	m2	0.490	0.586	0.302	0.494	0.642	0.299	0.621	0.832	0.312	0.801	1.044	0.380
	m3	0.692	0.848	0.475	0.661	0.875	0.451	0.745	1.019	0.446	0.908	1.162	0.476
μ5	m1	0.158	0.077	-0.199	0.025	-0.011	-0.221	-0.025	-0.087	-0.272	-0.082	-0.175	-0.312
	m2	0.108	0.044	-0.188	-0.056	-0.098	-0.228	-0.119	-0.178	-0.275	-0.276	-0.340	-0.375
	m3	0.074	-0.008	-0.207	-0.076	-0.134	-0.235	-0.142	-0.209	-0.280	-0.305	-0.370	-0.385
μ6	m1	0.185	0.130	0.016	0.124	0.145	0.006	0.109	0.107	-0.012	0.129	0.148	0.012
	m2	0.217	0.155	-0.015	0.148	0.155	-0.050	0.185	0.165	-0.049	0.158	0.160	-0.017
	m3	0.229	0.167	0.015	0.152	0.167	-0.005	0.194	0.170	0.005	0.158	0.166	0.021
μ7	m1	0.116	0.084	-0.069	0.055	0.049	-0.153	0.031	0.005	-0.143	0.002	-0.052	-0.194
	m2	0.159	0.107	-0.077	0.083	0.040	-0.146	0.049	0.018	-0.164	-0.048	-0.071	-0.203
	m3	0.144	0.074	-0.080	0.070	0.019	-0.139	0.035	-0.002	-0.148	-0.065	-0.093	-0.182
S2	no	-0.242	-0.444	-0.212	-0.242	-0.444	-0.212	-0.242	-0.444	-0.212	-0.242	-0.444	-0.212
μ1	m1	-0.041	-0.073	0.061	-0.010	-0.067	0.069	-0.001	-0.025	0.066	0.000	-0.016	0.087
	m2	-0.016	-0.077	0.045	0.044	0.025	0.080	0.000	-0.013	0.065	0.048	0.009	0.087
	m3	-0.021	-0.091	0.053	0.043	0.021	0.082	-0.002	-0.016	0.070	0.045	0.009	0.089
μ2	m1	-0.068	-0.093	0.055	0.011	0.030	0.107	0.084	0.159	0.142	0.095	0.166	0.170
	m2	-0.146	-0.190	-0.055	-0.092	-0.055	0.042	-0.039	0.037	0.073	-0.048	0.029	0.089
	m3	-0.161	-0.213	-0.037	-0.094	-0.063	0.050	-0.039	0.035	0.082	-0.056	0.027	0.101
μ3	m1	-0.150	-0.074	0.115	-0.002	0.053	0.131	0.078	0.197	0.166	0.121	0.195	0.225
	m2	0.603	0.552	0.436	0.645	0.594	0.395	0.628	0.630	0.415	0.621	0.619	0.428
	m3	0.603	0.528	0.440	0.640	0.585	0.403	0.626	0.628	0.421	0.615	0.618	0.432
μ4	m1	0.034	0.058	0.354	0.236	0.307	0.443	0.311	0.440	0.468	0.329	0.543	0.427
	m2	0.005	-0.010	0.184	0.155	0.145	0.264	0.219	0.263	0.273	0.296	0.447	0.315
	m3	0.075	0.095	0.268	0.214	0.244	0.327	0.248	0.333	0.334	0.318	0.496	0.364
μ5	m1	-0.020	-0.021	-0.004	0.005	-0.035	-0.001	-0.015	-0.071	-0.031	-0.051	-0.116	-0.033
	m2	-0.033	-0.029	0.023	0.023	0.024	0.039	-0.049	-0.077	0.021	-0.100	-0.105	-0.002
	m3	-0.047	-0.047	0.018	0.023	0.016	0.039	-0.056	-0.086	0.017	-0.107	-0.119	-0.013
μ6	m1	-0.065	-0.102	0.104	0.006	-0.031	0.127	0.047	0.038	0.101	0.033	0.069	0.150
	m2	-0.017	-0.072	0.092	0.069	0.033	0.115	0.052	0.044	0.099	0.079	0.076	0.146
	m3	-0.013	-0.075	0.105	0.070	0.033	0.131	0.056	0.045	0.114	0.077	0.080	0.153
μ7	m1	-0.012	-0.083	0.026	-0.002	-0.046	0.068	-0.015	-0.018	0.023	-0.033	-0.041	0.047
	m2	-0.023	-0.047	0.045	0.020	0.004	0.081	-0.003	-0.007	0.065	-0.006	-0.020	0.054
	m3	-0.028	-0.066	0.042	0.021	0.000	0.081	-0.005	-0.013	0.067	-0.012	-0.028	0.060
S3	no	0.864	0.738	0.485	0.864	0.738	0.485	0.864	0.738	0.485	0.864	0.738	0.485
μ1	m1	0.078	0.016	-0.019	0.047	0.064	-0.053	0.024	0.039	-0.063	0.015	0.022	-0.064
	m2	0.055	0.024	-0.007	-0.003	0.004	-0.079	-0.007	0.010	-0.057	0.007	-0.024	-0.058
	m3	0.068	0.032	-0.009	0.006	0.012	-0.082	-0.003	0.011	-0.054	0.009	-0.018	-0.060
μ2	m1	0.008	0.123	0.000	0.059	0.205	0.002	0.093	0.229	0.044	0.118	0.223	0.032
	m2	-0.047	-0.014	-0.095	-0.074	0.013	-0.100	-0.043	0.072	-0.066	-0.105	-0.009	-0.095

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
	m3	-0.051	-0.028	-0.089	-0.072	0.008	-0.100	-0.040	0.068	-0.059	-0.108	-0.001	-0.084
μ3	m1	0.176	-0.006	-0.004	0.053	-0.104	-0.125	-0.061	-0.188	-0.136	-0.119	-0.181	-0.143
	m2	-0.636	-0.706	-0.343	-0.650	-0.685	-0.398	-0.675	-0.664	-0.396	-0.648	-0.639	-0.363
	m3	-0.618	-0.713	-0.355	-0.646	-0.688	-0.403	-0.673	-0.663	-0.395	-0.649	-0.639	-0.370
μ4	m1	-0.277	-0.475	-0.035	-0.142	-0.327	0.005	-0.064	-0.139	0.040	-0.025	-0.076	0.061
	m2	-0.102	-0.241	-0.096	-0.093	-0.226	-0.112	-0.023	-0.088	-0.086	-0.011	-0.040	-0.016
	m3	-0.084	-0.257	-0.033	-0.068	-0.219	-0.066	-0.017	-0.080	-0.043	-0.004	-0.028	0.017
μ5	m1	0.230	0.349	0.053	0.149	0.241	-0.028	0.125	0.159	-0.048	0.065	0.136	-0.066
	m2	0.130	0.228	0.049	0.067	0.113	-0.031	0.041	0.083	-0.066	-0.002	0.049	-0.095
	m3	0.158	0.273	0.046	0.090	0.140	-0.035	0.055	0.102	-0.064	0.016	0.066	-0.100
μ6	m1	0.009	-0.061	-0.016	-0.013	-0.029	-0.058	0.007	-0.016	-0.045	-0.020	-0.062	-0.046
	m2	0.037	-0.026	-0.027	-0.034	-0.039	-0.086	-0.007	-0.011	-0.073	0.000	-0.021	-0.044
	m3	0.052	-0.023	-0.028	-0.027	-0.040	-0.075	-0.002	-0.011	-0.064	0.003	-0.017	-0.041
μ7	m1	0.115	0.171	0.017	0.090	0.108	-0.036	0.073	0.092	-0.051	0.042	0.031	-0.055
	m2	0.053	0.105	0.036	0.024	0.056	-0.054	0.013	0.056	-0.059	-0.003	-0.006	-0.060
	m3	0.072	0.118	0.026	0.040	0.071	-0.061	0.021	0.058	-0.059	-0.001	-0.004	-0.055

* Note: S1, S2, and S3 indicate subtests 1, 2, and 3, respectively. 'no' refers to no equating. $\mu 1 = (0,0,0)$, $\mu 2 = (0.1,0.1,0.1)$, $\mu 3 = (0,0.1,-0.1)$, μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. m1 is the first method of using subtest anchor score as the anchor; m2 is the second method of using equated total score; and m3 is the third method of using anchor total score.

Table 3.19. Bias compared with True Equating Function using Augmented Scores

		ρ = 0.9 Items & Common items			ρ = 0.8 Items & Common items			ρ = 0.6 Items & Common items			ρ = 0.4 Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
S1	no	0.341	0.367	0.386	0.341	0.367	0.386	0.341	0.367	0.386	0.341	0.367	0.386
μ1	m1	0.085	0.050	0.028	0.060	0.049	-0.005	0.100	0.093	0.006	0.116	0.068	0.011
	m2	0.111	0.068	0.012	0.049	0.030	-0.013	0.129	0.110	0.025	0.116	0.102	0.030
	m3	0.098	0.055	0.036	0.039	0.018	0.001	0.119	0.100	0.034	0.105	0.095	0.033
μ2	m1	0.200	0.234	0.110	0.105	0.187	0.080	0.194	0.235	0.061	0.206	0.297	0.085
	m2	0.087	0.087	-0.019	0.026	0.043	-0.008	0.039	0.067	-0.031	0.001	0.015	-0.061
	m3	0.077	0.079	0.009	0.020	0.034	0.008	0.029	0.056	-0.018	-0.014	0.006	-0.053
μ3	m1	0.078	0.021	0.051	0.033	0.017	-0.012	0.096	0.073	0.001	0.091	0.087	0.014
	m2	0.115	0.080	0.027	0.097	0.030	-0.022	0.126	0.128	0.020	0.073	0.056	0.024
	m3	0.106	0.064	0.040	0.087	0.019	-0.006	0.116	0.116	0.032	0.062	0.049	0.026
μ4	m1	0.617	0.882	0.509	0.573	0.854	0.506	0.634	0.941	0.475	0.636	0.954	0.490
	m2	0.304	0.486	0.153	0.311	0.501	0.211	0.545	0.723	0.318	0.743	0.933	0.461
	m3	0.391	0.670	0.275	0.403	0.669	0.322	0.620	0.865	0.417	0.797	1.019	0.532
μ5	m1	0.047	-0.080	-0.092	-0.019	-0.082	-0.115	0.037	-0.036	-0.113	0.049	-0.041	-0.095
	m2	0.008	-0.069	-0.075	-0.110	-0.160	-0.120	-0.093	-0.155	-0.126	-0.184	-0.239	-0.181
	m3	-0.003	-0.092	-0.072	-0.129	-0.186	-0.122	-0.116	-0.180	-0.138	-0.205	-0.254	-0.198
μ6	m1	0.120	0.099	0.088	0.088	0.126	0.068	0.120	0.125	0.062	0.172	0.176	0.095
	m2	0.140	0.110	0.051	0.108	0.132	0.030	0.207	0.192	0.067	0.211	0.199	0.106
	m3	0.138	0.110	0.076	0.102	0.136	0.059	0.209	0.188	0.096	0.207	0.199	0.119
μ7	m1	0.058	0.019	0.022	0.030	0.017	-0.057	0.079	0.047	-0.013	0.093	0.042	-0.033
	m2	0.091	0.049	0.009	0.048	0.008	-0.045	0.083	0.051	-0.023	0.036	0.012	-0.034
	m3	0.080	0.029	0.019	0.035	-0.008	-0.038	0.065	0.031	-0.021	0.019	-0.005	-0.029
S2	no	-0.242	-0.444	-0.212	-0.242	-0.444	-0.212	-0.242	-0.444	-0.212	-0.242	-0.444	-0.212
μ1	m1	0.008	0.001	0.022	0.019	-0.046	0.007	0.002	-0.025	0.027	0.000	-0.032	0.032
	m2	0.004	-0.018	0.007	0.062	0.037	0.016	0.005	-0.011	0.024	0.048	-0.006	0.034
	m3	-0.001	-0.024	0.016	0.060	0.035	0.022	0.002	-0.016	0.030	0.045	-0.006	0.032
μ2	m1	0.108	0.112	0.081	0.088	0.116	0.061	0.091	0.149	0.083	0.071	0.123	0.084
	m2	0.025	-0.004	-0.012	-0.003	0.036	0.008	-0.030	0.031	0.017	-0.088	-0.026	-0.005
	m3	0.020	-0.006	0.008	-0.003	0.034	0.015	-0.031	0.026	0.025	-0.096	-0.026	0.005
μ3	m1	0.100	0.159	0.106	0.089	0.124	0.069	0.061	0.174	0.078	0.079	0.144	0.113
	m2	0.609	0.578	0.338	0.622	0.572	0.290	0.612	0.607	0.332	0.609	0.599	0.359
	m3	0.609	0.572	0.343	0.618	0.568	0.300	0.607	0.603	0.337	0.603	0.597	0.355
μ4	m1	0.229	0.397	0.249	0.247	0.376	0.243	0.251	0.346	0.257	0.243	0.405	0.269
	m2	0.054	0.151	0.029	0.127	0.172	0.066	0.194	0.214	0.122	0.267	0.367	0.206
	m3	0.060	0.202	0.067	0.140	0.224	0.100	0.204	0.258	0.162	0.274	0.396	0.235
μ5	m1	-0.038	-0.078	-0.044	-0.031	-0.091	-0.077	-0.025	-0.085	-0.062	-0.035	-0.099	-0.056
	m2	-0.019	-0.057	-0.002	0.010	-0.012	-0.026	-0.062	-0.093	-0.024	-0.100	-0.103	-0.048
	m3	-0.022	-0.060	0.010	0.001	-0.021	-0.022	-0.071	-0.101	-0.024	-0.103	-0.108	-0.053
μ6	m1	0.025	0.042	0.062	0.039	0.014	0.053	0.049	0.041	0.046	0.023	0.034	0.073
	m2	0.019	0.012	0.037	0.085	0.055	0.033	0.059	0.050	0.049	0.073	0.050	0.085
	m3	0.017	0.008	0.045	0.083	0.054	0.047	0.060	0.047	0.062	0.070	0.052	0.086
μ7	m1	0.010	-0.037	-0.012	0.005	-0.051	-0.009	-0.011	-0.021	-0.015	-0.030	-0.043	0.008
	m2	0.004	-0.018	0.011	0.034	0.001	0.008	0.002	-0.008	0.019	-0.005	-0.028	0.009
	m3	0.006	-0.021	0.020	0.033	0.001	0.013	-0.003	-0.016	0.023	-0.011	-0.034	0.014
S3	no	0.864	0.738	0.485	0.864	0.738	0.485	0.864	0.738	0.485	0.864	0.738	0.485
μ1	m1	0.002	-0.031	-0.028	0.001	0.022	-0.041	0.005	0.019	-0.041	0.006	0.017	-0.045
	m2	-0.009	-0.017	-0.031	-0.036	-0.024	-0.073	-0.023	-0.007	-0.034	-0.001	-0.029	-0.036
	m3	-0.008	-0.019	-0.023	-0.032	-0.021	-0.069	-0.020	-0.008	-0.032	0.002	-0.022	-0.043
μ2	m1	0.100	0.192	0.058	0.086	0.209	0.040	0.084	0.217	0.060	0.103	0.212	0.037
	m2	0.031	0.054	-0.044	-0.038	0.040	-0.063	-0.051	0.061	-0.050	-0.131	-0.031	-0.096

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
	m3	0.033	0.053	-0.026	-0.033	0.038	-0.056	-0.048	0.058	-0.044	-0.132	-0.021	-0.089
μ3	m1	-0.082	-0.199	-0.086	-0.079	-0.209	-0.133	-0.096	-0.218	-0.109	-0.112	-0.173	-0.104
	m2	-0.670	-0.709	-0.348	-0.668	-0.693	-0.376	-0.689	-0.678	-0.373	-0.662	-0.650	-0.343
	m3	-0.663	-0.716	-0.350	-0.666	-0.697	-0.374	-0.689	-0.678	-0.373	-0.662	-0.649	-0.355
μ4	m1	-0.047	-0.109	0.034	-0.055	-0.165	0.025	-0.057	-0.119	0.031	-0.037	-0.102	0.043
	m2	-0.053	-0.099	-0.078	-0.065	-0.143	-0.099	-0.007	-0.066	-0.068	0.001	-0.041	-0.003
	m3	-0.055	-0.097	-0.054	-0.061	-0.140	-0.078	-0.011	-0.068	-0.040	0.002	-0.039	0.019
μ5	m1	0.060	0.131	-0.020	0.055	0.110	-0.055	0.083	0.111	-0.035	0.062	0.143	-0.042
	m2	0.038	0.103	-0.010	0.004	0.035	-0.061	-0.001	0.037	-0.065	-0.011	0.047	-0.082
	m3	0.048	0.116	0.004	0.018	0.047	-0.054	0.014	0.056	-0.059	0.012	0.071	-0.080
μ6	m1	-0.007	-0.041	-0.001	-0.022	-0.028	-0.031	-0.008	-0.028	-0.023	-0.027	-0.073	-0.033
	m2	-0.005	-0.038	-0.037	-0.051	-0.044	-0.067	-0.019	-0.021	-0.046	-0.002	-0.028	-0.016
	m3	-0.002	-0.042	-0.033	-0.049	-0.047	-0.056	-0.016	-0.023	-0.040	0.000	-0.025	-0.021
μ7	m1	0.013	0.052	-0.014	0.036	0.040	-0.045	0.047	0.064	-0.033	0.037	0.028	-0.033
	m2	-0.012	0.031	-0.001	-0.012	0.011	-0.069	-0.011	0.032	-0.043	-0.009	-0.011	-0.039
	m3	-0.003	0.029	0.006	-0.002	0.018	-0.066	-0.004	0.032	-0.042	-0.006	-0.008	-0.035

* Note: S1, S2, and S3 indicate subtests 1, 2, and 3, respectively. 'no' refers to no equating. $\mu1 = (0,0,0)$, $\mu2 = (0.1,0.1,0.1)$, $\mu3 = (0,0.1,-0.1)$, μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. m1 is the first method of using subtest anchor score as the anchor; m2 is the second method of using equated total score; and m3 is the third method of using anchor total score.

3.4.3.2 Absolute difference

Table 3.20, Table 3.21, and Table 3.22 present the average absolute difference between true and estimated equating results. Compared to no equating, in general, absolute average differences were smaller when using one of the three anchor score methods. With the high correlation (0.9), using either the equated total or the anchor total score approach showed the smallest absolute difference, except for the case with a shifted distribution, $\mu3 = (0,0.1,-0.1)$. When the correlation was dropped to 0.4, using the anchor score in each subtest produced smaller values, except for the condition of a positively skewed distribution (μ4). Decreasing the correlation made it more likely to have large difference between using the anchor score in each subtest and using either the equated total score or the anchor total score across the simulation conditions.

Under the shifted proficiency distributions and skewed distributions, values of absolute difference tend to become slightly larger compared to the case with the normal distribution with no mean shift (μ1). The positively skewed distribution with .75 skewness (μ4) showed the largest absolute difference. Under this distribution with the high correlation (0.9 or 0.8), using the equated total score as the anchor produced relatively smaller absolute difference than the other two methods. These values became larger as the correlation decreased to 0.4. In subtests 2 and 3, using the equated total or anchor total did not perform well when the mean vector was shifted to different directions, $\mu 3 = (0, 0.1, -0.1)$. When the skewness was relatively small (μ6 and μ7), the values were close to those from the normal distribution with a mean vector of 0s.

Instead of using observed scores, having augmented scores as equating input produced smaller absolute differences in many cases (see Table 3.22). As can be seen in Table 3.21, however, weighted averages did not perform better than observed scores, except for a few cases in the positively skewed distribution ($\mu 4$). Overall, a similar pattern was found across the three different approaches (using observed scores, weighted averages, and augmented scores). With high correlations, using either equated total or anchor total score approach showed better results, and having more common items yielded smaller absolute difference. Under the positively skewed distribution (μ4), using the equated total score outperformed the other two methods in all three subtests. In subtests 2 and 3, using the subtest anchor score performed better than the other two methods when the distribution was shifted from $\mu = (0,0,0)$ to $\mu 3 = (0, 0.1, -0.1)$.

Table 3.20. Absolute Difference compared with True Equating Function using Observed Scores

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
S1	no	0.632	0.605	0.482	0.632	0.605	0.482	0.632	0.605	0.482	0.632	0.605	0.482
μ1	m1	0.256	0.324	0.262	0.244	0.324	0.259	0.237	0.324	0.233	0.262	0.359	0.282
	m2	0.273	0.301	0.280	0.273	0.318	0.280	0.308	0.337	0.281	0.368	0.404	0.317
	m3	0.260	0.295	0.254	0.271	0.314	0.248	0.309	0.340	0.264	0.376	0.419	0.310
μ2	m1	0.325	0.405	0.292	0.251	0.369	0.261	0.293	0.364	0.257	0.296	0.449	0.283
	m2	0.294	0.319	0.288	0.280	0.316	0.283	0.295	0.346	0.297	0.388	0.402	0.331
	m3	0.289	0.316	0.256	0.276	0.315	0.258	0.301	0.351	0.276	0.393	0.420	0.317
μ3	m1	0.278	0.330	0.257	0.231	0.313	0.238	0.258	0.334	0.235	0.272	0.320	0.271
	m2	0.291	0.316	0.287	0.269	0.300	0.279	0.306	0.360	0.292	0.373	0.399	0.334
	m3	0.286	0.308	0.264	0.267	0.299	0.249	0.311	0.368	0.275	0.378	0.407	0.311
μ4	m1	1.286	1.686	0.747	1.173	1.569	0.712	1.112	1.504	0.612	1.157	1.577	0.599
	m2	0.604	0.940	0.554	0.555	0.896	0.512	0.686	1.010	0.490	0.953	1.295	0.542
	m3	0.763	1.255	0.581	0.724	1.171	0.540	0.835	1.221	0.529	1.081	1.430	0.586
μ5	m1	0.302	0.470	0.359	0.350	0.492	0.396	0.357	0.532	0.452	0.387	0.565	0.490
	m2	0.240	0.286	0.314	0.322	0.404	0.350	0.431	0.496	0.410	0.557	0.660	0.528
	m3	0.239	0.310	0.285	0.331	0.430	0.331	0.449	0.535	0.414	0.592	0.706	0.544
μ6	m1	0.340	0.427	0.276	0.287	0.388	0.272	0.280	0.371	0.221	0.285	0.399	0.269
	m2	0.316	0.347	0.286	0.279	0.338	0.283	0.325	0.362	0.279	0.397	0.421	0.299
	m3	0.315	0.361	0.270	0.274	0.345	0.268	0.334	0.369	0.271	0.401	0.432	0.295
μ7	m1	0.261	0.317	0.264	0.244	0.322	0.271	0.244	0.352	0.278	0.282	0.390	0.331
	m2	0.261	0.290	0.279	0.274	0.307	0.295	0.308	0.352	0.301	0.420	0.431	0.368
	m3	0.256	0.289	0.254	0.270	0.304	0.262	0.312	0.362	0.284	0.433	0.450	0.362
S2	no	0.248	0.447	0.233	0.248	0.447	0.233	0.248	0.447	0.233	0.248	0.447	0.233
μ1	m1	0.243	0.318	0.243	0.248	0.316	0.218	0.220	0.288	0.174	0.225	0.299	0.182
	m2	0.202	0.267	0.219	0.200	0.263	0.195	0.233	0.284	0.174	0.289	0.309	0.196
	m3	0.205	0.273	0.233	0.206	0.268	0.206	0.234	0.280	0.189	0.287	0.311	0.201
μ2	m1	0.264	0.359	0.240	0.251	0.350	0.219	0.251	0.354	0.214	0.259	0.351	0.218
	m2	0.201	0.270	0.200	0.201	0.281	0.185	0.271	0.315	0.190	0.306	0.340	0.209
	m3	0.202	0.284	0.224	0.199	0.284	0.201	0.266	0.319	0.210	0.302	0.336	0.217
μ3	m1	0.285	0.374	0.254	0.277	0.349	0.216	0.250	0.358	0.205	0.259	0.357	0.238
	m2	0.509	0.498	0.255	0.539	0.524	0.235	0.571	0.594	0.294	0.595	0.607	0.348
	m3	0.515	0.501	0.266	0.540	0.527	0.249	0.573	0.599	0.311	0.593	0.611	0.355
μ4	m1	0.857	1.098	0.557	0.826	1.063	0.505	0.762	0.953	0.470	0.724	1.001	0.447
	m2	0.276	0.473	0.261	0.295	0.528	0.268	0.384	0.552	0.288	0.514	0.744	0.330
	m3	0.365	0.691	0.319	0.376	0.690	0.308	0.444	0.668	0.314	0.561	0.816	0.359
μ5	m1	0.369	0.550	0.407	0.373	0.535	0.381	0.325	0.456	0.329	0.318	0.464	0.271
	m2	0.274	0.402	0.300	0.277	0.384	0.264	0.307	0.399	0.244	0.352	0.427	0.233
	m3	0.262	0.413	0.312	0.278	0.391	0.278	0.310	0.411	0.261	0.362	0.444	0.244
μ6	m1	0.237	0.323	0.243	0.235	0.315	0.222	0.245	0.309	0.207	0.243	0.340	0.216
	m2	0.190	0.256	0.196	0.201	0.246	0.191	0.249	0.295	0.193	0.303	0.337	0.215
	m3	0.186	0.262	0.211	0.198	0.253	0.205	0.252	0.300	0.208	0.304	0.345	0.224
μ7	m1	0.264	0.373	0.278	0.263	0.368	0.250	0.230	0.314	0.204	0.237	0.301	0.177
	m2	0.202	0.280	0.229	0.206	0.280	0.210	0.260	0.289	0.191	0.282	0.300	0.187
	m3	0.208	0.296	0.249	0.214	0.286	0.224	0.262	0.283	0.203	0.289	0.304	0.195
S3	no	0.864	0.738	0.485	0.864	0.738	0.485	0.864	0.738	0.485	0.864	0.738	0.485
μ1	m1	0.187	0.263	0.172	0.185	0.255	0.175	0.182	0.248	0.177	0.179	0.249	0.173
	m2	0.167	0.211	0.166	0.195	0.228	0.169	0.227	0.263	0.189	0.280	0.298	0.199
	m3	0.164	0.215	0.161	0.193	0.224	0.166	0.228	0.264	0.186	0.281	0.299	0.195
μ2	m1	0.261	0.341	0.216	0.229	0.344	0.201	0.231	0.348	0.192	0.226	0.336	0.195

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
	m2	0.209	0.247	0.175	0.187	0.253	0.179	0.242	0.283	0.194	0.266	0.306	0.200
	m3	0.207	0.248	0.167	0.188	0.254	0.171	0.242	0.282	0.189	0.264	0.297	0.199
μ3	m1	0.225	0.326	0.209	0.211	0.334	0.214	0.220	0.316	0.219	0.219	0.289	0.207
	m2	0.606	0.635	0.315	0.612	0.640	0.348	0.644	0.639	0.370	0.625	0.617	0.356
	m3	0.595	0.634	0.311	0.609	0.640	0.343	0.642	0.635	0.367	0.624	0.616	0.358
μ4	m1	0.484	0.671	0.468	0.452	0.630	0.401	0.406	0.573	0.326	0.401	0.573	0.310
	m2	0.193	0.306	0.215	0.230	0.348	0.203	0.277	0.363	0.213	0.349	0.459	0.240
	m3	0.211	0.380	0.251	0.242	0.409	0.219	0.293	0.427	0.222	0.367	0.504	0.248
μ5	m1	0.271	0.429	0.288	0.261	0.423	0.292	0.276	0.404	0.295	0.257	0.394	0.293
	m2	0.220	0.307	0.191	0.245	0.308	0.201	0.314	0.346	0.244	0.331	0.402	0.287
	m3	0.216	0.317	0.201	0.236	0.314	0.217	0.310	0.363	0.261	0.339	0.418	0.309
μ6	m1	0.201	0.281	0.189	0.206	0.289	0.176	0.192	0.266	0.164	0.193	0.269	0.175
	m2	0.167	0.213	0.174	0.184	0.237	0.166	0.241	0.279	0.185	0.271	0.303	0.209
	m3	0.164	0.214	0.165	0.181	0.238	0.166	0.241	0.285	0.180	0.270	0.301	0.202
μ7	m1	0.183	0.270	0.197	0.202	0.281	0.191	0.188	0.273	0.196	0.188	0.263	0.187
	m2	0.160	0.206	0.174	0.182	0.223	0.179	0.244	0.275	0.199	0.263	0.300	0.222
	m3	0.161	0.214	0.168	0.184	0.228	0.178	0.244	0.282	0.195	0.268	0.301	0.223

* Note: S1, S2, and S3 indicate subtests 1, 2, and 3, respectively. 'no' refers to no equating. $\mu1 = (0,0,0)$, $\mu2 = (0.1,0.1,0.1)$, $\mu3 = (0,0.1,-0.1)$, μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. m1 is the first method of using subtest anchor scores as the anchor; m2 is the second method of using equated total score; and m3 is the third method of using anchor total score.

Table 3.21. Absolute Difference compared with True Equating Function using Weighted Averages

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
S1	no	0.632	0.605	0.482	0.632	0.605	0.482	0.632	0.605	0.482	0.632	0.605	0.482
μ1	m1	0.326	0.361	0.286	0.276	0.343	0.277	0.254	0.339	0.243	0.261	0.353	0.278
	m2	0.340	0.326	0.294	0.288	0.341	0.299	0.318	0.341	0.289	0.342	0.393	0.313
	m3	0.329	0.317	0.267	0.282	0.339	0.259	0.318	0.341	0.264	0.348	0.407	0.297
μ2	m1	0.341	0.430	0.283	0.277	0.380	0.270	0.321	0.376	0.260	0.295	0.448	0.270
	m2	0.336	0.338	0.309	0.316	0.344	0.302	0.308	0.346	0.299	0.375	0.387	0.328
	m3	0.331	0.336	0.272	0.308	0.342	0.270	0.308	0.348	0.270	0.378	0.395	0.299
μ3	m1	0.321	0.386	0.269	0.261	0.347	0.273	0.279	0.343	0.246	0.265	0.310	0.262
	m2	0.341	0.345	0.301	0.307	0.321	0.304	0.321	0.359	0.301	0.351	0.384	0.326
	m3	0.335	0.337	0.271	0.301	0.318	0.267	0.323	0.366	0.276	0.355	0.386	0.291
μ4	m1	1.409	1.724	0.928	1.273	1.643	0.888	1.161	1.568	0.724	1.139	1.532	0.646
	m2	0.682	0.939	0.459	0.669	0.967	0.463	0.777	1.065	0.472	0.954	1.266	0.532
	m3	0.903	1.274	0.611	0.841	1.237	0.589	0.908	1.275	0.564	1.063	1.394	0.602
μ5	m1	0.386	0.576	0.454	0.341	0.506	0.421	0.303	0.479	0.425	0.329	0.511	0.451
	m2	0.283	0.297	0.316	0.276	0.359	0.328	0.361	0.411	0.360	0.488	0.588	0.480
	m3	0.271	0.366	0.343	0.290	0.409	0.340	0.375	0.451	0.383	0.521	0.633	0.502
μ6	m1	0.371	0.412	0.237	0.298	0.382	0.249	0.290	0.371	0.206	0.294	0.401	0.264
	m2	0.367	0.363	0.298	0.312	0.360	0.283	0.341	0.363	0.274	0.396	0.410	0.298
	m3	0.369	0.366	0.249	0.306	0.357	0.248	0.344	0.362	0.256	0.399	0.420	0.283
μ7	m1	0.317	0.374	0.310	0.271	0.350	0.307	0.250	0.354	0.280	0.272	0.379	0.320
	m2	0.331	0.322	0.290	0.299	0.325	0.311	0.305	0.342	0.293	0.391	0.402	0.362
	m3	0.320	0.323	0.272	0.292	0.333	0.288	0.307	0.353	0.274	0.400	0.417	0.343
S2	no	0.248	0.447	0.233	0.248	0.447	0.233	0.248	0.447	0.233	0.248	0.447	0.233
μ1	m1	0.349	0.398	0.232	0.290	0.368	0.230	0.250	0.330	0.201	0.243	0.321	0.206
	m2	0.254	0.282	0.185	0.242	0.295	0.215	0.249	0.313	0.212	0.294	0.321	0.224
	m3	0.289	0.341	0.211	0.254	0.315	0.229	0.252	0.314	0.212	0.293	0.323	0.226
μ2	m1	0.324	0.452	0.255	0.281	0.395	0.235	0.259	0.374	0.242	0.257	0.351	0.239
	m2	0.271	0.337	0.193	0.256	0.303	0.190	0.284	0.324	0.208	0.321	0.341	0.240
	m3	0.302	0.393	0.220	0.262	0.326	0.203	0.279	0.331	0.214	0.316	0.338	0.246
μ3	m1	0.335	0.437	0.309	0.285	0.367	0.262	0.250	0.340	0.246	0.263	0.365	0.279
	m2	0.610	0.581	0.440	0.648	0.610	0.404	0.641	0.655	0.421	0.639	0.646	0.442
	m3	0.612	0.582	0.448	0.643	0.605	0.413	0.640	0.656	0.428	0.635	0.648	0.447
μ4	m1	0.779	1.064	0.793	0.733	0.998	0.743	0.679	0.902	0.675	0.641	0.944	0.606
	m2	0.273	0.446	0.371	0.319	0.506	0.408	0.397	0.538	0.406	0.488	0.701	0.447
	m3	0.382	0.684	0.492	0.392	0.670	0.495	0.440	0.642	0.459	0.514	0.766	0.488
μ5	m1	0.484	0.664	0.300	0.336	0.495	0.242	0.284	0.389	0.226	0.279	0.415	0.199
	m2	0.277	0.365	0.202	0.248	0.341	0.183	0.266	0.342	0.201	0.311	0.385	0.195
	m3	0.328	0.486	0.238	0.261	0.370	0.193	0.273	0.359	0.209	0.318	0.403	0.195
μ6	m1	0.343	0.423	0.295	0.282	0.356	0.280	0.267	0.327	0.249	0.251	0.341	0.274
	m2	0.236	0.302	0.211	0.239	0.274	0.226	0.264	0.307	0.227	0.297	0.337	0.270
	m3	0.266	0.363	0.234	0.247	0.291	0.243	0.268	0.310	0.241	0.297	0.345	0.281
μ7	m1	0.360	0.479	0.232	0.284	0.393	0.220	0.246	0.333	0.188	0.244	0.318	0.197
	m2	0.245	0.295	0.188	0.232	0.295	0.201	0.272	0.303	0.210	0.281	0.309	0.205
	m3	0.285	0.369	0.213	0.247	0.316	0.210	0.275	0.301	0.210	0.287	0.313	0.211
S3	no	0.864	0.738	0.485	0.864	0.738	0.485	0.864	0.738	0.485	0.864	0.738	0.485
μ1	m1	0.344	0.412	0.234	0.260	0.367	0.227	0.210	0.290	0.197	0.201	0.274	0.191
	m2	0.237	0.251	0.192	0.240	0.270	0.198	0.236	0.283	0.205	0.294	0.311	0.211
	m3	0.275	0.303	0.211	0.255	0.295	0.208	0.240	0.293	0.203	0.297	0.318	0.208
μ2	m1	0.325	0.424	0.235	0.259	0.384	0.231	0.228	0.364	0.195	0.220	0.338	0.200

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
	m2	0.233	0.272	0.206	0.227	0.261	0.209	0.268	0.300	0.217	0.287	0.323	0.223
	m3	0.264	0.327	0.212	0.242	0.282	0.215	0.270	0.309	0.211	0.290	0.320	0.221
μ3	m1	0.394	0.443	0.232	0.271	0.367	0.237	0.217	0.312	0.217	0.228	0.303	0.207
	m2	0.644	0.714	0.355	0.654	0.688	0.406	0.683	0.675	0.405	0.657	0.651	0.374
	m3	0.633	0.735	0.364	0.651	0.698	0.408	0.682	0.673	0.405	0.659	0.653	0.383
μ4	m1	0.699	0.990	0.478	0.563	0.809	0.403	0.428	0.605	0.318	0.391	0.564	0.292
	m2	0.228	0.395	0.214	0.260	0.405	0.230	0.291	0.380	0.227	0.341	0.444	0.237
	m3	0.315	0.601	0.252	0.309	0.541	0.244	0.320	0.462	0.234	0.364	0.495	0.234
μ5	m1	0.392	0.514	0.332	0.322	0.437	0.272	0.269	0.373	0.255	0.250	0.353	0.247
	m2	0.257	0.336	0.201	0.237	0.279	0.186	0.277	0.298	0.223	0.312	0.359	0.248
	m3	0.311	0.413	0.238	0.281	0.330	0.210	0.288	0.330	0.242	0.323	0.386	0.275
μ6	m1	0.339	0.443	0.214	0.274	0.398	0.215	0.218	0.302	0.193	0.211	0.292	0.187
	m2	0.211	0.265	0.198	0.211	0.275	0.194	0.254	0.302	0.205	0.276	0.315	0.217
	m3	0.244	0.324	0.191	0.222	0.305	0.194	0.258	0.316	0.201	0.277	0.317	0.208
μ7	m1	0.351	0.431	0.244	0.282	0.371	0.226	0.219	0.291	0.212	0.212	0.285	0.197
	m2	0.236	0.280	0.199	0.236	0.277	0.195	0.249	0.278	0.208	0.277	0.309	0.220
	m3	0.278	0.344	0.212	0.262	0.315	0.205	0.251	0.292	0.211	0.285	0.318	0.228

* Note: S1, S2, and S3 indicate subtests 1, 2, and 3, respectively. 'no' refers to no equating. $\mu1 = (0,0,0)$, $\mu2 = (0.1,0.1,0.1)$, $\mu3 = (0,0.1,-0.1)$, μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. m1 is the first method of using subtest anchor score as the anchor; m2 is the second method of using equated total score; and m3 is the third method of using anchor total score.

Table 3.22. Absolute Difference compared with True Equating Function using Augmented Scores

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
S1	no	0.632	0.605	0.482	0.632	0.605	0.482	0.632	0.605	0.482	0.632	0.605	0.482
μ1	m1	0.246	0.300	0.222	0.242	0.311	0.217	0.246	0.320	0.203	0.251	0.319	0.215
	m2	0.271	0.288	0.238	0.282	0.325	0.236	0.321	0.334	0.250	0.344	0.383	0.262
	m3	0.259	0.282	0.217	0.278	0.325	0.210	0.320	0.332	0.237	0.343	0.393	0.251
μ2	m1	0.287	0.368	0.253	0.242	0.345	0.231	0.298	0.361	0.229	0.292	0.436	0.242
	m2	0.279	0.293	0.246	0.287	0.308	0.251	0.298	0.336	0.258	0.367	0.365	0.272
	m3	0.272	0.289	0.216	0.284	0.307	0.238	0.300	0.339	0.246	0.365	0.370	0.256
μ3	m1	0.257	0.308	0.215	0.231	0.302	0.203	0.262	0.321	0.198	0.259	0.298	0.206
	m2	0.283	0.286	0.252	0.288	0.310	0.242	0.326	0.362	0.263	0.355	0.370	0.271
	m3	0.278	0.278	0.229	0.286	0.312	0.219	0.325	0.364	0.256	0.355	0.368	0.251
μ4	m1	0.963	1.335	0.657	0.918	1.317	0.675	0.965	1.362	0.622	0.991	1.384	0.635
	m2	0.402	0.678	0.282	0.437	0.740	0.314	0.682	0.943	0.414	0.913	1.177	0.554
	m3	0.496	0.894	0.354	0.529	0.932	0.401	0.765	1.114	0.495	0.971	1.278	0.616
μ5	m1	0.270	0.448	0.272	0.272	0.428	0.276	0.247	0.416	0.293	0.265	0.432	0.286
	m2	0.272	0.298	0.227	0.290	0.353	0.229	0.353	0.382	0.260	0.416	0.507	0.324
	m3	0.263	0.321	0.219	0.294	0.380	0.234	0.362	0.415	0.278	0.443	0.546	0.349
μ6	m1	0.282	0.352	0.219	0.265	0.353	0.224	0.283	0.358	0.203	0.300	0.385	0.229
	m2	0.294	0.306	0.247	0.276	0.327	0.246	0.355	0.367	0.257	0.409	0.410	0.272
	m3	0.291	0.305	0.222	0.270	0.332	0.231	0.359	0.364	0.254	0.409	0.414	0.263
μ7	m1	0.254	0.317	0.221	0.232	0.303	0.222	0.225	0.321	0.215	0.245	0.336	0.245
	m2	0.279	0.295	0.238	0.281	0.300	0.244	0.314	0.330	0.242	0.366	0.373	0.291
	m3	0.273	0.294	0.220	0.277	0.300	0.226	0.312	0.336	0.230	0.372	0.382	0.284
S2	no	0.248	0.447	0.233	0.248	0.447	0.233	0.248	0.447	0.233	0.248	0.447	0.233
μ1	m1	0.215	0.278	0.155	0.223	0.289	0.165	0.216	0.299	0.151	0.217	0.300	0.160
	m2	0.204	0.242	0.152	0.221	0.267	0.166	0.248	0.297	0.174	0.298	0.316	0.192
	m3	0.205	0.242	0.157	0.223	0.270	0.174	0.248	0.295	0.176	0.296	0.316	0.191
μ2	m1	0.244	0.322	0.183	0.237	0.331	0.173	0.236	0.341	0.180	0.236	0.324	0.185
	m2	0.204	0.244	0.156	0.211	0.275	0.162	0.274	0.313	0.180	0.314	0.322	0.208
	m3	0.202	0.244	0.163	0.207	0.277	0.163	0.269	0.315	0.189	0.307	0.319	0.209
μ3	m1	0.232	0.334	0.183	0.235	0.310	0.166	0.224	0.327	0.172	0.234	0.333	0.190
	m2	0.614	0.599	0.346	0.628	0.595	0.312	0.625	0.636	0.349	0.626	0.631	0.381
	m3	0.615	0.594	0.351	0.625	0.592	0.320	0.624	0.637	0.355	0.622	0.633	0.375
μ4	m1	0.544	0.813	0.432	0.539	0.800	0.423	0.575	0.768	0.437	0.551	0.830	0.441
	m2	0.217	0.349	0.167	0.239	0.403	0.204	0.357	0.471	0.266	0.480	0.640	0.349
	m3	0.227	0.427	0.180	0.252	0.479	0.229	0.382	0.545	0.289	0.492	0.686	0.364
μ5	m1	0.257	0.414	0.185	0.280	0.429	0.206	0.264	0.382	0.218	0.258	0.408	0.198
	m2	0.201	0.298	0.157	0.230	0.316	0.158	0.277	0.348	0.185	0.313	0.391	0.191
	m3	0.196	0.306	0.157	0.235	0.328	0.160	0.284	0.361	0.193	0.319	0.404	0.196
μ6	m1	0.232	0.317	0.190	0.229	0.298	0.185	0.239	0.305	0.191	0.226	0.305	0.195
	m2	0.198	0.260	0.162	0.228	0.253	0.168	0.259	0.296	0.186	0.298	0.317	0.212
	m3	0.193	0.261	0.162	0.227	0.256	0.173	0.263	0.302	0.195	0.294	0.322	0.220
μ7	m1	0.214	0.299	0.156	0.233	0.317	0.164	0.220	0.316	0.160	0.228	0.301	0.156
	m2	0.200	0.236	0.149	0.210	0.258	0.165	0.271	0.296	0.181	0.280	0.308	0.180
	m3	0.201	0.239	0.156	0.212	0.258	0.169	0.272	0.293	0.183	0.284	0.311	0.185
S3	no	0.864	0.738	0.485	0.864	0.738	0.485	0.864	0.738	0.485	0.864	0.738	0.485
μ1	m1	0.194	0.281	0.163	0.189	0.273	0.167	0.191	0.258	0.162	0.191	0.269	0.170
	m2	0.185	0.223	0.164	0.218	0.249	0.171	0.238	0.274	0.187	0.295	0.310	0.203
	m3	0.182	0.228	0.162	0.212	0.243	0.169	0.238	0.278	0.176	0.298	0.316	0.196
μ2	m1	0.222	0.319	0.178	0.202	0.328	0.184	0.211	0.333	0.171	0.205	0.317	0.176

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
	m2	0.194	0.234	0.166	0.197	0.244	0.182	0.261	0.284	0.200	0.292	0.320	0.208
	m3	0.194	0.236	0.154	0.194	0.246	0.175	0.256	0.285	0.192	0.292	0.313	0.202
μ3	m1	0.218	0.334	0.176	0.208	0.336	0.196	0.216	0.312	0.183	0.218	0.296	0.175
	m2	0.673	0.711	0.357	0.671	0.697	0.384	0.698	0.688	0.386	0.673	0.661	0.358
	m3	0.665	0.718	0.356	0.668	0.701	0.381	0.698	0.686	0.385	0.674	0.663	0.370
μ4	m1	0.414	0.605	0.296	0.414	0.612	0.294	0.397	0.576	0.278	0.391	0.575	0.300
	m2	0.192	0.284	0.175	0.241	0.350	0.197	0.304	0.388	0.207	0.372	0.472	0.250
	m3	0.202	0.343	0.166	0.247	0.406	0.191	0.318	0.448	0.209	0.390	0.516	0.243
μ5	m1	0.221	0.368	0.233	0.227	0.374	0.217	0.237	0.359	0.224	0.235	0.351	0.222
	m2	0.193	0.275	0.171	0.211	0.263	0.173	0.280	0.303	0.215	0.312	0.362	0.242
	m3	0.203	0.294	0.176	0.222	0.288	0.176	0.285	0.331	0.225	0.322	0.387	0.264
μ6	m1	0.206	0.292	0.170	0.210	0.313	0.170	0.201	0.280	0.161	0.203	0.279	0.174
	m2	0.181	0.229	0.178	0.204	0.257	0.171	0.255	0.298	0.190	0.279	0.312	0.209
	m3	0.180	0.233	0.164	0.200	0.260	0.162	0.257	0.303	0.183	0.281	0.313	0.201
μ7	m1	0.184	0.281	0.170	0.203	0.289	0.175	0.194	0.272	0.175	0.200	0.277	0.169
	m2	0.172	0.215	0.173	0.202	0.242	0.181	0.252	0.279	0.200	0.278	0.311	0.217
	m3	0.173	0.227	0.166	0.204	0.249	0.181	0.251	0.291	0.195	0.285	0.316	0.221

* Note: S1, S2, and S3 indicate subtests 1, 2, and 3, respectively. 'no' refers to no equating. μ1 = (0,0,0), μ2 = (0.1,0.1,0.1), μ3 = (0,0.1, −0.1), μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. m1 is the first method of using subtest anchor score as the anchor; m2 is the second method of using equated total score; and m3 is the third method of using anchor total score.

3.4.3.3 RMSE

Table 3.23, Table 3.24, and Table 3.25 present RMSE values from three different approaches: using observed scores, weighted averages, and augmented scores, respectively. RMSEs illustrate very similar patterns of the results as presented in the previous section (3.4.3.2).

Table 3.23. RMSE compared with True Equating Function using Observed Scores

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
S1	no	0.535	0.486	0.307	0.535	0.486	0.307	0.535	0.486	0.307	0.535	0.486	0.307
μ1	m1	0.322	0.400	0.323	0.305	0.402	0.321	0.296	0.397	0.287	0.323	0.448	0.344
	m2	0.325	0.363	0.334	0.339	0.409	0.342	0.381	0.415	0.338	0.455	0.506	0.390
	m3	0.315	0.362	0.314	0.335	0.405	0.308	0.382	0.421	0.321	0.464	0.526	0.383
μ2	m1	0.400	0.514	0.358	0.310	0.468	0.328	0.363	0.450	0.322	0.366	0.554	0.350
	m2	0.356	0.387	0.343	0.345	0.397	0.345	0.369	0.437	0.358	0.487	0.499	0.408
	m3	0.354	0.389	0.317	0.341	0.396	0.326	0.371	0.444	0.341	0.495	0.519	0.395
μ3	m1	0.344	0.416	0.321	0.290	0.383	0.302	0.331	0.404	0.291	0.337	0.403	0.334
	m2	0.347	0.388	0.341	0.332	0.378	0.336	0.384	0.435	0.354	0.464	0.493	0.407
	m3	0.345	0.384	0.320	0.331	0.379	0.310	0.387	0.447	0.340	0.469	0.506	0.388
μ4	m1	1.364	1.767	0.810	1.259	1.649	0.783	1.192	1.597	0.683	1.249	1.679	0.685
	m2	0.673	1.007	0.617	0.648	0.990	0.576	0.793	1.123	0.569	1.064	1.406	0.642
	m3	0.848	1.334	0.661	0.843	1.269	0.619	0.959	1.335	0.611	1.200	1.541	0.687
μ5	m1	0.364	0.551	0.406	0.410	0.573	0.456	0.411	0.614	0.499	0.449	0.656	0.541
	m2	0.293	0.348	0.353	0.394	0.495	0.400	0.512	0.579	0.461	0.660	0.768	0.593
	m3	0.295	0.376	0.332	0.405	0.522	0.388	0.532	0.621	0.470	0.694	0.814	0.612
μ6	m1	0.414	0.514	0.347	0.351	0.485	0.347	0.347	0.471	0.282	0.352	0.497	0.337
	m2	0.375	0.418	0.342	0.343	0.430	0.352	0.405	0.457	0.338	0.500	0.535	0.377
	m3	0.380	0.437	0.330	0.340	0.443	0.342	0.417	0.470	0.332	0.506	0.551	0.377
μ7	m1	0.319	0.399	0.325	0.304	0.400	0.332	0.297	0.443	0.328	0.342	0.487	0.393
	m2	0.317	0.353	0.326	0.339	0.388	0.347	0.383	0.436	0.353	0.514	0.523	0.446
	m3	0.314	0.354	0.305	0.335	0.388	0.317	0.387	0.449	0.335	0.526	0.545	0.441
S2	no	0.084	0.213	0.074	0.084	0.213	0.074	0.084	0.213	0.074	0.084	0.213	0.074
μ1	m1	0.302	0.391	0.291	0.309	0.398	0.264	0.270	0.357	0.217	0.283	0.366	0.222
	m2	0.246	0.314	0.261	0.255	0.324	0.238	0.292	0.353	0.218	0.361	0.394	0.243
	m3	0.251	0.327	0.278	0.261	0.334	0.252	0.292	0.350	0.235	0.360	0.397	0.252
μ2	m1	0.327	0.438	0.289	0.319	0.430	0.268	0.311	0.442	0.273	0.320	0.440	0.278
	m2	0.241	0.322	0.245	0.252	0.335	0.228	0.335	0.392	0.243	0.390	0.441	0.264
	m3	0.245	0.339	0.269	0.249	0.344	0.247	0.330	0.396	0.265	0.386	0.439	0.276
μ3	m1	0.338	0.440	0.304	0.331	0.424	0.262	0.308	0.432	0.254	0.317	0.438	0.291
	m2	0.542	0.556	0.289	0.586	0.589	0.276	0.641	0.678	0.351	0.676	0.699	0.408
	m3	0.551	0.563	0.305	0.588	0.594	0.293	0.642	0.683	0.371	0.674	0.705	0.418
μ4	m1	0.931	1.178	0.610	0.900	1.140	0.560	0.834	1.027	0.524	0.801	1.078	0.497
	m2	0.325	0.540	0.312	0.357	0.608	0.321	0.466	0.638	0.354	0.601	0.843	0.391
	m3	0.433	0.771	0.383	0.447	0.771	0.374	0.529	0.757	0.385	0.652	0.916	0.422
μ5	m1	0.431	0.628	0.448	0.431	0.611	0.428	0.379	0.533	0.376	0.376	0.544	0.317
	m2	0.309	0.457	0.340	0.325	0.447	0.310	0.375	0.469	0.296	0.426	0.513	0.289
	m3	0.300	0.474	0.352	0.328	0.455	0.324	0.378	0.484	0.313	0.437	0.530	0.298
μ6	m1	0.295	0.402	0.297	0.301	0.396	0.273	0.308	0.391	0.257	0.296	0.419	0.272
	m2	0.234	0.310	0.238	0.253	0.296	0.230	0.314	0.365	0.241	0.378	0.420	0.275
	m3	0.231	0.322	0.254	0.250	0.310	0.248	0.321	0.371	0.263	0.381	0.428	0.284
μ7	m1	0.320	0.453	0.329	0.324	0.446	0.297	0.284	0.388	0.250	0.296	0.372	0.221
	m2	0.249	0.332	0.270	0.259	0.344	0.254	0.317	0.358	0.236	0.358	0.390	0.234
	m3	0.255	0.350	0.290	0.267	0.353	0.268	0.319	0.355	0.250	0.364	0.393	0.243
S3	no	0.855	0.636	0.308	0.855	0.636	0.308	0.855	0.636	0.308	0.855	0.636	0.308
μ1	m1	0.234	0.327	0.218	0.235	0.319	0.221	0.230	0.309	0.217	0.226	0.313	0.214
	m2	0.207	0.262	0.206	0.241	0.282	0.212	0.282	0.324	0.237	0.354	0.376	0.250
	m3	0.204	0.268	0.203	0.238	0.280	0.210	0.283	0.326	0.233	0.353	0.378	0.243
μ2	m1	0.314	0.421	0.270	0.280	0.418	0.253	0.283	0.423	0.243	0.279	0.404	0.245
	m2	0.255	0.301	0.210	0.234	0.306	0.221	0.304	0.351	0.244	0.333	0.387	0.250
	m3	0.252	0.304	0.205	0.233	0.310	0.217	0.302	0.352	0.239	0.331	0.383	0.248

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
μ3	m1	0.273	0.394	0.260	0.267	0.414	0.259	0.268	0.386	0.260	0.271	0.355	0.251
	m2	0.637	0.680	0.345	0.652	0.697	0.388	0.699	0.699	0.417	0.681	0.683	0.410
	m3	0.627	0.680	0.347	0.650	0.700	0.390	0.698	0.699	0.416	0.681	0.686	0.412
μ4	m1	0.532	0.742	0.519	0.507	0.705	0.454	0.461	0.644	0.380	0.453	0.641	0.362
	m2	0.235	0.367	0.263	0.282	0.415	0.248	0.350	0.444	0.266	0.425	0.540	0.294
	m3	0.255	0.442	0.313	0.294	0.479	0.275	0.368	0.509	0.281	0.446	0.587	0.306
μ5	m1	0.324	0.511	0.332	0.315	0.500	0.340	0.332	0.481	0.341	0.315	0.459	0.338
	m2	0.273	0.368	0.233	0.295	0.373	0.250	0.378	0.421	0.298	0.408	0.480	0.346
	m3	0.264	0.384	0.241	0.285	0.379	0.264	0.376	0.437	0.317	0.415	0.494	0.367
μ6	m1	0.245	0.345	0.240	0.260	0.367	0.225	0.244	0.330	0.205	0.240	0.334	0.219
	m2	0.205	0.268	0.215	0.229	0.296	0.212	0.299	0.346	0.228	0.329	0.391	0.260
	m3	0.202	0.272	0.205	0.226	0.302	0.213	0.300	0.353	0.221	0.327	0.386	0.253
μ7	m1	0.233	0.337	0.251	0.254	0.352	0.238	0.238	0.348	0.244	0.235	0.333	0.233
	m2	0.209	0.253	0.216	0.229	0.278	0.220	0.302	0.338	0.245	0.331	0.372	0.275
	m3	0.209	0.261	0.214	0.231	0.282	0.222	0.301	0.346	0.243	0.337	0.374	0.276

* Note: S1, S2, and S3 indicate subtests 1, 2, and 3, respectively. 'no' refers to no equating. $\mu1 = (0,0,0)$, $\mu2 = (0.1,0.1,0.1)$, $\mu3 = (0,0.1,-0.1)$, μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. m1 is the first method of using subtest anchor score as the anchor; m2 is the second method of using equated total score; and m3 is the third method of using anchor total score.

Table 3.24. RMSE compared with True Equating Function using Weighted Averages

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
S1	no	0.535	0.486	0.307	0.535	0.486	0.307	0.535	0.486	0.307	0.535	0.486	0.307
μ1	m1	0.406	0.445	0.348	0.340	0.430	0.339	0.317	0.412	0.294	0.319	0.446	0.334
	m2	0.405	0.391	0.352	0.358	0.439	0.360	0.389	0.417	0.343	0.428	0.490	0.382
	m3	0.399	0.384	0.328	0.349	0.434	0.319	0.389	0.420	0.320	0.433	0.508	0.364
μ2	m1	0.425	0.534	0.351	0.345	0.492	0.337	0.393	0.466	0.325	0.368	0.554	0.336
	m2	0.408	0.416	0.370	0.388	0.433	0.366	0.382	0.436	0.356	0.472	0.482	0.400
	m3	0.409	0.419	0.336	0.378	0.429	0.335	0.381	0.435	0.332	0.476	0.491	0.373
μ3	m1	0.397	0.484	0.334	0.329	0.423	0.339	0.356	0.419	0.299	0.333	0.394	0.318
	m2	0.406	0.425	0.360	0.378	0.405	0.365	0.400	0.438	0.357	0.446	0.475	0.398
	m3	0.402	0.422	0.328	0.373	0.403	0.328	0.401	0.445	0.334	0.449	0.481	0.366
μ4	m1	1.478	1.812	0.991	1.348	1.732	0.953	1.234	1.658	0.788	1.221	1.637	0.723
	m2	0.761	1.028	0.544	0.758	1.068	0.535	0.874	1.170	0.544	1.053	1.375	0.625
	m3	0.972	1.347	0.695	0.936	1.336	0.668	1.012	1.382	0.639	1.168	1.503	0.694
μ5	m1	0.463	0.676	0.507	0.413	0.601	0.482	0.366	0.567	0.474	0.395	0.607	0.499
	m2	0.349	0.371	0.364	0.346	0.460	0.379	0.436	0.495	0.410	0.591	0.695	0.543
	m3	0.344	0.449	0.393	0.363	0.513	0.401	0.453	0.541	0.435	0.623	0.741	0.567
μ6	m1	0.448	0.514	0.306	0.366	0.486	0.316	0.359	0.468	0.259	0.361	0.500	0.332
	m2	0.433	0.442	0.359	0.377	0.454	0.354	0.418	0.461	0.334	0.494	0.526	0.381
	m3	0.439	0.449	0.305	0.372	0.457	0.317	0.426	0.465	0.313	0.500	0.539	0.367
μ7	m1	0.386	0.469	0.372	0.336	0.432	0.373	0.308	0.448	0.332	0.337	0.477	0.379
	m2	0.402	0.399	0.344	0.372	0.413	0.367	0.376	0.431	0.343	0.485	0.498	0.438
	m3	0.393	0.402	0.327	0.363	0.422	0.344	0.377	0.444	0.325	0.493	0.518	0.426
S2	no	0.084	0.213	0.074	0.084	0.213	0.074	0.084	0.213	0.074	0.084	0.213	0.074
μ1	m1	0.423	0.497	0.291	0.360	0.460	0.297	0.307	0.411	0.256	0.307	0.394	0.257
	m2	0.310	0.344	0.232	0.304	0.368	0.275	0.312	0.389	0.264	0.367	0.405	0.278
	m3	0.353	0.416	0.259	0.318	0.394	0.288	0.316	0.392	0.268	0.367	0.408	0.281
μ2	m1	0.402	0.542	0.314	0.357	0.482	0.294	0.325	0.461	0.306	0.317	0.440	0.308
	m2	0.337	0.403	0.243	0.313	0.374	0.242	0.347	0.403	0.267	0.400	0.440	0.307
	m3	0.374	0.472	0.274	0.323	0.404	0.257	0.344	0.410	0.275	0.394	0.437	0.313
μ3	m1	0.416	0.542	0.369	0.364	0.454	0.324	0.316	0.426	0.297	0.326	0.451	0.346
	m2	0.663	0.670	0.492	0.706	0.697	0.468	0.717	0.745	0.481	0.717	0.742	0.508
	m3	0.674	0.684	0.505	0.707	0.699	0.479	0.716	0.748	0.490	0.714	0.747	0.514
μ4	m1	0.855	1.163	0.843	0.808	1.094	0.800	0.748	0.987	0.728	0.721	1.027	0.663
	m2	0.337	0.529	0.422	0.383	0.589	0.464	0.476	0.625	0.473	0.572	0.802	0.512
	m3	0.457	0.771	0.541	0.461	0.761	0.555	0.521	0.731	0.529	0.602	0.868	0.554
μ5	m1	0.571	0.784	0.364	0.405	0.585	0.296	0.352	0.478	0.279	0.340	0.499	0.250
	m2	0.338	0.442	0.256	0.307	0.410	0.232	0.336	0.418	0.256	0.387	0.468	0.251
	m3	0.395	0.584	0.293	0.324	0.444	0.241	0.343	0.436	0.259	0.396	0.489	0.249
μ6	m1	0.417	0.528	0.357	0.351	0.448	0.345	0.333	0.413	0.309	0.308	0.420	0.340
	m2	0.292	0.374	0.271	0.298	0.333	0.285	0.327	0.378	0.291	0.373	0.420	0.334
	m3	0.326	0.448	0.291	0.306	0.354	0.304	0.333	0.383	0.309	0.373	0.429	0.348
μ7	m1	0.446	0.585	0.288	0.355	0.481	0.277	0.307	0.413	0.244	0.306	0.394	0.249
	m2	0.303	0.369	0.233	0.289	0.370	0.256	0.335	0.375	0.263	0.357	0.399	0.263
	m3	0.352	0.457	0.260	0.303	0.396	0.266	0.339	0.373	0.262	0.363	0.404	0.270
S3	no	0.855	0.636	0.308	0.855	0.636	0.308	0.855	0.636	0.308	0.855	0.636	0.308
μ1	m1	0.423	0.517	0.296	0.327	0.453	0.278	0.269	0.366	0.245	0.253	0.342	0.234
	m2	0.290	0.319	0.240	0.295	0.333	0.240	0.296	0.354	0.257	0.367	0.387	0.261
	m3	0.334	0.385	0.263	0.317	0.369	0.253	0.305	0.368	0.257	0.370	0.395	0.258
μ2	m1	0.422	0.531	0.293	0.325	0.473	0.287	0.287	0.451	0.249	0.275	0.410	0.253
	m2	0.289	0.329	0.247	0.284	0.328	0.251	0.332	0.371	0.268	0.356	0.404	0.271

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
	m3	0.327	0.394	0.260	0.301	0.359	0.267	0.334	0.383	0.263	0.358	0.404	0.270
μ3	m1	0.477	0.543	0.294	0.338	0.456	0.286	0.267	0.380	0.264	0.290	0.376	0.254
	m2	0.689	0.770	0.392	0.704	0.754	0.447	0.741	0.745	0.454	0.721	0.720	0.430
	m3	0.686	0.800	0.407	0.708	0.770	0.455	0.743	0.748	0.457	0.723	0.726	0.439
μ4	m1	0.770	1.081	0.534	0.632	0.895	0.462	0.489	0.688	0.372	0.450	0.642	0.343
	m2	0.280	0.459	0.255	0.325	0.481	0.271	0.365	0.466	0.282	0.414	0.527	0.290
	m3	0.376	0.676	0.309	0.378	0.623	0.293	0.397	0.554	0.290	0.443	0.582	0.287
μ5	m1	0.470	0.642	0.400	0.390	0.531	0.330	0.326	0.457	0.308	0.308	0.424	0.297
	m2	0.309	0.420	0.252	0.295	0.342	0.232	0.342	0.371	0.275	0.384	0.437	0.306
	m3	0.370	0.513	0.294	0.345	0.402	0.260	0.361	0.411	0.296	0.400	0.466	0.332
μ6	m1	0.414	0.537	0.271	0.339	0.483	0.268	0.274	0.377	0.240	0.266	0.363	0.231
	m2	0.259	0.334	0.244	0.270	0.346	0.240	0.319	0.374	0.252	0.333	0.403	0.270
	m3	0.298	0.405	0.237	0.284	0.383	0.244	0.324	0.392	0.247	0.333	0.404	0.261
μ7	m1	0.430	0.541	0.310	0.355	0.458	0.281	0.274	0.374	0.268	0.267	0.357	0.247
	m2	0.290	0.350	0.249	0.295	0.338	0.239	0.310	0.347	0.262	0.344	0.382	0.276
	m3	0.340	0.428	0.265	0.326	0.384	0.252	0.317	0.367	0.267	0.356	0.392	0.284

* Note: S1, S2, and S3 indicate subtests 1, 2, and 3, respectively. 'no' refers to no equating. $\mu 1 = (0,0,0)$, $\mu 2 = (0.1,0.1,0.1)$, $\mu 3 = (0,0.1,-0.1)$, μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. m1 is the first method of using subtest anchor score as the anchor; m2 is the second method of using equated total score; and m3 is the third method of using anchor total score.

Table 3.25. RMSE compared with True Equating Function using Augmented Scores

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
S1	no	0.535	0.486	0.307	0.535	0.486	0.307	0.535	0.486	0.307	0.535	0.486	0.307
μ1	m1	0.311	0.376	0.281	0.303	0.394	0.276	0.310	0.386	0.251	0.312	0.407	0.263
	m2	0.337	0.358	0.297	0.353	0.419	0.296	0.399	0.416	0.307	0.437	0.482	0.324
	m3	0.325	0.355	0.280	0.348	0.418	0.264	0.397	0.416	0.292	0.437	0.495	0.313
μ2	m1	0.356	0.470	0.313	0.295	0.441	0.293	0.370	0.446	0.284	0.364	0.536	0.296
	m2	0.349	0.367	0.300	0.355	0.390	0.313	0.374	0.428	0.313	0.463	0.456	0.336
	m3	0.343	0.363	0.270	0.351	0.389	0.299	0.373	0.429	0.302	0.462	0.461	0.317
μ3	m1	0.320	0.385	0.271	0.289	0.374	0.256	0.337	0.395	0.250	0.322	0.377	0.261
	m2	0.340	0.363	0.306	0.360	0.388	0.299	0.402	0.446	0.324	0.451	0.462	0.344
	m3	0.335	0.358	0.280	0.360	0.391	0.277	0.400	0.447	0.313	0.449	0.464	0.321
μ4	m1	1.033	1.415	0.716	0.987	1.398	0.735	1.032	1.446	0.678	1.056	1.473	0.685
	m2	0.495	0.758	0.358	0.540	0.835	0.379	0.779	1.042	0.485	1.005	1.275	0.639
	m3	0.585	0.972	0.443	0.640	1.028	0.473	0.866	1.218	0.567	1.065	1.377	0.696
μ5	m1	0.330	0.531	0.326	0.335	0.512	0.335	0.304	0.497	0.343	0.328	0.516	0.335
	m2	0.333	0.363	0.276	0.366	0.458	0.279	0.423	0.464	0.310	0.518	0.608	0.387
	m3	0.325	0.391	0.270	0.374	0.485	0.287	0.434	0.500	0.331	0.545	0.647	0.413
μ6	m1	0.358	0.431	0.288	0.324	0.438	0.290	0.352	0.455	0.256	0.364	0.470	0.286
	m2	0.365	0.379	0.302	0.350	0.418	0.313	0.438	0.467	0.319	0.508	0.522	0.346
	m3	0.365	0.381	0.275	0.345	0.424	0.298	0.445	0.467	0.315	0.510	0.525	0.338
μ7	m1	0.314	0.393	0.279	0.292	0.379	0.277	0.285	0.413	0.266	0.309	0.426	0.293
	m2	0.348	0.365	0.289	0.359	0.386	0.297	0.393	0.420	0.297	0.463	0.467	0.361
	m3	0.340	0.363	0.271	0.351	0.387	0.277	0.387	0.428	0.281	0.466	0.478	0.354
S2	no	0.084	0.213	0.074	0.084	0.213	0.074	0.084	0.213	0.074	0.084	0.213	0.074
μ1	m1	0.264	0.345	0.198	0.283	0.363	0.208	0.265	0.368	0.192	0.272	0.363	0.196
	m2	0.251	0.295	0.194	0.278	0.326	0.208	0.310	0.371	0.225	0.370	0.395	0.236
	m3	0.253	0.296	0.199	0.281	0.331	0.215	0.310	0.369	0.225	0.368	0.396	0.238
μ2	m1	0.303	0.396	0.225	0.298	0.404	0.211	0.293	0.424	0.229	0.291	0.402	0.236
	m2	0.254	0.299	0.198	0.260	0.328	0.206	0.337	0.389	0.228	0.393	0.421	0.259
	m3	0.252	0.302	0.207	0.255	0.331	0.210	0.333	0.391	0.235	0.384	0.416	0.261
μ3	m1	0.287	0.402	0.229	0.289	0.391	0.203	0.275	0.398	0.213	0.287	0.413	0.235
	m2	0.658	0.668	0.395	0.681	0.670	0.366	0.696	0.721	0.408	0.700	0.720	0.440
	m3	0.660	0.666	0.401	0.678	0.669	0.377	0.693	0.720	0.413	0.696	0.722	0.435
μ4	m1	0.602	0.889	0.476	0.595	0.877	0.473	0.634	0.844	0.486	0.614	0.905	0.483
	m2	0.270	0.424	0.218	0.294	0.475	0.257	0.434	0.549	0.327	0.561	0.734	0.411
	m3	0.280	0.504	0.233	0.314	0.555	0.287	0.457	0.626	0.354	0.573	0.780	0.424
μ5	m1	0.313	0.493	0.226	0.334	0.507	0.253	0.323	0.460	0.267	0.314	0.485	0.240
	m2	0.250	0.363	0.205	0.282	0.385	0.200	0.347	0.421	0.234	0.387	0.475	0.244
	m3	0.246	0.374	0.202	0.289	0.399	0.201	0.353	0.438	0.240	0.395	0.491	0.250
μ6	m1	0.284	0.388	0.237	0.285	0.372	0.230	0.298	0.383	0.234	0.277	0.377	0.245
	m2	0.245	0.316	0.208	0.288	0.304	0.211	0.329	0.367	0.237	0.371	0.397	0.272
	m3	0.237	0.319	0.207	0.284	0.311	0.216	0.335	0.373	0.248	0.368	0.401	0.277
μ7	m1	0.259	0.368	0.193	0.291	0.390	0.202	0.271	0.386	0.199	0.285	0.372	0.196
	m2	0.245	0.295	0.190	0.267	0.330	0.210	0.336	0.368	0.230	0.357	0.392	0.227
	m3	0.246	0.298	0.199	0.270	0.330	0.214	0.336	0.366	0.230	0.362	0.395	0.231
S3	no	0.855	0.636	0.308	0.855	0.636	0.308	0.855	0.636	0.308	0.855	0.636	0.308
μ1	m1	0.242	0.350	0.206	0.239	0.336	0.208	0.238	0.322	0.200	0.237	0.336	0.208
	m2	0.229	0.279	0.203	0.266	0.303	0.210	0.297	0.340	0.235	0.370	0.389	0.248
	m3	0.227	0.290	0.202	0.263	0.301	0.207	0.300	0.346	0.226	0.371	0.396	0.241
μ2	m1	0.274	0.395	0.225	0.250	0.400	0.229	0.260	0.412	0.212	0.254	0.386	0.220
	m2	0.241	0.289	0.200	0.248	0.301	0.222	0.325	0.352	0.245	0.363	0.397	0.255

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
	m3	0.239	0.293	0.190	0.245	0.306	0.219	0.320	0.355	0.237	0.362	0.392	0.251
μ3	m1	0.264	0.407	0.220	0.262	0.414	0.233	0.260	0.382	0.226	0.273	0.364	0.214
	m2	0.706	0.758	0.390	0.713	0.756	0.424	0.755	0.755	0.434	0.732	0.729	0.410
	m3	0.699	0.766	0.390	0.712	0.761	0.424	0.755	0.757	0.434	0.734	0.735	0.421
μ4	m1	0.463	0.676	0.339	0.471	0.685	0.341	0.452	0.648	0.325	0.443	0.647	0.345
	m2	0.237	0.343	0.212	0.298	0.417	0.236	0.383	0.471	0.256	0.449	0.557	0.302
	m3	0.245	0.404	0.207	0.304	0.478	0.231	0.399	0.534	0.258	0.471	0.603	0.297
μ5	m1	0.273	0.460	0.275	0.274	0.455	0.262	0.290	0.437	0.268	0.293	0.420	0.266
	m2	0.238	0.338	0.214	0.255	0.327	0.216	0.339	0.373	0.261	0.383	0.441	0.293
	m3	0.247	0.369	0.217	0.268	0.353	0.220	0.352	0.405	0.273	0.397	0.468	0.316
μ6	m1	0.251	0.358	0.210	0.266	0.388	0.208	0.253	0.345	0.203	0.250	0.348	0.216
	m2	0.220	0.285	0.219	0.255	0.320	0.215	0.320	0.366	0.235	0.339	0.405	0.263
	m3	0.221	0.293	0.204	0.250	0.328	0.208	0.321	0.375	0.227	0.340	0.401	0.253
μ7	m1	0.233	0.350	0.216	0.255	0.362	0.219	0.241	0.349	0.219	0.250	0.347	0.209
	m2	0.222	0.265	0.215	0.254	0.300	0.222	0.314	0.342	0.246	0.345	0.383	0.266
	m3	0.224	0.280	0.206	0.258	0.307	0.222	0.314	0.355	0.242	0.355	0.390	0.271

* Note: S1, S2, and S3 indicate subtests 1, 2, and 3, respectively. 'no' refers to no equating. $\mu1 = (0,0,0)$, $\mu2 = (0.1,0.1,0.1)$, $\mu3 = (0,0.1,-0.1)$, μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. m1 is the first method of using subtest anchor score as the anchor; m2 is the second method of using equated total score; and m3 is the third method of using anchor total score.

3.4.4 Number of Score Points Equal between True and Estimated Conversion and score difference that matters (DTM)

3.4.4.1 Score points equal between true and estimated equating output

Table 3.26, Table 3.27, and Table 3.28 present the percentages which score points equal when true and estimated conversion tables were compared. Larger numbers indicate more accurate equating results. It means equating results are similar to true conversions.

When observed scores were used and proficiency distributions remained the same in two groups, μ1=(0,0,0), using the equated total or anchor total score approach outperformed the subtest anchor score approach in the conditions of high correlations and smaller number of items (16 items) or common items (32 items with 4 common items) in subtest 1. As the correlation decreased, using the anchor set in each subtest produced better results. In most cases, percentages dropped as proficiency distributions differed as mean shifts and skewness were introduced. However, the positively skewed distribution with skewness of 0.25 (μ6) yielded very similar results to the normal distribution with identical mean vectors in two groups (μ1), or even showed slightly higher percentages in a few cases. Skewed distributions with high skewness (μ4 and μ5) produced less accurate equating results compared to the other distributions. When there was a mean shift to $\mu 3 = (0{,}0.1, -0.1)$, adopting the equated total or anchor total score approach did not perform well in subtests 2 and 3.

No equating in subtest 3, where the mean score differences between FormX and FormY were the largest among three subtests, showed that only 15 or 21 percentages out of 33 score points were equal in the true and estimated conversion tables. Regardless of simulation factors and methods—using different anchor sets or input scores—performing equating produced more accurate results than no equating in all cases from subtest 3. In addition, most cases showed better results than no equating in subtest 1. In subtest 2, on the other hand, no equating showed higher percentages than conducting equating.

A similar pattern was found in weighted average and augmented score results (see Table 3.27 and Table 3.28). The weighted average approach did not perform as well as the observed score approach or augmented score approach, which was also reported in the previous sections. In a few cases when the correlation was relatively low, weighted averages produced slightly higher percentages than observed scores or augmented scores. In most cases, observed scores showed better results than weighted averages. Moreover, more score points were equal when augmented scores were adopted compared to using either observed scores or weighted averages, especially in the case of having a smaller number of items (16 items).

Table 3.26. Average Percentages of the Number of Score Points Equal between True and Estimated Conversion using Observed Scores

Item	Dist*	ρ	Subtest1			Subtest2			Subtest3		
			m1	m2	m3	m1	m2	m3	m1	m2	m3
32	No EQ		48.48	48.48	48.48	60.61	60.61	60.61	21.21	21.21	21.21
&	μ1	0.9	65.73	65.85	67.79	58.33	58.21	59.18	68.79	73.76	72.64
4		0.8	62.91	62.76	63.18	57.94	56.94	58.15	70.06	72.03	72.88
		0.6	63.52	60.88	61.52	60.21	57.36	58.91	71.73	68.97	68.39
		0.4	63.21	59.06	58.58	59.52	59.33	58.79	71.27	66.85	66.06
	μ2	0.9	57.00	64.55	66.73	55.91	57.64	59.09	66.58	72.85	73.42
		0.8	57.33	63.39	64.82	55.76	55.55	57.06	63.67	71.55	71.15
		0.6	56.97	62.76	63.94	54.15	56.30	56.30	63.52	69.15	68.76
		0.4	52.24	62.12	61.85	55.73	56.79	57.00	64.73	66.03	66.55
	μ3	0.9	64.33	67.27	67.91	52.79	49.18	49.03	67.33	36.82	36.85
		0.8	63.55	63.36	63.64	54.76	48.70	48.15	67.15	39.27	38.82
		0.6	61.21	59.42	59.91	52.73	45.88	46.39	68.70	41.03	41.39
		0.4	64.94	61.03	62.48	54.27	46.76	46.79	69.67	42.18	42.76
	μ4	0.9	43.00	50.70	46.97	43.79	53.27	53.24	55.91	81.76	76.64
		0.8	43.03	48.06	46.42	44.18	52.64	50.85	55.70	76.00	70.85
		0.6	42.39	44.03	43.36	44.55	52.58	50.79	58.00	72.12	66.70
		0.4	42.42	42.94	42.55	44.39	48.70	47.97	58.21	63.82	61.45
	μ5	0.9	44.33	57.58	57.15	48.48	54.27	53.73	50.88	62.18	57.94
		0.8	45.24	54.61	53.91	48.97	54.91	53.79	49.12	63.73	57.27
		0.6	45.09	52.88	52.15	48.33	53.21	51.58	49.91	59.91	53.00
		0.4	44.97	45.09	44.70	47.94	51.48	50.18	49.18	51.76	48.85
	μ6	0.9	66.39	67.18	68.00	57.36	56.91	57.67	72.64	77.88	77.94
		0.8	66.12	66.64	67.06	57.76	57.55	58.45	71.67	74.09	74.48
		0.6	65.45	63.00	63.85	57.24	56.27	57.39	74.09	71.64	71.48
		0.4	63.48	60.45	60.36	56.67	57.30	58.00	73.45	71.18	71.64
	μ7	0.9	57.39	61.82	63.24	56.97	56.48	57.61	65.00	73.18	71.61
		0.8	59.33	62.09	62.52	55.55	54.97	55.18	63.94	69.61	69.76
		0.6	59.00	58.24	59.00	55.82	55.88	55.52	65.27	65.55	63.79
		0.4	56.88	55.30	55.85	56.82	56.58	56.64	64.85	64.24	63.33
32	No EQ		45.45	45.45	45.45	87.88	87.88	87.88	15.15	15.15	15.15
&	μ1	0.9	68.39	64.39	67.88	76.15	76.61	76.97	79.36	76.97	77.06
8		0.8	68.91	64.09	65.21	76.24	74.15	73.24	77.36	75.94	76.42
		0.6	68.94	62.06	62.03	74.58	72.15	71.76	77.73	74.09	74.61
		0.4	66.94	60.24	61.09	74.91	67.33	67.30	79.00	65.76	65.67
	μ2	0.9	64.88	66.82	67.61	66.94	74.09	74.27	75.00	75.91	75.88
		0.8	67.00	63.18	64.15	69.70	78.73	79.00	75.39	74.73	75.24
		0.6	65.15	64.94	64.85	68.30	72.27	72.42	76.91	69.91	70.97
		0.4	64.91	60.61	61.67	68.27	73.39	74.09	78.36	67.48	68.36
	μ3	0.9	67.94	63.06	64.12	68.09	32.24	32.70	76.45	33.82	34.39
		0.8	69.79	63.91	65.73	69.03	31.33	31.73	76.39	34.91	35.33
		0.6	68.06	61.61	62.48	71.82	33.39	33.27	76.09	36.00	35.73
		0.4	67.18	60.85	61.36	69.06	32.42	32.55	75.03	36.21	36.30
	μ4	0.9	40.42	59.36	55.30	60.79	72.82	73.33	67.42	79.09	82.09
		0.8	41.21	57.58	55.48	59.12	66.36	65.48	68.00	76.30	78.48
		0.6	41.00	49.64	49.06	59.67	59.52	58.70	70.48	74.91	74.48

Item	Dist*	ρ	Subtest1			Subtest2			Subtest3		
			m1	m2	m3	m1	m2	m3	m1	m2	m3
		0.4	40.45	45.03	45.06	59.73	54.06	53.88	70.94	71.36	69.18
	μ5	0.9	58.03	62.27	64.06	59.79	71.97	74.18	60.24	71.18	69.67
		0.8	57.67	59.36	60.67	56.45	69.73	70.36	61.55	67.42	66.12
		0.6	59.58	54.58	56.03	60.18	66.03	65.91	60.36	59.21	58.21
		0.4	58.82	52.12	51.88	61.39	63.70	62.76	59.79	56.79	55.48
	μ6	0.9	70.48	66.94	67.27	78.00	79.45	80.00	82.09	78.06	79.15
		0.8	69.64	66.73	67.82	78.73	74.24	73.76	79.76	78.91	79.52
		0.6	69.00	63.06	62.36	75.48	70.97	70.09	80.79	73.12	72.85
		0.4	67.21	59.97	60.27	78.06	67.45	68.42	82.76	70.64	70.76
	μ7	0.9	62.97	62.03	63.97	73.21	77.21	77.70	75.24	77.79	76.97
		0.8	65.85	62.55	64.27	68.24	74.27	74.39	74.00	75.67	75.97
		0.6	67.12	60.58	62.76	70.91	69.36	69.48	74.30	70.36	69.76
		0.4	65.67	58.61	59.42	69.97	67.21	67.06	73.76	64.82	64.88
16 & 4	No EQ		48.48	48.48	48.48	60.61	60.61	60.61	21.21	21.21	21.21
	μ1	0.9	65.73	65.85	67.79	58.33	58.21	59.18	68.79	73.76	72.64
		0.8	62.91	62.76	63.18	57.94	56.94	58.15	70.06	72.03	72.88
		0.6	63.52	60.88	61.52	60.21	57.36	58.91	71.73	68.97	68.39
		0.4	63.21	59.06	58.58	59.52	59.33	58.79	71.27	66.85	66.06
	μ2	0.9	57.00	64.55	66.73	55.91	57.64	59.09	66.58	72.85	73.42
		0.8	57.33	63.39	64.82	55.76	55.55	57.06	63.67	71.55	71.15
		0.6	56.97	62.76	63.94	54.15	56.30	56.30	63.52	69.15	68.76
		0.4	52.24	62.12	61.85	55.73	56.79	57.00	64.73	66.03	66.55
	μ3	0.9	64.33	67.27	67.91	52.79	49.18	49.03	67.33	36.82	36.85
		0.8	63.55	63.36	63.64	54.76	48.70	48.15	67.15	39.27	38.82
		0.6	61.21	59.42	59.91	52.73	45.88	46.39	68.70	41.03	41.39
		0.4	64.94	61.03	62.48	54.27	46.76	46.79	69.67	42.18	42.76
	μ4	0.9	43.00	50.70	46.97	43.79	53.27	53.24	55.91	81.76	76.64
		0.8	43.03	48.06	46.42	44.18	52.64	50.85	55.70	76.00	70.85
		0.6	42.39	44.03	43.36	44.55	52.58	50.79	58.00	72.12	66.70
		0.4	42.42	42.94	42.55	44.39	48.70	47.97	58.21	63.82	61.45
	μ5	0.9	44.33	57.58	57.15	48.48	54.27	53.73	50.88	62.18	57.94
		0.8	45.24	54.61	53.91	48.97	54.91	53.79	49.12	63.73	57.27
		0.6	45.09	52.88	52.15	48.33	53.21	51.58	49.91	59.91	53.00
		0.4	44.97	45.09	44.70	47.94	51.48	50.18	49.18	51.76	48.85
	μ6	0.9	66.39	67.18	68.00	57.36	56.91	57.67	72.64	77.88	77.94
		0.8	66.12	66.64	67.06	57.76	57.55	58.45	71.67	74.09	74.48
		0.6	65.45	63.00	63.85	57.24	56.27	57.39	74.09	71.64	71.48
		0.4	63.48	60.45	60.36	56.67	57.30	58.00	73.45	71.18	71.64
	μ7	0.9	57.39	61.82	63.24	56.97	56.48	57.61	65.00	73.18	71.61
		0.8	59.33	62.09	62.52	55.55	54.97	55.18	63.94	69.61	69.76
		0.6	59.00	58.24	59.00	55.82	55.88	55.52	65.27	65.55	63.79
		0.4	56.88	55.30	55.85	56.82	56.58	56.64	64.85	64.24	63.33

* Note: μ1 = (0,0,0), μ2 = (0.1,0.1,0.1), μ3 = (0,0.1, −0.1), μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. m1 is the first method of using subtest anchor score as the anchor; m2 is the second method of using equated total score; and m3 is the third method of using anchor total score.

Table 3.27. Average Percentages of the Number of Score Points Equal between True and Estimated Conversion using Weighted Averages

Item	Dist*	ρ	Subtest1 m1	Subtest1 m2	Subtest1 m3	Subtest2 m1	Subtest2 m2	Subtest2 m3	Subtest3 m1	Subtest3 m2	Subtest3 m3
32 & 4	No EQ		48.48	48.48	48.48	60.61	60.61	60.61	21.21	21.21	21.21
	μ1	0.9	55.82	59.00	60.18	49.33	48.36	52.12	55.18	67.73	63.52
		0.8	57.64	60.48	60.67	52.09	49.70	52.12	56.09	65.61	64.82
		0.6	61.64	60.70	61.48	58.58	55.00	56.15	64.76	64.33	64.91
		0.4	62.03	60.03	58.70	57.70	57.82	57.48	70.21	66.06	65.45
	μ2	0.9	51.06	61.76	63.24	50.64	48.67	51.24	53.88	64.06	61.55
		0.8	55.70	61.12	61.61	50.21	49.79	53.85	54.73	64.39	63.64
		0.6	56.58	62.67	62.36	51.21	54.45	55.18	58.03	64.03	65.03
		0.4	53.33	61.24	60.67	54.39	53.36	53.94	61.33	65.03	64.48
	μ3	0.9	55.91	59.70	62.27	48.94	38.76	39.91	54.24	33.79	33.42
		0.8	59.18	62.55	63.27	52.27	41.97	42.15	58.70	34.00	33.64
		0.6	60.52	59.18	59.24	52.67	44.39	44.36	64.09	39.21	39.79
		0.4	65.12	59.97	61.42	52.18	46.36	46.64	67.18	44.15	44.45
	μ4	0.9	37.24	48.70	41.39	36.30	53.70	48.85	38.55	65.45	50.09
		0.8	39.24	49.12	43.67	39.94	49.27	48.82	44.88	67.64	57.70
		0.6	41.88	45.76	44.15	42.76	47.73	47.48	56.21	73.21	65.76
		0.4	45.21	46.82	46.61	42.52	44.33	43.97	59.64	65.85	62.33
	μ5	0.9	31.42	50.12	45.88	36.24	45.39	45.61	45.97	56.85	51.70
		0.8	38.33	49.94	47.64	42.85	47.18	48.61	44.52	62.61	53.94
		0.6	42.97	50.45	49.09	47.00	51.76	50.33	48.03	58.36	52.06
		0.4	45.39	43.97	43.73	50.18	54.09	51.97	49.06	53.30	48.79
	μ6	0.9	62.12	63.52	62.64	50.24	49.67	53.70	53.91	67.33	63.39
		0.8	65.36	65.39	65.03	53.52	51.36	56.33	57.70	65.42	64.42
		0.6	65.45	64.15	64.24	57.30	56.09	56.94	68.61	67.30	67.06
		0.4	62.91	61.24	60.88	53.91	55.52	56.21	71.76	70.18	69.82
	μ7	0.9	50.67	56.33	56.30	46.64	48.15	50.03	52.33	63.06	59.52
		0.8	53.27	56.91	57.15	48.09	49.27	51.70	52.55	64.00	59.00
		0.6	56.88	57.30	57.94	53.91	54.52	54.94	61.85	62.76	61.48
		0.4	55.76	53.97	54.48	56.27	57.00	57.91	63.94	63.82	63.12
32 & 8	No EQ		45.45	45.45	45.45	87.88	87.88	87.88	15.15	15.15	15.15
	μ1	0.9	59.12	56.91	59.12	52.88	60.76	56.09	66.70	71.79	70.85
		0.8	64.24	60.42	62.42	59.24	63.09	62.03	69.70	70.55	70.73
		0.6	68.33	62.21	63.64	70.30	71.24	71.00	72.48	72.42	72.24
		0.4	66.88	61.58	61.42	73.61	69.45	69.58	78.21	65.61	65.94
	μ2	0.9	59.18	59.33	60.36	56.15	67.67	63.88	64.33	69.39	70.03
		0.8	62.94	60.76	62.21	59.58	69.91	70.18	69.03	68.88	70.30
		0.6	64.15	64.88	65.21	66.91	71.18	71.73	72.39	67.64	68.70
		0.4	64.73	60.12	60.61	65.30	71.97	72.39	76.15	67.82	68.45
	μ3	0.9	58.52	56.18	57.39	64.39	23.48	24.88	63.03	34.55	34.73
		0.8	65.67	61.00	62.67	64.39	25.21	27.18	69.64	32.67	31.97
		0.6	66.67	63.36	64.03	71.82	32.00	32.33	74.27	33.52	34.48
		0.4	67.64	62.00	61.70	65.91	31.58	31.88	71.24	38.94	38.91
	μ4	0.9	33.52	49.70	47.03	46.70	76.76	63.91	44.33	75.15	70.48
		0.8	37.30	50.76	48.79	54.64	65.88	59.36	53.76	77.27	74.67
		0.6	41.67	48.94	47.91	59.21	59.67	57.15	67.52	76.09	75.15
		0.4	44.27	46.27	46.30	60.73	55.45	54.61	71.09	71.39	69.48
	μ5	0.9	45.70	55.97	57.97	35.73	58.03	51.55	57.09	69.18	66.39
		0.8	49.94	56.58	56.88	45.94	62.06	60.45	60.21	69.70	66.33
		0.6	55.33	54.94	55.39	57.42	65.97	67.24	59.42	60.33	59.45
		0.4	55.85	49.61	48.97	62.64	65.91	65.88	61.15	58.36	57.30

Item	Dist*	ρ	Subtest1			Subtest2			Subtest3		
			m1	m2	m3	m1	m2	m3	m1	m2	m3
	μ6	0.9	60.39	60.15	60.42	57.88	64.79	61.15	64.79	72.82	72.55
		0.8	66.33	62.88	64.03	66.27	68.88	67.06	71.06	73.97	75.39
		0.6	68.33	63.58	63.85	71.88	71.33	71.42	76.15	70.09	70.09
		0.4	67.58	60.61	60.36	76.88	69.61	69.85	80.94	70.52	70.85
	μ7	0.9	54.70	54.70	57.09	49.00	64.61	58.67	62.91	70.58	71.09
		0.8	60.64	58.79	60.24	57.12	64.03	62.27	65.76	71.06	70.82
		0.6	64.24	60.55	62.73	67.03	67.12	68.12	70.45	69.03	68.36
		0.4	63.79	57.70	57.55	71.55	69.85	69.15	72.27	65.27	65.79
16 & 4	No EQ		48.48	48.48	48.48	60.61	60.61	60.61	21.21	21.21	21.21
	μ1	0.9	55.82	59.00	60.18	49.33	48.36	52.12	55.18	67.73	63.52
		0.8	57.64	60.48	60.67	52.09	49.70	52.12	56.09	65.61	64.82
		0.6	61.64	60.70	61.48	58.58	55.00	56.15	64.76	64.33	64.91
		0.4	62.03	60.03	58.70	57.70	57.82	57.48	70.21	66.06	65.45
	μ2	0.9	51.06	61.76	63.24	50.64	48.67	51.24	53.88	64.06	61.55
		0.8	55.70	61.12	61.61	50.21	49.79	53.85	54.73	64.39	63.64
		0.6	56.58	62.67	62.36	51.21	54.45	55.18	58.03	64.03	65.03
		0.4	53.33	61.24	60.67	54.39	53.36	53.94	61.33	65.03	64.48
	μ3	0.9	55.91	59.70	62.27	48.94	38.76	39.91	54.24	33.79	33.42
		0.8	59.18	62.55	63.27	52.27	41.97	42.15	58.70	34.00	33.64
		0.6	60.52	59.18	59.24	52.67	44.39	44.36	64.09	39.21	39.79
		0.4	65.12	59.97	61.42	52.18	46.36	46.64	67.18	44.15	44.45
	μ4	0.9	37.24	48.70	41.39	36.30	53.70	48.85	38.55	65.45	50.09
		0.8	39.24	49.12	43.67	39.94	49.27	48.82	44.88	67.64	57.70
		0.6	41.88	45.76	44.15	42.76	47.73	47.48	56.21	73.21	65.76
		0.4	45.21	46.82	46.61	42.52	44.33	43.97	59.64	65.85	62.33
	μ5	0.9	31.42	50.12	45.88	36.24	45.39	45.61	45.97	56.85	51.70
		0.8	38.33	49.94	47.64	42.85	47.18	48.61	44.52	62.61	53.94
		0.6	42.97	50.45	49.09	47.00	51.76	50.33	48.03	58.36	52.06
		0.4	45.39	43.97	43.73	50.18	54.09	51.97	49.06	53.30	48.79
	μ6	0.9	62.12	63.52	62.64	50.24	49.67	53.70	53.91	67.33	63.39
		0.8	65.36	65.39	65.03	53.52	51.36	56.33	57.70	65.42	64.42
		0.6	65.45	64.15	64.24	57.30	56.09	56.94	68.61	67.30	67.06
		0.4	62.91	61.24	60.88	53.91	55.52	56.21	71.76	70.18	69.82
	μ7	0.9	50.67	56.33	56.30	46.64	48.15	50.03	52.33	63.06	59.52
		0.8	53.27	56.91	57.15	48.09	49.27	51.70	52.55	64.00	59.00
		0.6	56.88	57.30	57.94	53.91	54.52	54.94	61.85	62.76	61.48
		0.4	55.76	53.97	54.48	56.27	57.00	57.91	63.94	63.82	63.12

* Note: μ1 $= (0,0,0)$, μ2 $= (0.1,0.1,0.1)$, μ3 $= (0,0.1,-0.1)$, μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. m1 is the first method of using subtest anchor score as the anchor; m2 is the second method of using equated total score; and m3 is the third method of using anchor total score.

Table 3.28. Average Percentages of the Number of Score Points Equal between True and Estimated Conversion using Augmented Scores

Item	Dist*	ρ	Subtest1 m1	Subtest1 m2	Subtest1 m3	Subtest2 m1	Subtest2 m2	Subtest2 m3	Subtest3 m1	Subtest3 m2	Subtest3 m3
32	No EQ		48.48	48.48	48.48	60.61	60.61	60.61	21.21	21.21	21.21
&	μ1	0.9	67.24	68.03	68.76	58.30	58.79	59.70	73.79	78.45	77.67
4		0.8	64.45	65.91	65.91	59.42	57.58	58.94	74.48	78.55	78.82
		0.6	64.33	61.97	62.39	60.61	57.06	57.36	75.97	74.42	73.88
		0.4	62.85	60.15	58.64	58.94	57.82	57.85	75.48	71.88	71.52
	μ2	0.9	57.39	65.82	65.36	55.21	58.82	58.33	67.39	75.09	74.55
		0.8	56.91	63.06	64.33	55.45	55.21	56.67	66.61	74.30	73.48
		0.6	56.33	61.03	61.27	53.85	54.91	55.30	66.00	73.45	72.91
		0.4	51.58	59.88	59.30	54.27	54.30	54.97	66.79	71.42	71.36
	μ3	0.9	67.15	68.67	69.15	52.00	48.52	48.06	70.85	46.85	47.06
		0.8	64.85	67.45	67.48	50.64	47.58	46.76	71.15	48.91	47.94
		0.6	61.45	60.33	59.70	45.97	46.00	45.18	73.27	49.88	49.27
		0.4	64.76	60.21	60.55	54.85	48.24	47.88	73.85	51.91	51.27
	μ4	0.9	39.91	48.61	45.18	41.73	48.27	47.88	56.73	75.88	71.64
		0.8	42.03	48.58	46.64	42.30	45.52	47.03	57.39	73.82	69.09
		0.6	46.03	47.24	46.55	42.39	46.79	46.39	60.36	72.39	68.36
		0.4	45.94	46.18	46.06	41.58	43.61	43.03	60.24	66.18	63.70
	μ5	0.9	49.61	63.67	62.55	51.42	54.67	55.55	51.88	63.03	59.24
		0.8	47.76	55.48	54.45	52.76	55.27	56.48	52.15	63.36	59.27
		0.6	46.09	50.18	48.24	53.88	56.85	55.85	53.30	61.55	57.55
		0.4	46.88	44.76	43.91	53.61	56.64	55.24	53.00	55.79	53.21
	μ6	0.9	61.48	65.67	64.42	57.70	57.15	57.48	75.85	81.70	81.18
		0.8	64.76	66.79	66.48	57.48	56.33	58.30	77.67	80.12	80.45
		0.6	64.82	62.48	62.82	57.61	55.70	56.70	78.82	77.27	76.15
		0.4	63.27	62.03	61.45	54.36	55.61	55.97	77.55	75.15	74.79
	μ7	0.9	64.12	67.27	67.58	56.27	60.27	58.82	68.88	77.58	75.24
		0.8	62.42	64.88	65.48	54.55	57.18	55.61	67.03	75.73	75.00
		0.6	59.33	58.85	58.67	57.79	57.06	57.09	69.64	70.09	69.24
		0.4	57.24	55.33	54.85	57.76	58.36	57.82	70.15	70.24	69.48
32	No EQ		45.45	45.45	45.45	87.88	87.88	87.88	15.15	15.15	15.15
&	μ1	0.9	70.36	66.73	68.42	74.18	76.97	76.88	81.03	78.52	78.67
8		0.8	70.15	66.03	66.55	73.73	76.21	74.24	80.39	79.30	79.06
		0.6	70.42	64.48	64.61	76.33	75.52	74.76	81.12	77.36	77.73
		0.4	66.36	60.03	60.24	74.88	69.48	69.18	82.24	70.21	70.30
	μ2	0.9	63.18	66.15	67.03	62.73	72.91	72.61	73.12	76.97	77.27
		0.8	67.36	65.76	65.73	64.67	75.18	74.00	75.91	77.91	77.48
		0.6	64.91	65.45	64.94	65.45	69.06	69.91	76.91	74.06	74.09
		0.4	62.85	59.52	59.61	63.61	69.97	70.55	78.39	73.09	73.39
	μ3	0.9	67.91	65.82	66.18	64.52	38.12	37.91	76.82	43.15	43.79
		0.8	70.24	66.58	67.30	63.70	35.12	34.79	78.15	43.42	43.97
		0.6	67.64	64.88	64.45	65.73	34.67	33.91	78.64	45.24	45.21
		0.4	65.70	60.48	59.70	63.12	34.67	34.27	76.91	46.21	45.91
	μ4	0.9	40.09	47.61	46.94	56.91	67.79	65.42	65.36	80.64	81.52
		0.8	41.88	50.24	49.64	57.48	63.42	61.48	66.85	77.00	78.06
		0.6	44.91	49.79	49.64	59.64	60.06	59.67	70.24	74.91	74.52
		0.4	44.88	47.45	47.36	60.76	56.15	56.00	71.03	71.70	70.03
	μ5	0.9	59.91	65.76	66.12	54.21	62.48	64.52	56.12	64.45	64.30
		0.8	56.94	60.03	59.94	52.27	63.12	63.24	57.88	63.73	63.42
		0.6	57.03	56.15	55.21	57.82	64.15	63.73	58.30	59.30	58.15
		0.4	55.45	49.52	48.94	60.70	63.15	61.36	59.21	58.61	58.12

Item	Dist*	ρ	Subtest1			Subtest2			Subtest3		
			m1	m2	m3	m1	m2	m3	m1	m2	m3
	μ6	0.9	64.58	65.76	65.64	76.27	80.12	79.88	84.79	81.61	82.18
		0.8	66.94	65.48	65.91	76.12	77.70	76.00	84.06	82.61	82.79
		0.6	66.73	63.09	62.45	76.09	73.18	72.61	84.06	77.27	77.27
		0.4	66.55	60.12	60.61	75.21	68.67	69.15	85.39	74.61	75.15
	μ7	0.9	66.39	66.70	67.94	68.42	75.70	74.82	74.27	76.73	76.76
		0.8	67.09	64.36	64.97	67.36	74.64	72.82	74.67	78.12	77.30
		0.6	66.09	64.09	64.00	72.21	71.61	70.36	75.52	73.09	71.91
		0.4	64.06	57.03	56.48	73.03	70.06	69.21	75.09	69.73	69.24
16 & 4	No EQ		48.48	48.48	48.48	60.61	60.61	60.61	21.21	21.21	21.21
	μ1	0.9	67.24	68.03	68.76	58.30	58.79	59.70	73.79	78.45	77.67
		0.8	64.45	65.91	65.91	59.42	57.58	58.94	74.48	78.55	78.82
		0.6	64.33	61.97	62.39	60.61	57.06	57.36	75.97	74.42	73.88
		0.4	62.85	60.15	58.64	58.94	57.82	57.85	75.48	71.88	71.52
	μ2	0.9	57.39	65.82	65.36	55.21	58.82	58.33	67.39	75.09	74.55
		0.8	56.91	63.06	64.33	55.45	55.21	56.67	66.61	74.30	73.48
		0.6	56.33	61.03	61.27	53.85	54.91	55.30	66.00	73.45	72.91
		0.4	51.58	59.88	59.30	54.27	54.30	54.97	66.79	71.42	71.36
	μ3	0.9	67.15	68.67	69.15	52.00	48.52	48.06	70.85	46.85	47.06
		0.8	64.85	67.45	67.48	50.64	47.58	46.76	71.15	48.91	47.94
		0.6	61.45	60.33	59.70	45.97	46.00	45.18	73.27	49.88	49.27
		0.4	64.76	60.21	60.55	54.85	48.24	47.88	73.85	51.91	51.27
	μ4	0.9	39.91	48.61	45.18	41.73	48.27	47.88	56.73	75.88	71.64
		0.8	42.03	48.58	46.64	42.30	45.52	47.03	57.39	73.82	69.09
		0.6	46.03	47.24	46.55	42.39	46.79	46.39	60.36	72.39	68.36
		0.4	45.94	46.18	46.06	41.58	43.61	43.03	60.24	66.18	63.70
	μ5	0.9	49.61	63.67	62.55	51.42	54.67	55.55	51.88	63.03	59.24
		0.8	47.76	55.48	54.45	52.76	55.27	56.48	52.15	63.36	59.27
		0.6	46.09	50.18	48.24	53.88	56.85	55.85	53.30	61.55	57.55
		0.4	46.88	44.76	43.91	53.61	56.64	55.24	53.00	55.79	53.21
	μ6	0.9	61.48	65.67	64.42	57.70	57.15	57.48	75.85	81.70	81.18
		0.8	64.76	66.79	66.48	57.48	56.33	58.30	77.67	80.12	80.45
		0.6	64.82	62.48	62.82	57.61	55.70	56.70	78.82	77.27	76.15
		0.4	63.27	62.03	61.45	54.36	55.61	55.97	77.55	75.15	74.79
	μ7	0.9	64.12	67.27	67.58	56.27	60.27	58.82	68.88	77.58	75.24
		0.8	62.42	64.88	65.48	54.55	57.18	55.61	67.03	75.73	75.00
		0.6	59.33	58.85	58.67	57.79	57.06	57.09	69.64	70.09	69.24
		0.4	57.24	55.33	54.85	57.76	58.36	57.82	70.15	70.24	69.48

* Note: $\mu 1 = (0,0,0)$, $\mu 2 = (0.1,0.1,0.1)$, $\mu 3 = (0,0.1,-0.1)$, μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. m1 is the first method of using subtest anchor score as the anchor; m2 is the second method of using equated total score; and m3 is the third method of using anchor total score.

3.4.4.2 Difference that matters (DTM)

Table 3.29, Table 3.30, and Table 3.31 include average percentages of DTM which demonstrate similar patterns of the results presented in the previous sections (3.4.4.1). Values in these tables indicate percentages of score points out of the total score points where the difference between true and estimated conversions was less than 0.5. When the number of items was 32 including 8 common items, the percentages were higher than having less items or common items regardless of the correlations or proficiency distributions. When the proficiency distribution was shifted to $\mu 3=(0,0.1,-0.1)$, the anchor total and equated total methods outperformed the subtest anchor score method in subtests 2 and 3. On the other hand, under the skewed distribution with a skewness of .75 ($\mu 4$), the subtest anchor score method did not perform as well as the other two methods. In general, with high correlations, anchor total or equated total score methods are more likely to produce more accurate equating results, having higher percentages. As the correlation dropped to 0.4, the subtest anchor score method tends to yield better equating results.

In most cases, observed scores produced higher percentages of DTM than weighted averages; in some cases, however, weighted averages showed slightly higher percentages when the correlation was 0.4. In subtest 3, augmented scores showed better results than observed scores whereas observed scores performed slightly better than augmented scores in subtest 2.

Table 3.29. Percentages of DTM using Observed Scores

Item	Dist*	ρ	Subtest1			Subtest2			Subtest3		
			m1	m2	m3	m1	m2	m3	m1	m2	m3
32 & 4	No EQ		48.48	48.48	48.48	60.61	60.61	60.61	21.21	21.21	21.21
	μ1	0.9	75.30	75.58	77.55	79.88	83.00	82.88	75.52	81.76	80.42
		0.8	70.48	71.36	72.42	79.06	81.03	81.33	76.85	79.94	79.45
		0.6	73.30	71.33	71.58	76.94	78.30	79.48	79.00	75.97	76.21
		0.4	72.94	69.09	68.42	78.67	75.85	75.42	78.42	72.36	72.03
	μ2	0.9	70.18	74.24	75.76	75.64	83.76	83.48	73.94	82.09	81.97
		0.8	70.85	73.58	74.85	73.55	79.64	79.48	71.73	80.55	79.67
		0.6	71.03	72.70	73.48	73.70	74.48	74.70	70.94	75.94	76.67
		0.4	66.39	70.18	70.06	74.42	74.76	76.06	72.82	72.67	73.48
	μ3	0.9	73.70	75.61	77.03	75.30	49.82	50.00	69.45	32.64	32.42
		0.8	73.67	70.82	71.58	77.24	49.85	49.58	70.09	35.91	36.52
		0.6	70.82	69.45	69.67	75.00	48.21	47.45	72.09	38.42	38.79
		0.4	75.39	69.30	70.18	73.36	48.91	48.27	73.52	38.85	38.03
	μ4	0.9	45.73	59.03	57.00	50.67	71.67	66.79	50.97	79.48	72.76
		0.8	46.45	57.18	55.21	50.27	68.94	64.03	49.79	71.82	66.00
		0.6	46.91	54.06	53.33	51.52	62.70	59.42	52.30	68.03	61.91
		0.4	47.64	51.88	51.58	49.30	56.36	55.06	52.64	58.61	56.03
	μ5	0.9	58.33	69.27	69.12	63.27	74.12	74.30	59.18	72.48	68.61
		0.8	60.94	66.67	66.09	63.30	73.76	72.76	58.42	73.12	69.91
		0.6	60.88	64.03	64.55	66.15	71.39	69.73	60.64	69.27	66.00
		0.4	61.64	56.70	56.27	63.61	66.39	66.12	59.21	61.70	59.58
	μ6	0.9	71.64	76.15	76.85	74.67	82.91	82.12	75.91	84.03	83.45
		0.8	73.79	75.06	75.21	77.09	82.36	81.52	74.61	79.97	79.55
		0.6	72.79	71.97	73.03	75.67	79.06	78.64	78.61	76.36	75.39
		0.4	71.91	68.42	68.70	76.64	75.15	74.24	76.76	73.42	73.94
	μ7	0.9	69.12	71.36	73.12	76.82	81.76	82.45	73.48	81.73	81.24
		0.8	70.12	71.94	72.91	73.91	79.52	79.91	71.55	78.97	77.94
		0.6	70.15	69.52	69.64	75.42	76.70	77.55	74.18	75.09	74.21
		0.4	69.79	65.24	65.94	76.27	75.52	76.55	73.82	72.30	72.33
32 & 8	No EQ		45.45	45.45	45.45	87.88	87.88	87.88	15.15	15.15	15.15
	μ1	0.9	78.09	75.09	77.58	85.55	86.09	86.21	86.24	85.36	85.76
		0.8	78.82	75.21	76.45	85.03	85.06	84.33	84.24	83.85	84.58
		0.6	78.15	72.33	72.45	84.45	81.24	82.45	84.48	81.76	82.39
		0.4	76.52	69.85	70.52	85.61	76.76	77.12	85.42	73.52	72.67
	μ2	0.9	75.67	76.88	77.70	81.00	85.70	85.88	81.85	83.91	84.03
		0.8	78.00	73.52	74.52	82.06	86.39	86.85	83.18	82.67	82.42
		0.6	74.24	74.45	74.76	82.27	79.27	79.82	84.00	77.94	78.42
		0.4	74.48	69.58	70.42	82.70	76.55	77.67	84.79	73.42	74.91
	μ3	0.9	78.36	74.15	75.76	82.36	47.58	46.97	84.55	31.30	33.58
		0.8	79.30	74.39	75.85	82.61	47.15	46.70	83.30	34.88	36.00
		0.6	76.73	72.15	73.42	85.18	46.85	47.79	83.64	35.48	36.30
		0.4	76.58	69.03	69.67	82.30	47.42	47.94	82.12	36.61	36.58
	μ4	0.9	51.42	67.79	64.82	60.03	81.94	81.82	62.12	90.21	90.42
		0.8	51.58	67.42	65.55	60.64	80.55	79.36	63.15	85.94	85.30
		0.6	51.67	59.73	59.45	58.36	70.33	68.24	65.36	78.61	77.21
		0.4	51.27	54.39	54.33	59.64	62.52	61.52	64.94	70.09	67.48
	μ5	0.9	72.18	74.15	75.76	73.73	82.85	84.52	74.97	84.55	82.64
		0.8	71.30	70.70	71.61	74.21	83.67	84.00	75.39	80.12	78.67
		0.6	73.24	65.79	66.97	75.67	74.48	75.55	75.09	72.09	70.45
		0.4	73.09	61.88	61.33	77.27	71.82	72.42	73.21	68.15	66.52
	μ6	0.9	79.18	77.12	77.67	82.97	87.27	87.42	86.70	88.06	88.52

Item	Dist*	ρ	Subtest1			Subtest2			Subtest3		
			m1	m2	m3	m1	m2	m3	m1	m2	m3
		0.8	79.64	76.18	76.48	84.06	85.82	85.27	83.24	85.52	85.97
		0.6	77.24	72.21	72.39	82.39	80.39	79.55	84.30	80.06	80.27
		0.4	76.61	69.21	68.88	83.91	76.67	76.61	86.58	78.48	78.48
	μ7	0.9	75.27	71.76	74.33	85.91	85.70	86.39	83.36	85.42	85.58
		0.8	76.61	73.30	75.03	80.73	84.30	84.79	82.15	83.06	82.79
		0.6	77.52	70.18	72.55	83.12	77.48	78.79	83.03	77.33	78.06
		0.4	75.76	67.48	68.15	81.18	76.27	77.36	82.36	74.42	74.06
16 & 4	No EQ		48.48	48.48	48.48	60.61	60.61	60.61	21.21	21.21	21.21
	μ1	0.9	75.30	75.58	77.55	79.88	83.00	82.88	75.52	81.76	80.42
		0.8	70.48	71.36	72.42	79.06	81.03	81.33	76.85	79.94	79.45
		0.6	73.30	71.33	71.58	76.94	78.30	79.48	79.00	75.97	76.21
		0.4	72.94	69.09	68.42	78.67	75.85	75.42	78.42	72.36	72.03
	μ2	0.9	70.18	74.24	75.76	75.64	83.76	83.48	73.94	82.09	81.97
		0.8	70.85	73.58	74.85	73.55	79.64	79.48	71.73	80.55	79.67
		0.6	71.03	72.70	73.48	73.70	74.48	74.70	70.94	75.94	76.67
		0.4	66.39	70.18	70.06	74.42	74.76	76.06	72.82	72.67	73.48
	μ3	0.9	73.70	75.61	77.03	75.30	49.82	50.00	69.45	32.64	32.42
		0.8	73.67	70.82	71.58	77.24	49.85	49.58	70.09	35.91	36.52
		0.6	70.82	69.45	69.67	75.00	48.21	47.45	72.09	38.42	38.79
		0.4	75.39	69.30	70.18	73.36	48.91	48.27	73.52	38.85	38.03
	μ4	0.9	45.73	59.03	57.00	50.67	71.67	66.79	50.97	79.48	72.76
		0.8	46.45	57.18	55.21	50.27	68.94	64.03	49.79	71.82	66.00
		0.6	46.91	54.06	53.33	51.52	62.70	59.42	52.30	68.03	61.91
		0.4	47.64	51.88	51.58	49.30	56.36	55.06	52.64	58.61	56.03
	μ5	0.9	58.33	69.27	69.12	63.27	74.12	74.30	59.18	72.48	68.61
		0.8	60.94	66.67	66.09	63.30	73.76	72.76	58.42	73.12	69.91
		0.6	60.88	64.03	64.55	66.15	71.39	69.73	60.64	69.27	66.00
		0.4	61.64	56.70	56.27	63.61	66.39	66.12	59.21	61.70	59.58
	μ6	0.9	71.64	76.15	76.85	74.67	82.91	82.12	75.91	84.03	83.45
		0.8	73.79	75.06	75.21	77.09	82.36	81.52	74.61	79.97	79.55
		0.6	72.79	71.97	73.03	75.67	79.06	78.64	78.61	76.36	75.39
		0.4	71.91	68.42	68.70	76.64	75.15	74.24	76.76	73.42	73.94
	μ7	0.9	69.12	71.36	73.12	76.82	81.76	82.45	73.48	81.73	81.24
		0.8	70.12	71.94	72.91	73.91	79.52	79.91	71.55	78.97	77.94
		0.6	70.15	69.52	69.64	75.42	76.70	77.55	74.18	75.09	74.21
		0.4	69.79	65.24	65.94	76.27	75.52	76.55	73.82	72.30	72.33

* Note: $\mu 1 = (0,0,0)$, $\mu 2 = (0.1, 0.1, 0.1)$, $\mu 3 = (0, 0.1, -0.1)$, μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. m1 is the first method of using subtest anchor score as the anchor; m2 is the second method of using equated total score; and m3 is the third method of using anchor total score.

Table 3.30. Percentages of DTM using Weighted Averages

Item	Dist*	ρ	Subtest1			Subtest2			Subtest3		
			m1	m2	m3	m1	m2	m3	m1	m2	m3
32 & 4	No EQ		48.48	48.48	48.48	60.61	60.61	60.61	21.21	21.21	21.21
	μ1	0.9	63.03	66.39	68.15	60.52	70.00	65.39	61.79	75.15	70.58
		0.8	66.67	66.91	67.91	65.97	72.39	70.97	62.91	73.12	70.21
		0.6	71.82	71.06	70.58	74.03	75.76	76.18	71.88	72.88	72.55
		0.4	71.73	70.21	69.70	76.58	77.18	76.45	77.91	71.94	71.76
	μ2	0.9	59.36	65.85	68.67	56.18	64.91	60.36	57.97	72.18	67.12
		0.8	65.03	67.09	69.18	64.36	70.79	69.76	61.73	73.18	69.82
		0.6	70.39	71.94	72.73	69.67	73.48	73.33	65.88	73.61	73.61
		0.4	67.06	70.00	69.06	72.09	74.55	75.79	69.82	71.91	71.94
	μ3	0.9	62.15	65.00	69.21	59.00	44.88	43.09	57.24	28.91	30.18
		0.8	66.61	68.91	70.09	66.64	46.00	44.85	61.97	30.27	31.76
		0.6	70.27	69.64	69.55	73.06	45.76	45.55	69.36	35.27	36.55
		0.4	76.03	68.21	68.85	69.79	48.79	48.73	71.85	39.36	40.45
	μ4	0.9	35.94	48.06	44.12	36.64	62.39	46.30	37.76	62.33	46.48
		0.8	39.06	49.52	47.15	41.15	61.06	51.58	42.33	63.64	52.73
		0.6	44.12	51.48	50.42	47.94	59.70	55.85	51.00	69.48	61.42
		0.4	48.12	50.97	51.18	49.39	57.00	55.85	53.88	60.61	57.06
	μ5	0.9	40.39	62.91	59.39	38.52	62.61	50.85	52.30	65.45	60.70
		0.8	50.39	63.15	60.64	52.48	64.88	63.27	52.97	71.70	65.79
		0.6	57.18	62.61	61.97	64.06	69.24	68.45	57.94	68.00	64.42
		0.4	61.36	56.88	56.67	65.88	68.45	68.61	60.39	62.73	60.30
	μ6	0.9	65.73	67.09	68.94	57.91	69.52	64.36	54.48	73.73	66.97
		0.8	71.85	70.42	71.52	67.39	74.79	74.61	60.36	72.09	68.45
		0.6	72.27	71.97	73.03	73.79	77.42	76.30	73.61	72.97	71.73
		0.4	70.94	68.94	69.30	73.76	73.12	72.61	75.36	72.36	72.67
	μ7	0.9	58.55	62.79	65.09	53.33	67.76	62.00	59.09	73.03	67.12
		0.8	63.82	66.76	66.03	61.45	70.58	70.52	60.94	72.45	67.15
		0.6	68.30	68.88	68.82	71.88	74.79	75.94	71.00	72.03	71.79
		0.4	69.33	66.15	66.30	74.76	76.12	76.15	74.03	72.64	72.21
32 & 8	No EQ		45.45	45.45	45.45	87.88	87.88	87.88	15.15	15.15	15.15
	μ1	0.9	67.45	64.67	67.00	65.73	72.36	70.06	66.82	75.52	73.30
		0.8	74.06	70.27	72.79	70.45	75.79	76.42	72.73	76.73	74.70
		0.6	78.12	73.03	74.03	80.21	80.82	81.06	79.24	78.88	78.33
		0.4	77.70	72.48	72.21	82.91	78.09	78.94	85.06	72.85	73.21
	μ2	0.9	67.79	66.73	68.91	65.52	70.33	68.61	65.27	75.52	73.64
		0.8	71.91	69.39	70.82	71.76	75.82	75.76	72.15	75.39	75.12
		0.6	73.36	73.97	74.21	80.09	77.15	77.48	78.36	75.91	75.88
		0.4	74.97	69.61	69.70	80.33	75.70	76.27	83.61	74.36	74.82
	μ3	0.9	67.03	63.48	65.67	66.79	38.33	36.21	62.55	31.88	34.39
		0.8	75.30	70.76	72.91	71.00	38.94	40.06	73.52	30.64	32.48
		0.6	76.45	73.09	74.30	81.79	47.18	47.39	80.06	32.73	33.73
		0.4	77.61	71.45	71.79	81.42	48.48	49.03	80.00	38.09	38.73
	μ4	0.9	38.64	54.33	52.15	42.61	79.61	67.15	41.91	81.64	69.67
		0.8	42.45	57.24	55.61	48.36	73.52	67.18	49.12	81.21	74.73
		0.6	48.33	56.00	55.39	54.94	68.18	65.39	62.42	79.24	76.03
		0.4	51.24	54.27	54.52	58.06	62.91	62.33	65.58	72.70	69.85
	μ5	0.9	54.03	66.88	68.06	49.24	69.45	65.27	62.18	74.15	69.61
		0.8	61.85	67.82	68.00	63.15	74.94	75.67	66.33	76.85	73.06
		0.6	69.15	65.97	67.27	72.91	75.70	76.48	71.91	72.30	70.24
		0.4	69.52	59.45	58.76	76.48	73.88	73.67	73.33	69.52	67.58
	μ6	0.9	69.21	66.52	67.61	65.06	73.67	72.27	63.97	78.91	75.48

Item	Dist*	ρ	Subtest1			Subtest2			Subtest3		
			m1	m2	m3	m1	m2	m3	m1	m2	m3
		0.8	74.61	72.09	73.36	74.27	80.79	80.21	73.21	79.91	78.64
		0.6	78.61	73.67	74.09	78.48	80.36	80.58	80.18	77.39	77.21
		0.4	77.45	69.94	69.36	81.55	77.27	77.70	85.30	77.52	78.27
	μ7	0.9	65.67	62.33	65.33	63.18	74.45	71.76	64.48	75.24	71.97
		0.8	71.70	68.03	69.82	70.00	75.64	75.97	71.06	77.45	75.30
		0.6	76.21	71.12	73.00	77.30	76.82	78.15	77.97	75.15	76.00
		0.4	75.15	67.61	67.82	81.27	78.21	78.15	81.79	74.45	74.42
16 & 4	No EQ		48.48	48.48	48.48	60.61	60.61	60.61	21.21	21.21	21.21
	μ1	0.9	63.03	66.39	68.15	60.52	70.00	65.39	61.79	75.15	70.58
		0.8	66.67	66.91	67.91	65.97	72.39	70.97	62.91	73.12	70.21
		0.6	71.82	71.06	70.58	74.03	75.76	76.18	71.88	72.88	72.55
		0.4	71.73	70.21	69.70	76.58	77.18	76.45	77.91	71.94	71.76
	μ2	0.9	59.36	65.85	68.67	56.18	64.91	60.36	57.97	72.18	67.12
		0.8	65.03	67.09	69.18	64.36	70.79	69.76	61.73	73.18	69.82
		0.6	70.39	71.94	72.73	69.67	73.48	73.33	65.88	73.61	73.61
		0.4	67.06	70.00	69.06	72.09	74.55	75.79	69.82	71.91	71.94
	μ3	0.9	62.15	65.00	69.21	59.00	44.88	43.09	57.24	28.91	30.18
		0.8	66.61	68.91	70.09	66.64	46.00	44.85	61.97	30.27	31.76
		0.6	70.27	69.64	69.55	73.06	45.76	45.55	69.36	35.27	36.55
		0.4	76.03	68.21	68.85	69.79	48.79	48.73	71.85	39.36	40.45
	μ4	0.9	35.94	48.06	44.12	36.64	62.39	46.30	37.76	62.33	46.48
		0.8	39.06	49.52	47.15	41.15	61.06	51.58	42.33	63.64	52.73
		0.6	44.12	51.48	50.42	47.94	59.70	55.85	51.00	69.48	61.42
		0.4	48.12	50.97	51.18	49.39	57.00	55.85	53.88	60.61	57.06
	μ5	0.9	40.39	62.91	59.39	38.52	62.61	50.85	52.30	65.45	60.70
		0.8	50.39	63.15	60.64	52.48	64.88	63.27	52.97	71.70	65.79
		0.6	57.18	62.61	61.97	64.06	69.24	68.45	57.94	68.00	64.42
		0.4	61.36	56.88	56.67	65.88	68.45	68.61	60.39	62.73	60.30
	μ6	0.9	65.73	67.09	68.94	57.91	69.52	64.36	54.48	73.73	66.97
		0.8	71.85	70.42	71.52	67.39	74.79	74.61	60.36	72.09	68.45
		0.6	72.27	71.97	73.03	73.79	77.42	76.30	73.61	72.97	71.73
		0.4	70.94	68.94	69.30	73.76	73.12	72.61	75.36	72.36	72.67
	μ7	0.9	58.55	62.79	65.09	53.33	67.76	62.00	59.09	73.03	67.12
		0.8	63.82	66.76	66.03	61.45	70.58	70.52	60.94	72.45	67.15
		0.6	68.30	68.88	68.82	71.88	74.79	75.94	71.00	72.03	71.79
		0.4	69.33	66.15	66.30	74.76	76.12	76.15	74.03	72.64	72.21

* Note: $\mu1 = (0,0,0)$, $\mu2 = (0.1,0.1,0.1)$, $\mu3 = (0,0.1,-0.1)$, μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. m1 is the first method of using subtest anchor score as the anchor; m2 is the second method of using equated total score; and m3 is the third method of using anchor total score.

Table 3.31. Percentages of DTM using Augmented Scores

Item	Dist*	ρ	Subtest1			Subtest2			Subtest3		
			m1	m2	m3	m1	m2	m3	m1	m2	m3
32	No EQ		48.48	48.48	48.48	60.61	60.61	60.61	21.21	21.21	21.21
&	μ1	0.9	75.24	76.45	76.97	77.76	79.36	79.00	80.64	86.58	85.09
4		0.8	73.18	73.91	74.48	77.61	81.39	81.30	81.85	85.64	85.09
		0.6	74.97	72.91	72.39	77.15	78.70	78.61	83.18	80.97	80.91
		0.4	72.06	70.39	68.12	78.00	76.52	75.91	83.45	78.27	77.27
	μ2	0.9	67.70	73.45	73.03	73.21	79.27	78.55	74.79	85.79	85.45
		0.8	72.18	74.12	74.79	73.79	79.91	79.33	74.27	84.67	83.91
		0.6	72.24	72.64	72.82	72.79	74.39	74.36	74.52	81.27	81.33
		0.4	65.52	69.24	68.12	70.91	73.70	74.18	75.30	77.12	76.94
	μ3	0.9	74.06	74.70	75.88	71.33	57.76	58.18	71.97	41.21	41.52
		0.8	74.30	74.18	74.79	73.03	55.18	54.94	72.91	44.24	45.03
		0.6	72.18	71.67	71.45	71.06	51.45	50.88	75.12	46.03	46.61
		0.4	75.42	69.70	69.03	70.94	50.85	50.06	77.58	47.36	47.52
	μ4	0.9	42.30	50.61	50.09	47.00	65.00	59.09	50.94	73.33	66.55
		0.8	45.06	52.15	51.82	49.45	62.48	59.18	51.21	68.76	63.42
		0.6	48.64	53.82	54.00	52.15	61.76	59.12	53.70	68.85	63.42
		0.4	47.73	50.91	50.70	50.55	57.24	56.33	54.76	59.94	57.82
	μ5	0.9	63.58	74.21	73.52	59.39	65.61	67.79	59.36	72.03	69.64
		0.8	64.00	67.33	66.79	62.06	68.61	70.03	59.06	70.82	70.42
		0.6	58.91	61.85	61.30	66.24	68.48	68.15	61.09	68.67	66.33
		0.4	59.73	55.82	55.82	65.67	68.18	67.06	61.45	62.91	61.82
	μ6	0.9	66.73	71.42	71.42	73.52	80.88	80.15	79.39	87.61	86.64
		0.8	71.30	73.94	73.39	75.61	80.09	79.85	80.42	85.06	84.09
		0.6	71.36	72.15	72.30	74.52	78.21	77.36	83.18	82.09	80.70
		0.4	70.88	68.64	68.27	74.39	73.21	72.61	80.85	76.94	76.73
	μ7	0.9	72.42	75.09	76.36	70.82	76.67	76.36	78.58	86.64	85.21
		0.8	73.52	74.12	74.67	72.48	78.79	78.55	76.00	84.45	83.91
		0.6	72.58	71.15	70.79	75.55	78.42	78.58	77.00	77.85	77.30
		0.4	70.52	66.85	66.48	76.06	76.55	75.79	78.24	78.36	78.06
32	No EQ		45.45	45.45	45.45	87.88	87.88	87.88	15.15	15.15	15.15
&	μ1	0.9	79.12	76.79	78.55	82.79	86.06	85.06	89.82	90.42	89.58
8		0.8	80.91	77.73	78.15	81.55	86.12	85.36	87.76	88.64	88.24
		0.6	81.45	74.76	75.15	83.85	83.36	83.06	88.06	86.45	86.27
		0.4	77.76	72.03	71.85	83.70	77.18	77.12	89.76	78.27	78.42
	μ2	0.9	73.64	76.88	76.58	81.61	86.36	85.79	82.91	89.12	89.33
		0.8	79.09	75.76	76.39	81.67	85.30	85.30	86.33	89.30	88.91
		0.6	75.58	76.52	76.24	81.42	78.30	79.06	86.58	83.39	82.73
		0.4	74.21	69.70	69.88	81.24	77.18	77.18	87.30	80.42	81.03
	μ3	0.9	77.58	74.52	75.33	78.21	58.06	56.79	84.94	37.94	39.18
		0.8	80.33	77.61	78.18	79.03	56.06	54.85	86.12	42.76	43.79
		0.6	78.61	75.61	75.79	80.42	52.36	52.03	86.36	44.70	44.82
		0.4	78.09	71.45	71.67	80.88	52.58	52.58	86.18	45.61	46.06
	μ4	0.9	45.52	55.12	55.06	53.48	74.15	71.91	60.30	86.24	84.03
		0.8	48.36	59.00	58.61	56.97	72.58	71.70	62.03	83.00	82.24
		0.6	51.88	58.00	58.03	57.15	68.18	66.94	66.06	77.85	76.36
		0.4	51.09	53.70	53.73	58.12	62.85	62.27	65.73	71.06	69.48
	μ5	0.9	73.61	78.85	78.91	69.12	73.76	76.30	73.06	81.06	81.91
		0.8	70.45	71.55	71.76	71.61	76.45	77.55	74.24	79.73	80.55
		0.6	69.79	65.15	65.94	74.73	74.64	75.39	75.03	72.18	71.91
		0.4	67.00	58.06	57.48	75.09	73.18	72.70	74.58	71.27	70.76
	μ6	0.9	72.94	74.55	74.21	82.61	86.09	86.30	89.33	92.67	92.67

Item	Dist*	ρ	Subtest1			Subtest2			Subtest3		
			m1	m2	m3	m1	m2	m3	m1	m2	m3
		0.8	77.03	75.79	76.27	81.91	86.76	86.58	87.79	91.12	90.61
		0.6	77.79	73.79	73.18	81.39	81.39	79.70	88.64	85.39	85.58
		0.4	77.64	70.09	69.58	81.36	76.39	76.52	91.03	82.52	82.70
	μ7	0.9	78.15	77.76	78.76	78.91	84.03	83.91	85.06	88.97	89.39
		0.8	78.58	76.42	77.45	79.27	84.58	84.45	85.21	87.94	87.85
		0.6	79.52	74.76	75.91	81.94	80.18	79.70	86.82	82.58	83.85
		0.4	75.82	66.94	66.58	81.61	78.97	77.88	86.12	79.94	79.76
16 & 4	No EQ		48.48	48.48	48.48	60.61	60.61	60.61	21.21	21.21	21.21
	μ1	0.9	75.24	76.45	76.97	77.76	79.36	79.00	80.64	86.58	85.09
		0.8	73.18	73.91	74.48	77.61	81.39	81.30	81.85	85.64	85.09
		0.6	74.97	72.91	72.39	77.15	78.70	78.61	83.18	80.97	80.91
		0.4	72.06	70.39	68.12	78.00	76.52	75.91	83.45	78.27	77.27
	μ2	0.9	67.70	73.45	73.03	73.21	79.27	78.55	74.79	85.79	85.45
		0.8	72.18	74.12	74.79	73.79	79.91	79.33	74.27	84.67	83.91
		0.6	72.24	72.64	72.82	72.79	74.39	74.36	74.52	81.27	81.33
		0.4	65.52	69.24	68.12	70.91	73.70	74.18	75.30	77.12	76.94
	μ3	0.9	74.06	74.70	75.88	71.33	57.76	58.18	71.97	41.21	41.52
		0.8	74.30	74.18	74.79	73.03	55.18	54.94	72.91	44.24	45.03
		0.6	72.18	71.67	71.45	71.06	51.45	50.88	75.12	46.03	46.61
		0.4	75.42	69.70	69.03	70.94	50.85	50.06	77.58	47.36	47.52
	μ4	0.9	42.30	50.61	50.09	47.00	65.00	59.09	50.94	73.33	66.55
		0.8	45.06	52.15	51.82	49.45	62.48	59.18	51.21	68.76	63.42
		0.6	48.64	53.82	54.00	52.15	61.76	59.12	53.70	68.85	63.42
		0.4	47.73	50.91	50.70	50.55	57.24	56.33	54.76	59.94	57.82
	μ5	0.9	63.58	74.21	73.52	59.39	65.61	67.79	59.36	72.03	69.64
		0.8	64.00	67.33	66.79	62.06	68.61	70.03	59.06	70.82	70.42
		0.6	58.91	61.85	61.30	66.24	68.48	68.15	61.09	68.67	66.33
		0.4	59.73	55.82	55.82	65.67	68.18	67.06	61.45	62.91	61.82
	μ6	0.9	66.73	71.42	71.42	73.52	80.88	80.15	79.39	87.61	86.64
		0.8	71.30	73.94	73.39	75.61	80.09	79.85	80.42	85.06	84.09
		0.6	71.36	72.15	72.30	74.52	78.21	77.36	83.18	82.09	80.70
		0.4	70.88	68.64	68.27	74.39	73.21	72.61	80.85	76.94	76.73
	μ7	0.9	72.42	75.09	76.36	70.82	76.67	76.36	78.58	86.64	85.21
		0.8	73.52	74.12	74.67	72.48	78.79	78.55	76.00	84.45	83.91
		0.6	72.58	71.15	70.79	75.55	78.42	78.58	77.00	77.85	77.30
		0.4	70.52	66.85	66.48	76.06	76.55	75.79	78.24	78.36	78.06

* Note: $\mu 1 = (0,0,0)$, $\mu 2 = (0.1,0.1,0.1)$, $\mu 3 = (0,0.1,-0.1)$, μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. m1 is the first method of using subtest anchor score as the anchor; m2 is the second method of using equated total score; and m3 is the third method of using anchor total score.

3.5 Discussion

Subtest score equating results using observed scores, weighted averages, and augmented scores were evaluated and compared based on bias, absolute difference, RMSE, and number of score points equal between true and estimated conversion tables. The findings from this study are as follows:

First, the values of PRMSE were used to examine whether subtest scores or weighted averages have added value. With higher correlations such as 0.9, not all cases of subtest scores proved to have added value. However, with moderate and low correlations such as 0.6 and 0.4, subtest scores provided added value based on the PRMSE values. Multidimensional scaling was also applied as a preliminary study to examine dimensionality prior to equating. R-squared values and stress values as well as a visual inspection helped investigate the structure of the data. As correlation decreased, it was easier to see clusters for each subtest. These preliminary results supported the conclusion that with moderate correlations, it is worthwhile considering reporting subtest scores and doing statistical analyses based on subtests.

Second, when weighted averages were used in subtest score equating and despite being shown to have added value, observed scores and augmented scores produced smaller bias, absolute differences, and RMSEs than weighted averages. Augmented scores performed better than observed scores in terms of having smaller values for the equating error. Because weighted averages considered only the relationship between the subtest that is of interest and the total test when computing weights for each subtest, it was possible that the weights did not facilitate accomplishing effective equating with weighted averages. On the other hand, augmented scores used different weights which account for the relationship between all subtests and the total test and produced better equating results than observed scores and weighted averages.

Third, as correlations among subtests became smaller, using the subtest anchor score (Method 1) outperformed the other two methods regardless of the proficiency distributions, number of items and common items, and subtests. Although correlations were high such as 0.9 or 0.8, having more items and common items tends to work better with Method 1 than Methods 2 or 3 in several cases. This implies that with sufficient number of items and common items, using only subtest anchor scores as the anchor could provide relatively accurate equating results.

Fourth, when the proficiency distribution of one group was skewed, equating results were less accurate compared to when the two distributions were normal and equal to a mean vector of 0s. In most cases, the distribution with the relatively large skewness (.75) produced the largest bias, absolute difference, and RMSE. The condition with a skewness of .25 did not affect the accuracy of equating results very much. However, when the mean of the distributions was shifted to different directions for each dimension, shifting the mean of the distributions yielded relatively large equating errors compared to other conditions of the proficiency distribution.

The current study examined possible ways to equate subtest scores when correlations among dimensions varied. This study could provide a guideline for tests applying subtest score equating, especially for determining appropriate anchor sets depending on the correlations among subtests. However, several limitations should be noted in relation to the area of future research. First, the current study considered a simple structure model to generate item responses; each item was loaded on only one dimension, which is rarely practical. Future research needs to be conducted incorporating more complex structures such as a bifactor model and multidimensional compensatory model. Second, three subtests used in this study had similar mean scores and item difficulties resulting in not very effective equating. It would be worthwhile considering different levels of mean shifts and item difficulty in the two test forms (FormX and FormY) and examining how they affect subtest score equating. Third, the current study employed only one equating method—chained equipercentile equating. Future research may consider using other equating methods and compare them.

CHAPTER 4

COMPARING SUBTEST SCORE EQUATING METHODS UNDER THE ITEM RESPONSE THEORY FRAMEWORK

4.1 Introduction

The previous chapter focused on subtest score equating from the classical test theory perspective. The focal point of this chapter is to investigate possible subtest score equating methods from the item response theory framework. Item response theory has been widely used to deal with a variety of measurement problems in educational and psychological testing such as equating and scaling. Unlike traditional equating methods, IRT equating involves item parameter estimation and linking. Thus, different estimation and linking methods affect the results of equating.

As aforementioned, equating has been usually done at the total test level applying unidimensional IRT models. In reality, however, the total test may not be close to unidimensional although each subtest within the total test is unidimensional. Multidimensional IRT models can be more appropriate if the total test measures more than one trait or factor. Because of the estimation time and complexity of MIRT models, equating has not been done with MIRT models. In addition, there is no available software that produces raw-to-scale score or raw-to-raw conversion tables from MIRT equating. Alternatively, this study adopted unidimensional approximation to transform MIRT parameters to unidimensional IRT parameters. Unidimensional IRT equating was followed after parameter estimation and transformation. According to Brossman (2010), unidimensional approximation methods produced very similar results to full MIRT equating. This study applied IRT observed score equating from a unidimensional IRT model as well as a multidimensional IRT model applying unidimensional approximation and compared equating results under several simulation conditions including correlations among dimensions, test length, and different proficiency distributions.

In addition, two different approaches were adopted prior to equating: concurrent calibration and separate calibration with linking. Before equating is performed, all item parameters from different forms or different calibration runs need to be on the same scale. When the concurrent calibration approach is used, linking is not required because all item parameters are estimated at once and they are automatically placed on the same scale. In separate calibration, however, item parameters from different calibration runs are not on the same scale because of the scale indeterminacy. Linking methods are applied to resolve this issue although it is possible that there is a certain amount of linking errors. Concurrent calibration does not introduce linking errors, but scale contamination could be an issue if each form or each calibration has a unique trait. This study also compared these two calibration approaches.

For linking methods, both unidimensional and multidimensional linking were implemented. Yao (2011) compared three MIRT linking methods in a simulation study. Conditions included in her study were correlations between domains and different ability distributions. The author found that higher correlations between domains produced better item parameter recovery results. She also used several population distributions based on multivariate normal distribution and multivariate *t* distribution and examined the accuracy and effects of MIRT linking methods under different population types. The author concluded that population distribution had an impact on the parameter recovery. Populations with mean values closer to zero showed better item parameter recovery whereas populations including extreme values did not perform as well as the other populations. These conditions—correlations among dimensions and proficiency distributions—are also examined in this chapter.

4.2 Purpose of the Study

Unidimensional and multidimensional IRT based methods were applied and compared with each other and within each approach. Item parameters were first estimated based on the unidimensional and multidimensional IRT models. Two different approaches within each IRT model were adopted for IRT parameter calibration: separate and concurrent calibrations. Under the unidimensional approach, item parameters were estimated at the subtest level, which does not take into account the correlation among dimensions, and at the total test level, all subtests were treated as one test. When item parameters were separately estimated for each subtest, item parameter estimates from a new form needed to be placed on the scale of a reference form via unidimensional IRT linking methods prior to equating. On the other hand, when item parameters from both forms were calibrated concurrently, no linking procedure was required.

For IRT equating using multidimensional IRT parameters, all item parameters from different subtests were estimated at once based on a multidimensional IRT model. Within the multidiemsional IRT approach, both the separate and concurrent calibrations were also applied. After the separate calibration was performed, multidimensional linking methods were implemented to place item parameter estimates from the new form onto the old form scale. All these item parameter estimates were then transformed to unidimensional IRT item parameter estimates via the unidimensional approximation method. Using these transformed unidimensional IRT parameters, IRT equating was conducted. The same procedure was applied to item parameters from the concurrent calibration but without linking. Specific research questions are as follows:

1. Which method produces the most accurate equating results between unidimensional and multidimensional IRT based methods?
2. Which method—separate calibration with linking or concurrent calibration without linking—yields less equating errors?
3. When the number of items is small, does equating at the total test level outperform equating at the subtest level?
4. Which method produces better equating results when correlations among dimensions are high, moderate, or low?
5. When ability distributions differ, which method performs the best among unidimensional IRT equating at the total test level, unidimensional IRT equating at the subtest level, or equating using MIRT parameters with the unidimensional approximation method?

4.3 Methodology

4.3.1 Data

Item response data sets generated in Chapter 3 were also used in this study. Data generation procedures are described in Section 3.3.1. Data sets were reorganized and prepared for calibration in order to estimate item parameters using different test levels (total test level and subtest level), IRT models, and calibration methods. For example, when a unidimensional IRT model and separate calibration at the subtest level were used, six different data sets were created: three subtests for FormX and three for FormY. For a unidimensional IRT model and concurrent calibration at the subtest level, one data set including both FormX and FormY responses per each subtest or a total of three data sets, was created. Data for multidimensional IRT cases and unidimensional IRT at the total test level included all three subtests together—two data sets for separate calibration, one for FormX and one for FormY, and one file for the concurrent calibration. A configuration of data files is presented in Figure 4.1.

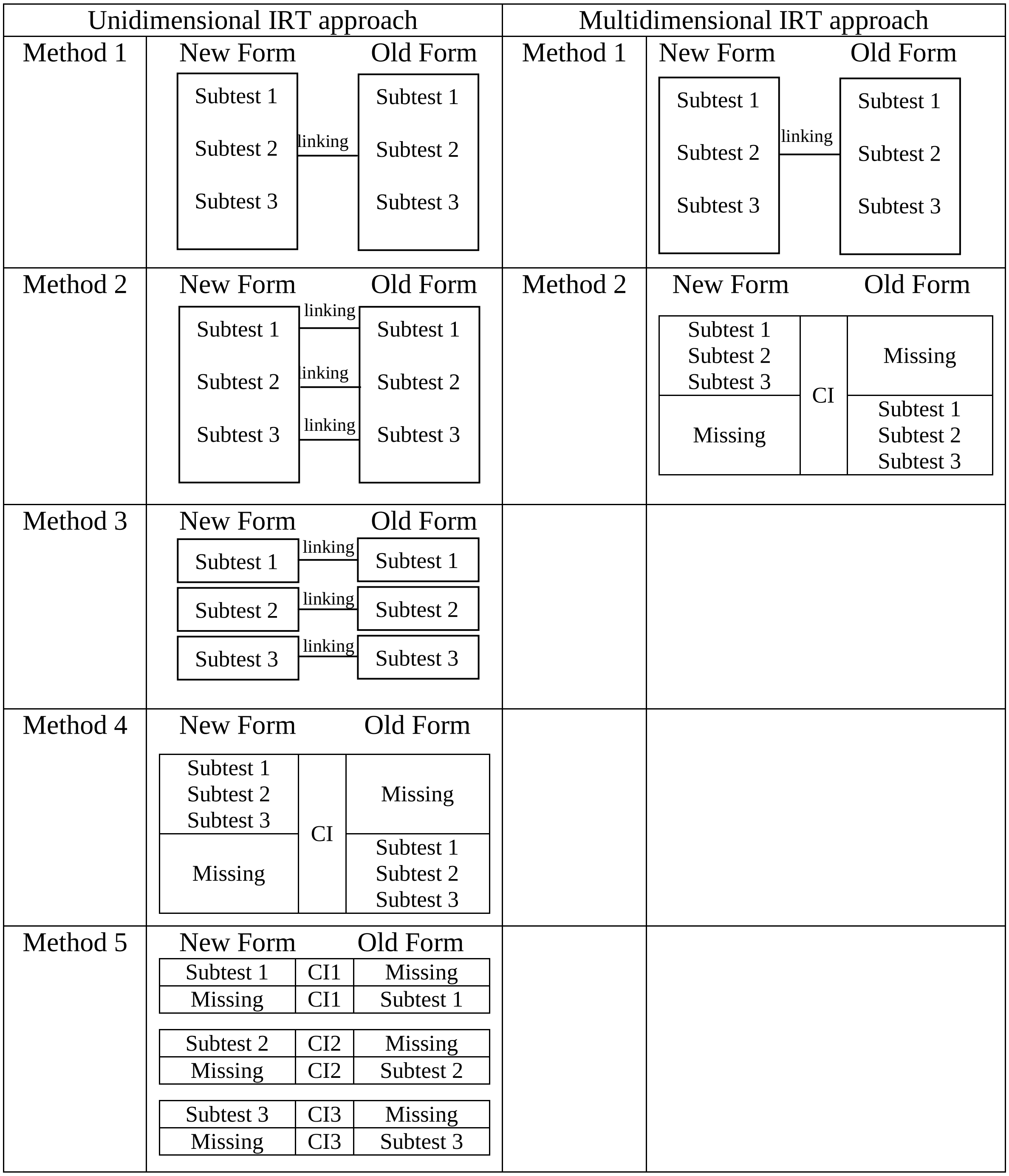


Figure 4.1. Overview of Subtest Score Equating Methods under the IRT framework

4.3.2 Simulation Conditions

Identical manipulation factors that were introduced in the previous chapter remained in this study. Three factors were considered: 1) correlations (0.4, 0.6, 0.8, and 0.9), 2) test length (32 items including 4 common items, 32 items including 8 common items, and 16 items including 4 common items), and 3) proficiency distributions (normal distributions—$\mu_1 = (0, 0, 0)$, $\mu_2 = (0.1, 0.1, 0.1)$, and $\mu_3 = (0, 0.1, -0.1)$ —and skewed distributions—μ_4, μ_5, μ_6, and μ_7 with skewness of .75, -.75, .25, and -.25). Possible combinations of study conditions were presented in Table 3.4.

4.3.3 Analysis

4.3.3.1 Item parameter estimation

Item parameters were estimated based on either a unidimensional or a multidimensional IRT model. Figure 4.1 illustrates methods applied to IRT equating. First, item parameters were estimated from both unidimensional and multidimensional 3PL models using flexMIRT (Cai, 2012). In the first and the second methods from the unidimensional approach, item parameters were calibrated at the total test level; thus, only two calibration runs were performed—one for FormX and another one for FormY. The first method used one linking constant to place all the item parameters on the same scale whereas the second method used three different linking constants using item parameters from each subtest although they were calibrated together in each form. The third method under the unidimensional approach involved separate items parameter calibration in each subtest. In this case, six different calibration runs were conducted—three for FormX and another three for FormY. The fourth and fifth methods used concurrent calibration. In the fourth method, item parameters from each subtest were estimated separately, but parameters from both forms were calibrated simultaneously. For instance, 1,000 examinees in Group A took FormY consisting of 32 items in a subtest, and another 1,000 examinees in Group B took FormX with the same number of items. Group A examinees did not respond to items appearing in FormX; those cases were treated as missing. The same logic applies to Group B examinees. Likewise, three separate runs of item calibration were conducted for the fourth method. In the fifth method, only one run was performed because all items in the three subtests for both forms were calibrated at once.

Using a multidimensional IRT model, all item parameters from the three subtests were estimated at once—one run for FormX and another run for FormY in the first method. The second method from the multidimensional model required only one run since it used concurrent calibration.

4.3.3.2 Item parameter recovery

As true item parameters based on a simple structure MIRT model were already known, estimated item parameters were compared to true item parameters. To evaluate the accuracy of item parameter recovery, correlations between true and estimated parameters from different methods described in the previous section were calculated. To compute more accurate correlations, Fisher's *r*-to-*z* transformation method was applied. According to Silver and Dunlap (1987), average *z* back transformed to *r* is less biased. Correlations of each replication were first transformed via Fisher's *r*-to-*z* transformation. The average *z* scores were then transformed to *r*.

Because this study adopted a simple structure model, *a* parameters were loaded on only one dimension. Thus, it was possible to compute correlations of *a* parameters from a unidimensional model and *a* parameters obtained from MIRT (*a*1, *a*2, and *a*3). For instance, *a* parameters from subtest 1, which were estimated from a unidimensional 3PL model, were compared to true *a* parameters loaded on dimension 1 (*a*1) from the MIRT 3PL model.

4.3.3.3 Linking and equating

After estimating item parameters from both unidimensional and multidimensional IRT models, parameter estimates from FormX were placed onto the same scale as those from FormY through linking methods—Stocking and Lord (1983) and Haebara (1980) — for both unidimensional and multidimensional IRT methods. In MIRT linking, the method of Li and Lissitz (2000) was used to account for rotational indeterminacy. Both for unidimensional and multidimensional IRT linking, an R package called *plink* (Weeks, 2011) was used.

IRT observed score equating was performed using the computer program, PIE (Hanson & Zeng, 2004). For unidimensional IRT equating, item parameters that were estimated from separate calibration and placed on the same scale after unidimensional linking, or item parameters estimated from concurrent calibration were used as input in PIE. A conversion table was derived as a result of equating. For multidimensional IRT equating, however, there is no available software. Item parameters put on the same scale from multidimensional linking were transformed to unidimensional item parameters via the unidimensional approximation method (Zhang, 1996; Brossman, 2010). Based on approximated item parameters, unidimensional IRT observed score equating using PIE was implemented. Unidimensional approximation procedure was described in Section 2.4.2.3. Equations (2.50), (2.51), and (2.52) were applied to compute approximated parameters.

As a criterion, true equating functions described in Section 3.3.4 were applied. Based on true equating, bias, absolute differences between true and estimated equating results, RMSE, score points equal between true and estimated conversions, and DTM were computed.

4.4 Results

4.4.1 Item parameter recovery

Tables from Table 4.1 to Table 4.6 present correlations between true parameters that were used to generate item responses and estimated item parameters from a unidimensional model. Table 4.1 shows item parameter recovery results from the first method of using separate calibration at the total test level. Correlations between true and estimated location parameters ranged from 0.921 to 0.990. When the distribution was positively skewed with a skewness of .75 (μ4), the correlations were slightly lower than those in the other distributions. In the case of the low correlation among dimensions, the correlations between true and estimated location parameters were smaller in subtest 3. Compared to the location parameters, the slope parameters showed lower correlations—from 0.655 to 0.888. There was a clear pattern showing that the higher the correlations among dimensions, the better the item parameter recovery results for the slope parameters.

Table 4.1. Correlations between True and Estimated Item Parameters from the Unidimensional Model based on Separate Calibration at the Total Test Level

ρ				0.9			0.8			0.6			0.4		
Item & Common				32&4	32&8	16&4	32&4	32&8	16&4	32&4	32&8	16&4	32&4	32&8	16&4
d	Y	μ1	S1	0.977	0.977	0.974	0.978	0.978	0.976	0.978	0.978	0.974	0.978	0.978	0.976
			S2	0.990	0.990	0.988	0.989	0.989	0.988	0.988	0.988	0.987	0.984	0.984	0.983
			S3	0.964	0.964	0.977	0.962	0.962	0.975	0.957	0.957	0.971	0.944	0.944	0.966
	X	μ1	S1	0.977	0.966	0.974	0.977	0.966	0.974	0.978	0.967	0.973	0.976	0.967	0.973
			S2	0.977	0.978	0.982	0.977	0.976	0.980	0.974	0.974	0.980	0.970	0.971	0.981
			S3	0.975	0.974	0.968	0.974	0.972	0.963	0.968	0.965	0.955	0.961	0.957	0.948
		μ2	S1	0.975	0.963	0.971	0.976	0.965	0.970	0.976	0.966	0.974	0.977	0.967	0.973
			S2	0.975	0.974	0.979	0.973	0.975	0.981	0.970	0.974	0.982	0.970	0.971	0.978
			S3	0.974	0.971	0.967	0.974	0.970	0.966	0.967	0.964	0.959	0.964	0.960	0.951
		μ3	S1	0.975	0.966	0.973	0.977	0.967	0.974	0.977	0.967	0.975	0.977	0.966	0.974
			S2	0.975	0.976	0.981	0.974	0.974	0.980	0.973	0.974	0.981	0.970	0.974	0.980
			S3	0.976	0.972	0.968	0.974	0.971	0.964	0.967	0.962	0.955	0.955	0.950	0.938
		μ4	S1	0.968	0.954	0.953	0.971	0.956	0.958	0.973	0.961	0.959	0.973	0.961	0.963
			S2	0.967	0.970	0.971	0.966	0.969	0.971	0.963	0.964	0.970	0.963	0.963	0.970
			S3	0.962	0.960	0.950	0.961	0.956	0.949	0.952	0.948	0.934	0.935	0.928	0.921
		μ5	S1	0.976	0.966	0.971	0.976	0.966	0.972	0.976	0.965	0.973	0.977	0.966	0.973
			S2	0.978	0.978	0.981	0.976	0.978	0.982	0.973	0.975	0.980	0.969	0.970	0.979
			S3	0.977	0.973	0.971	0.974	0.971	0.966	0.970	0.965	0.959	0.959	0.959	0.946
		μ6	S1	0.976	0.964	0.970	0.976	0.964	0.972	0.977	0.968	0.973	0.977	0.966	0.971
			S2	0.976	0.977	0.982	0.974	0.976	0.982	0.971	0.973	0.979	0.968	0.970	0.977
			S3	0.974	0.973	0.965	0.973	0.968	0.966	0.967	0.964	0.956	0.957	0.953	0.945
		μ7	S1	0.976	0.967	0.973	0.978	0.968	0.973	0.978	0.967	0.971	0.978	0.968	0.973
			S2	0.978	0.977	0.982	0.976	0.977	0.982	0.974	0.976	0.981	0.969	0.972	0.978
			S3	0.975	0.973	0.968	0.976	0.972	0.967	0.970	0.966	0.958	0.960	0.957	0.946
a	Y	μ1	S1	0.885	0.885	0.798	0.860	0.860	0.794	0.823	0.823	0.763	0.734	0.734	0.713
			S2	0.861	0.861	0.832	0.846	0.846	0.819	0.808	0.808	0.781	0.759	0.759	0.727
			S3	0.886	0.886	0.845	0.864	0.864	0.827	0.838	0.838	0.754	0.801	0.801	0.698
	X	μ1	S1	0.868	0.854	0.871	0.853	0.846	0.858	0.830	0.820	0.822	0.790	0.772	0.765
			S2	0.862	0.866	0.874	0.834	0.844	0.859	0.789	0.801	0.788	0.724	0.719	0.768
			S3	0.863	0.865	0.849	0.848	0.859	0.832	0.820	0.823	0.810	0.788	0.787	0.787
		μ2	S1	0.866	0.852	0.864	0.854	0.853	0.853	0.824	0.820	0.822	0.793	0.774	0.780
			S2	0.865	0.867	0.882	0.834	0.849	0.856	0.784	0.800	0.815	0.712	0.737	0.751
			S3	0.871	0.871	0.857	0.858	0.854	0.836	0.823	0.825	0.817	0.794	0.801	0.797
		μ3	S1	0.865	0.854	0.859	0.854	0.855	0.847	0.830	0.821	0.826	0.765	0.781	0.763
			S2	0.868	0.870	0.879	0.836	0.839	0.851	0.790	0.802	0.824	0.705	0.733	0.759
			S3	0.860	0.863	0.850	0.850	0.852	0.835	0.820	0.813	0.812	0.782	0.780	0.780
		μ4	S1	0.780	0.774	0.770	0.785	0.765	0.772	0.757	0.756	0.765	0.722	0.690	0.695
			S2	0.798	0.802	0.772	0.773	0.786	0.752	0.722	0.729	0.722	0.675	0.655	0.683
			S3	0.853	0.850	0.849	0.848	0.832	0.839	0.818	0.807	0.827	0.796	0.792	0.800
		μ5	S1	0.880	0.874	0.888	0.861	0.860	0.862	0.828	0.826	0.831	0.784	0.782	0.783
			S2	0.847	0.858	0.828	0.809	0.825	0.822	0.748	0.767	0.746	0.667	0.706	0.688
			S3	0.832	0.840	0.813	0.808	0.822	0.781	0.752	0.767	0.773	0.715	0.724	0.725
		μ6	S1	0.853	0.848	0.859	0.843	0.837	0.843	0.825	0.812	0.796	0.780	0.765	0.768
			S2	0.848	0.858	0.869	0.826	0.845	0.850	0.776	0.794	0.803	0.696	0.710	0.741
			S3	0.866	0.866	0.855	0.857	0.851	0.841	0.822	0.825	0.820	0.804	0.794	0.792
		μ7	S1	0.870	0.868	0.868	0.860	0.853	0.864	0.827	0.814	0.819	0.789	0.772	0.773
			S2	0.864	0.872	0.864	0.830	0.849	0.845	0.785	0.802	0.794	0.716	0.737	0.737
			S3	0.863	0.856	0.849	0.853	0.840	0.837	0.808	0.822	0.791	0.774	0.777	0.763

* Note: d refers to the location parameter and a is the slope parameter.

In Table 4.2, parameter recovery results from separate calibration at the subtest level are presented. The correlations between true and estimated location parameters were still high, all above 0.93 throughout the conditions. Unlike values in Table 4.1, the correlations between true and estimated slope parameters were consistent although the correlations among the dimensions dropped to 0.4. Under the positively skewed distribution (μ4), the correlations of location parameters were still lower than those from the other distributions, and the correlations of slope parameters were the lowest in subtest 1.

Table 4.2. Correlations between True and Estimated Item Parameters from the Unidimensional Model based on Separate Calibration at the Subtest Level

ρ				0.9			0.8			0.6			0.4		
Item & Common				32&4	32&8	16&4	32&4	32&8	16&4	32&4	32&8	16&4	32&4	32&8	16&4
d	Y	μ1	S1	0.975	0.975	0.971	0.976	0.976	0.972	0.973	0.973	0.970	0.974	0.974	0.972
			S2	0.989	0.989	0.987	0.988	0.988	0.987	0.988	0.988	0.988	0.987	0.987	0.987
			S3	0.962	0.962	0.972	0.963	0.963	0.973	0.964	0.964	0.971	0.962	0.962	0.972
	X	μ1	S1	0.975	0.964	0.970	0.974	0.963	0.969	0.973	0.962	0.968	0.973	0.965	0.970
			S2	0.975	0.976	0.981	0.975	0.975	0.980	0.975	0.975	0.979	0.976	0.976	0.979
			S3	0.974	0.971	0.963	0.973	0.972	0.962	0.972	0.971	0.963	0.974	0.971	0.963
		μ2	S1	0.970	0.959	0.966	0.970	0.958	0.962	0.969	0.959	0.964	0.969	0.957	0.963
			S2	0.973	0.971	0.977	0.971	0.972	0.977	0.972	0.973	0.979	0.971	0.972	0.978
			S3	0.971	0.968	0.962	0.973	0.970	0.963	0.971	0.969	0.960	0.972	0.970	0.961
		μ3	S1	0.972	0.964	0.967	0.974	0.964	0.969	0.974	0.964	0.969	0.973	0.962	0.969
			S2	0.973	0.973	0.979	0.971	0.973	0.977	0.971	0.973	0.978	0.972	0.973	0.978
			S3	0.974	0.970	0.960	0.974	0.972	0.958	0.975	0.971	0.960	0.972	0.972	0.961
		μ4	S1	0.955	0.938	0.944	0.958	0.938	0.940	0.958	0.934	0.940	0.955	0.937	0.940
			S2	0.959	0.960	0.963	0.960	0.961	0.960	0.958	0.960	0.963	0.961	0.959	0.961
			S3	0.949	0.945	0.932	0.948	0.947	0.933	0.951	0.951	0.933	0.953	0.948	0.932
		μ5	S1	0.971	0.960	0.960	0.970	0.960	0.961	0.970	0.959	0.959	0.972	0.960	0.959
			S2	0.974	0.973	0.977	0.974	0.975	0.977	0.974	0.974	0.975	0.975	0.974	0.978
			S3	0.974	0.971	0.967	0.973	0.971	0.965	0.973	0.971	0.969	0.975	0.973	0.967
		μ6	S1	0.973	0.962	0.966	0.972	0.961	0.969	0.973	0.963	0.968	0.973	0.960	0.967
			S2	0.974	0.974	0.980	0.974	0.975	0.981	0.972	0.973	0.978	0.974	0.975	0.980
			S3	0.971	0.970	0.960	0.972	0.969	0.961	0.972	0.970	0.961	0.972	0.970	0.958
		μ7	S1	0.973	0.963	0.968	0.974	0.964	0.967	0.973	0.963	0.967	0.972	0.964	0.967
			S2	0.976	0.975	0.980	0.976	0.976	0.981	0.976	0.976	0.981	0.975	0.975	0.980
			S3	0.973	0.973	0.963	0.975	0.973	0.965	0.974	0.972	0.966	0.975	0.972	0.965
a	Y	μ1	S1	0.886	0.886	0.768	0.886	0.886	0.777	0.882	0.882	0.799	0.883	0.883	0.787
			S2	0.846	0.846	0.811	0.847	0.847	0.804	0.842	0.842	0.816	0.844	0.844	0.818
			S3	0.880	0.880	0.816	0.869	0.869	0.819	0.882	0.882	0.821	0.879	0.879	0.816
	X	μ1	S1	0.862	0.854	0.870	0.858	0.856	0.857	0.853	0.852	0.855	0.856	0.858	0.857
			S2	0.851	0.852	0.849	0.832	0.853	0.843	0.836	0.848	0.828	0.843	0.852	0.831
			S3	0.851	0.856	0.823	0.850	0.857	0.821	0.857	0.858	0.828	0.853	0.859	0.827
		μ2	S1	0.858	0.849	0.846	0.860	0.852	0.840	0.853	0.852	0.855	0.859	0.847	0.861
			S2	0.842	0.848	0.852	0.833	0.856	0.837	0.843	0.848	0.843	0.838	0.852	0.840
			S3	0.857	0.861	0.836	0.859	0.856	0.828	0.861	0.859	0.836	0.861	0.862	0.834
		μ3	S1	0.859	0.852	0.844	0.858	0.855	0.844	0.862	0.853	0.847	0.851	0.844	0.842
			S2	0.856	0.858	0.846	0.835	0.853	0.840	0.844	0.856	0.860	0.848	0.854	0.849
			S3	0.853	0.853	0.816	0.854	0.855	0.813	0.855	0.855	0.824	0.840	0.848	0.814
		μ4	S1	0.680	0.677	0.678	0.688	0.655	0.669	0.674	0.680	0.684	0.670	0.670	0.659
			S2	0.751	0.740	0.724	0.742	0.756	0.698	0.739	0.734	0.713	0.762	0.743	0.710
			S3	0.811	0.796	0.786	0.803	0.804	0.777	0.806	0.796	0.782	0.807	0.802	0.787
		μ5	S1	0.881	0.875	0.890	0.871	0.876	0.867	0.875	0.876	0.865	0.878	0.871	0.866
			S2	0.793	0.816	0.752	0.797	0.821	0.763	0.794	0.827	0.748	0.802	0.820	0.762
			S3	0.807	0.823	0.779	0.804	0.812	0.760	0.804	0.815	0.792	0.809	0.825	0.782
		μ6	S1	0.838	0.831	0.827	0.842	0.830	0.819	0.835	0.836	0.819	0.842	0.832	0.835
			S2	0.835	0.838	0.829	0.835	0.845	0.856	0.826	0.843	0.846	0.836	0.838	0.830
			S3	0.855	0.856	0.826	0.853	0.850	0.825	0.852	0.855	0.832	0.849	0.841	0.824
		μ7	S1	0.862	0.867	0.861	0.871	0.860	0.864	0.868	0.857	0.854	0.863	0.863	0.876
			S2	0.845	0.858	0.826	0.839	0.859	0.820	0.835	0.855	0.825	0.832	0.853	0.835
			S3	0.849	0.852	0.824	0.855	0.849	0.828	0.851	0.857	0.825	0.849	0.857	0.826

* Note: d refers to the location parameter and a is the slope parameter.

Table 4.3, Table 4.4, Table 4.5, and Table 4.6 display item parameter recovery results from concurrent calibration. Table 4.3 and Table 4.4 are from the total test level while Table 4.5 and Table 4.6 are from the subtest level. Using separate calibration, item parameters on FormY were estimated once. With concurrent calibration, however, FormY responses were combined into FormX responses, and FormY item parameters were estimated in seven different runs along with each FormX proficiency distribution condition. The correlations of the location parameters were all exceeded 0.92 except when the correlation was 0.4 and the distribution was positively skewed (μ4) in subtest 3 (see Table 4.3). For the slope parameters, the range of correlations lay between 0.675 and 0.901. The correlations between true and estimated slope parameters decreased as the correlation among dimensions dropped from 0.9 to 0.4. Under the skewed distribution (μ4), the values were lower compared to the other distributions. Table 4.4 shows the results from FormY. The correlations between true and estimated parameters fell between 0.918 and 0.991 for the location parameters and between 0.678 and 0.901 for the slope parameters.

Table 4.5 and Table 4.6 present item parameter recovery results from the concurrent calibration at the subtest level (Method 5 in Figure 4.1). The correlations of the location parameters were all high—from 0.935 to 0.989. The slope parameters were all above 0.8, except for the conditions with the positively skewed distribution (μ4) and several cases with 16 items. The correlations between true and estimated parameters were consistent across the four different conditions of correlations (0.9, 0.8, 0.6, and 0.4).

Table 4.3. Correlations between True and Estimated Item Parameters from the Unidimensional Model based on Concurrent Calibration at the Total Test Level (FormX)

ρ			0.9			0.8			0.6			0.4		
Item&Common			32&4	32&8	16&4	32&4	32&8	16&4	32&4	32&8	16&4	32&4	32&8	16&4
d	μ1	S1	0.978	0.968	0.976	0.978	0.969	0.977	0.978	0.969	0.974	0.977	0.969	0.975
		S2	0.978	0.979	0.983	0.977	0.978	0.981	0.974	0.976	0.981	0.971	0.973	0.981
		S3	0.976	0.975	0.973	0.974	0.974	0.967	0.966	0.966	0.958	0.946	0.948	0.941
	μ2	S1	0.976	0.967	0.974	0.977	0.968	0.973	0.977	0.968	0.975	0.978	0.969	0.975
		S2	0.977	0.976	0.981	0.974	0.977	0.983	0.971	0.975	0.983	0.971	0.972	0.978
		S3	0.975	0.974	0.973	0.975	0.973	0.971	0.966	0.964	0.961	0.954	0.952	0.947
	μ3	S1	0.976	0.968	0.974	0.978	0.969	0.975	0.978	0.969	0.976	0.978	0.968	0.975
		S2	0.975	0.977	0.982	0.975	0.975	0.981	0.973	0.974	0.981	0.970	0.973	0.979
		S3	0.975	0.973	0.967	0.972	0.972	0.963	0.962	0.960	0.952	0.934	0.938	0.922
	μ4	S1	0.968	0.958	0.956	0.972	0.961	0.958	0.973	0.962	0.961	0.973	0.963	0.964
		S2	0.967	0.972	0.974	0.968	0.971	0.973	0.965	0.967	0.973	0.968	0.968	0.978
		S3	0.958	0.960	0.947	0.955	0.956	0.946	0.940	0.943	0.922	0.878	0.892	0.879
	μ5	S1	0.977	0.969	0.974	0.977	0.969	0.974	0.977	0.968	0.975	0.978	0.968	0.975
		S2	0.979	0.979	0.983	0.976	0.979	0.982	0.973	0.976	0.981	0.969	0.971	0.979
		S3	0.979	0.976	0.978	0.976	0.975	0.972	0.970	0.967	0.964	0.947	0.955	0.942
	μ6	S1	0.976	0.966	0.972	0.976	0.967	0.973	0.977	0.969	0.975	0.978	0.968	0.973
		S2	0.977	0.978	0.984	0.975	0.977	0.983	0.973	0.974	0.981	0.970	0.972	0.979
		S3	0.974	0.975	0.969	0.973	0.971	0.968	0.965	0.965	0.958	0.939	0.943	0.937
	μ7	S1	0.977	0.969	0.974	0.979	0.971	0.975	0.978	0.969	0.972	0.979	0.971	0.975
		S2	0.979	0.979	0.983	0.977	0.979	0.983	0.975	0.978	0.981	0.970	0.974	0.980
		S3	0.976	0.976	0.974	0.976	0.976	0.972	0.968	0.968	0.963	0.948	0.950	0.940
a	μ1	S1	0.873	0.868	0.884	0.860	0.857	0.874	0.839	0.835	0.842	0.797	0.787	0.786
		S2	0.867	0.879	0.886	0.840	0.858	0.873	0.795	0.815	0.805	0.728	0.737	0.771
		S3	0.865	0.873	0.867	0.850	0.864	0.845	0.816	0.827	0.816	0.760	0.772	0.776
	μ2	S1	0.868	0.864	0.877	0.860	0.863	0.869	0.831	0.838	0.841	0.799	0.789	0.807
		S2	0.870	0.878	0.892	0.842	0.862	0.873	0.790	0.813	0.832	0.716	0.751	0.758
		S3	0.870	0.878	0.870	0.859	0.863	0.846	0.820	0.828	0.818	0.771	0.783	0.785
	μ3	S1	0.868	0.867	0.873	0.860	0.865	0.865	0.837	0.836	0.843	0.772	0.795	0.782
		S2	0.872	0.881	0.888	0.841	0.850	0.864	0.796	0.814	0.833	0.704	0.747	0.748
		S3	0.863	0.875	0.858	0.850	0.858	0.840	0.813	0.817	0.814	0.749	0.759	0.755
	μ4	S1	0.785	0.797	0.803	0.788	0.784	0.792	0.756	0.769	0.785	0.713	0.691	0.694
		S2	0.804	0.819	0.793	0.785	0.805	0.778	0.734	0.751	0.749	0.688	0.691	0.726
		S3	0.838	0.841	0.832	0.828	0.825	0.822	0.787	0.793	0.793	0.731	0.738	0.743
	μ5	S1	0.887	0.884	0.901	0.870	0.872	0.881	0.839	0.842	0.855	0.795	0.795	0.801
		S2	0.855	0.871	0.849	0.821	0.843	0.847	0.760	0.784	0.774	0.675	0.728	0.706
		S3	0.848	0.861	0.851	0.825	0.842	0.818	0.760	0.786	0.797	0.696	0.729	0.734
	μ6	S1	0.858	0.859	0.873	0.851	0.852	0.860	0.833	0.827	0.820	0.785	0.781	0.784
		S2	0.853	0.869	0.880	0.832	0.859	0.861	0.782	0.809	0.819	0.695	0.726	0.751
		S3	0.866	0.873	0.865	0.857	0.857	0.847	0.811	0.827	0.819	0.770	0.771	0.769
	μ7	S1	0.876	0.878	0.880	0.867	0.867	0.884	0.833	0.832	0.835	0.796	0.788	0.794
		S2	0.869	0.880	0.875	0.837	0.862	0.858	0.790	0.817	0.812	0.717	0.748	0.744
		S3	0.867	0.870	0.863	0.859	0.857	0.852	0.810	0.832	0.806	0.752	0.766	0.760

* Note: d refers to the location parameter and a is the slope parameter. $\mu1 = (0,0,0)$, $\mu2 = (0.1,0.1,0.1)$, $\mu3 = (0,0.1,-0.1)$, μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. S1 is subtest 1; S2 is subtest 2; and S3 is subtest 3.

Table 4.4. Correlations between True and Estimated Item Parameters from the Unidimensional Model based on Concurrent Calibration at the Total Test Level (FormY)

ρ			0.9			0.8			0.6			0.4		
Item&Common			32&4	32&8	16&4	32&4	32&8	16&4	32&4	32&8	16&4	32&4	32&8	16&4
d	μ1	S1	0.977	0.978	0.977	0.979	0.979	0.977	0.978	0.979	0.975	0.979	0.979	0.976
		S2	0.991	0.991	0.989	0.990	0.990	0.989	0.988	0.988	0.987	0.985	0.985	0.984
		S3	0.967	0.968	0.979	0.965	0.967	0.978	0.959	0.960	0.974	0.938	0.935	0.964
	μ2	S1	0.976	0.977	0.975	0.978	0.978	0.976	0.977	0.978	0.974	0.978	0.978	0.975
		S2	0.991	0.991	0.989	0.990	0.990	0.989	0.988	0.989	0.988	0.985	0.986	0.984
		S3	0.966	0.968	0.978	0.964	0.966	0.978	0.959	0.960	0.973	0.942	0.942	0.967
	μ3	S1	0.977	0.978	0.976	0.979	0.979	0.977	0.978	0.979	0.975	0.978	0.979	0.977
		S2	0.990	0.991	0.989	0.989	0.990	0.989	0.988	0.988	0.987	0.985	0.985	0.984
		S3	0.967	0.967	0.980	0.965	0.965	0.978	0.958	0.957	0.973	0.931	0.929	0.959
	μ4	S1	0.978	0.978	0.976	0.979	0.979	0.977	0.979	0.979	0.976	0.979	0.980	0.977
		S2	0.990	0.990	0.988	0.989	0.989	0.987	0.987	0.987	0.986	0.985	0.984	0.983
		S3	0.966	0.966	0.978	0.965	0.963	0.976	0.957	0.955	0.970	0.925	0.918	0.955
	μ5	S1	0.977	0.977	0.975	0.978	0.979	0.975	0.977	0.978	0.973	0.978	0.978	0.974
		S2	0.991	0.991	0.989	0.990	0.990	0.989	0.988	0.989	0.987	0.985	0.985	0.984
		S3	0.966	0.968	0.978	0.964	0.966	0.978	0.959	0.960	0.973	0.937	0.940	0.965
	μ6	S1	0.978	0.978	0.977	0.979	0.979	0.977	0.978	0.979	0.976	0.979	0.979	0.977
		S2	0.991	0.991	0.989	0.990	0.990	0.989	0.988	0.988	0.987	0.985	0.985	0.983
		S3	0.967	0.968	0.979	0.965	0.966	0.978	0.959	0.959	0.974	0.937	0.934	0.963
	μ7	S1	0.977	0.978	0.976	0.978	0.979	0.977	0.978	0.979	0.974	0.978	0.979	0.976
		S2	0.991	0.991	0.989	0.990	0.990	0.989	0.988	0.988	0.987	0.985	0.985	0.984
		S3	0.967	0.968	0.979	0.965	0.967	0.978	0.960	0.960	0.974	0.938	0.936	0.965
a	μ1	S1	0.889	0.894	0.811	0.866	0.873	0.809	0.832	0.836	0.785	0.738	0.748	0.735
		S2	0.864	0.870	0.847	0.850	0.855	0.837	0.812	0.819	0.800	0.764	0.768	0.744
		S3	0.896	0.899	0.867	0.873	0.880	0.857	0.847	0.851	0.791	0.796	0.798	0.716
	μ2	S1	0.888	0.895	0.813	0.865	0.874	0.809	0.830	0.837	0.788	0.739	0.746	0.730
		S2	0.864	0.871	0.848	0.851	0.855	0.838	0.813	0.820	0.798	0.765	0.769	0.748
		S3	0.895	0.901	0.868	0.873	0.879	0.856	0.848	0.852	0.793	0.800	0.803	0.722
	μ3	S1	0.889	0.896	0.813	0.865	0.874	0.806	0.831	0.840	0.786	0.739	0.745	0.734
		S2	0.864	0.869	0.851	0.851	0.856	0.837	0.813	0.818	0.801	0.764	0.769	0.740
		S3	0.895	0.900	0.869	0.875	0.881	0.857	0.846	0.849	0.794	0.785	0.792	0.699
	μ4	S1	0.888	0.893	0.802	0.865	0.869	0.798	0.829	0.833	0.778	0.737	0.735	0.711
		S2	0.861	0.867	0.835	0.848	0.850	0.819	0.811	0.816	0.788	0.764	0.762	0.736
		S3	0.895	0.897	0.864	0.875	0.875	0.851	0.846	0.844	0.783	0.776	0.775	0.678
	μ5	S1	0.889	0.898	0.817	0.865	0.875	0.810	0.828	0.837	0.787	0.732	0.742	0.728
		S2	0.863	0.870	0.845	0.851	0.855	0.837	0.812	0.819	0.796	0.764	0.766	0.739
		S3	0.895	0.901	0.863	0.872	0.879	0.852	0.846	0.849	0.782	0.793	0.803	0.707
	μ6	S1	0.889	0.894	0.814	0.865	0.875	0.804	0.831	0.839	0.786	0.738	0.747	0.733
		S2	0.863	0.870	0.846	0.850	0.855	0.833	0.813	0.819	0.799	0.764	0.766	0.742
		S3	0.895	0.899	0.868	0.874	0.879	0.855	0.848	0.851	0.794	0.796	0.799	0.718
	μ7	S1	0.889	0.896	0.817	0.865	0.874	0.809	0.831	0.837	0.787	0.738	0.747	0.732
		S2	0.863	0.871	0.845	0.851	0.856	0.836	0.812	0.818	0.800	0.765	0.769	0.739
		S3	0.895	0.899	0.869	0.874	0.877	0.855	0.848	0.851	0.788	0.798	0.800	0.718

* Note: d refers to the location parameter and a is the slope parameter. $\mu1 = (0,0,0)$, $\mu2 = (0.1,0.1,0.1)$, $\mu3 = (0,0.1,-0.1)$, μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. S1 is subtest 1; S2 is subtest 2; and S3 is subtest 3.

Table 4.5. Correlations between True and Estimated Item Parameters from the Unidimensional Model based on Concurrent Calibration at the Subtest Level (FormX)

ρ			0.9			0.8			0.6			0.4		
Item&Common			32&4	32&8	16&4	32&4	32&8	16&4	32&4	32&8	16&4	32&4	32&8	16&4
d	μ1	S1	0.975	0.967	0.971	0.974	0.966	0.972	0.974	0.965	0.970	0.974	0.967	0.972
		S2	0.976	0.977	0.982	0.976	0.977	0.981	0.976	0.977	0.980	0.976	0.978	0.981
		S3	0.976	0.975	0.967	0.975	0.975	0.967	0.975	0.975	0.969	0.975	0.975	0.969
	μ2	S1	0.971	0.963	0.968	0.971	0.962	0.965	0.970	0.962	0.967	0.971	0.961	0.965
		S2	0.974	0.974	0.978	0.973	0.975	0.979	0.974	0.975	0.981	0.973	0.975	0.980
		S3	0.974	0.972	0.968	0.975	0.974	0.969	0.975	0.973	0.967	0.975	0.974	0.968
	μ3	S1	0.973	0.966	0.969	0.975	0.966	0.970	0.974	0.966	0.971	0.974	0.964	0.971
		S2	0.975	0.975	0.980	0.973	0.975	0.979	0.973	0.975	0.980	0.973	0.975	0.979
		S3	0.976	0.974	0.964	0.976	0.976	0.963	0.976	0.974	0.964	0.974	0.975	0.965
	μ4	S1	0.955	0.945	0.946	0.960	0.946	0.939	0.960	0.942	0.945	0.957	0.942	0.941
		S2	0.960	0.964	0.963	0.962	0.965	0.964	0.960	0.962	0.965	0.962	0.964	0.962
		S3	0.952	0.955	0.935	0.950	0.955	0.939	0.954	0.959	0.940	0.955	0.955	0.935
	μ5	S1	0.972	0.963	0.962	0.971	0.962	0.962	0.971	0.962	0.961	0.973	0.962	0.961
		S2	0.975	0.975	0.978	0.975	0.977	0.978	0.975	0.976	0.977	0.976	0.976	0.979
		S3	0.977	0.975	0.974	0.976	0.975	0.973	0.977	0.975	0.974	0.977	0.976	0.973
	μ6	S1	0.974	0.965	0.968	0.972	0.963	0.970	0.974	0.965	0.970	0.974	0.963	0.970
		S2	0.975	0.975	0.981	0.975	0.976	0.981	0.973	0.974	0.979	0.975	0.976	0.980
		S3	0.972	0.974	0.964	0.974	0.973	0.965	0.974	0.973	0.965	0.974	0.973	0.963
	μ7	S1	0.974	0.966	0.969	0.975	0.967	0.969	0.973	0.965	0.968	0.973	0.967	0.969
		S2	0.977	0.977	0.981	0.977	0.978	0.981	0.977	0.978	0.982	0.976	0.977	0.981
		S3	0.976	0.977	0.969	0.978	0.977	0.971	0.976	0.976	0.970	0.977	0.975	0.970
a	μ1	S1	0.867	0.868	0.882	0.865	0.869	0.873	0.858	0.865	0.870	0.861	0.868	0.869
		S2	0.859	0.868	0.870	0.839	0.863	0.863	0.845	0.861	0.848	0.850	0.864	0.849
		S3	0.860	0.871	0.837	0.861	0.871	0.839	0.867	0.875	0.849	0.864	0.874	0.848
	μ2	S1	0.862	0.861	0.859	0.868	0.866	0.854	0.858	0.866	0.868	0.865	0.860	0.870
		S2	0.848	0.862	0.866	0.841	0.869	0.853	0.850	0.859	0.860	0.846	0.867	0.862
		S3	0.865	0.874	0.856	0.869	0.872	0.844	0.870	0.878	0.854	0.867	0.876	0.852
	μ3	S1	0.864	0.864	0.857	0.865	0.868	0.861	0.868	0.865	0.863	0.859	0.856	0.857
		S2	0.862	0.869	0.865	0.841	0.865	0.854	0.853	0.870	0.872	0.853	0.868	0.867
		S3	0.862	0.868	0.839	0.863	0.871	0.832	0.864	0.870	0.843	0.850	0.866	0.830
	μ4	S1	0.699	0.715	0.725	0.710	0.700	0.700	0.695	0.715	0.743	0.692	0.707	0.697
		S2	0.763	0.760	0.745	0.758	0.784	0.739	0.753	0.753	0.743	0.776	0.770	0.742
		S3	0.821	0.819	0.799	0.809	0.822	0.791	0.817	0.820	0.805	0.817	0.818	0.799
	μ5	S1	0.888	0.887	0.903	0.878	0.886	0.880	0.881	0.890	0.881	0.884	0.882	0.880
		S2	0.804	0.832	0.781	0.807	0.838	0.790	0.804	0.842	0.784	0.813	0.836	0.784
		S3	0.820	0.842	0.810	0.818	0.832	0.795	0.821	0.835	0.823	0.822	0.843	0.806
	μ6	S1	0.846	0.846	0.843	0.849	0.847	0.837	0.843	0.850	0.839	0.849	0.848	0.850
		S2	0.842	0.854	0.851	0.842	0.860	0.867	0.836	0.856	0.862	0.847	0.858	0.846
		S3	0.862	0.868	0.839	0.862	0.864	0.842	0.859	0.871	0.850	0.856	0.858	0.839
	μ7	S1	0.868	0.878	0.870	0.878	0.875	0.879	0.874	0.871	0.867	0.868	0.875	0.888
		S2	0.852	0.868	0.842	0.845	0.872	0.837	0.842	0.868	0.843	0.839	0.867	0.846
		S3	0.861	0.869	0.844	0.868	0.867	0.842	0.864	0.873	0.848	0.860	0.870	0.845

* Note: d refers to the location parameter and a is the slope parameter. $\mu1 = (0,0,0)$, $\mu2 = (0.1,0.1,0.1)$, $\mu3 = (0,0.1,-0.1)$, μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. S1 is subtest 1; S2 is subtest 2; and S3 is subtest 3.

Table 4.6. Correlations between True and Estimated Item Parameters from the Unidimensional Model based on Concurrent Calibration at the Subtest Level (FormY)

ρ			0.9			0.8			0.6			0.4		
Item&Common			32&4	32&8	16&4	32&4	32&8	16&4	32&4	32&8	16&4	32&4	32&8	16&4
d	μ1	S1	0.976	0.976	0.973	0.976	0.977	0.974	0.974	0.975	0.972	0.975	0.976	0.973
		S2	0.989	0.989	0.988	0.989	0.989	0.988	0.988	0.989	0.989	0.987	0.987	0.987
		S3	0.964	0.965	0.975	0.965	0.967	0.976	0.966	0.967	0.974	0.964	0.965	0.975
	μ2	S1	0.974	0.975	0.971	0.975	0.975	0.972	0.973	0.973	0.971	0.974	0.974	0.971
		S2	0.989	0.989	0.988	0.989	0.989	0.987	0.988	0.989	0.988	0.987	0.987	0.987
		S3	0.963	0.964	0.974	0.964	0.966	0.975	0.964	0.966	0.973	0.963	0.964	0.974
	μ3	S1	0.975	0.977	0.972	0.976	0.977	0.973	0.974	0.975	0.972	0.975	0.976	0.974
		S2	0.989	0.989	0.987	0.989	0.989	0.987	0.988	0.988	0.988	0.987	0.987	0.987
		S3	0.965	0.965	0.976	0.965	0.967	0.976	0.966	0.967	0.975	0.964	0.966	0.975
	μ4	S1	0.976	0.976	0.971	0.976	0.976	0.972	0.974	0.974	0.970	0.975	0.976	0.973
		S2	0.988	0.988	0.986	0.988	0.988	0.986	0.988	0.987	0.987	0.986	0.986	0.986
		S3	0.964	0.962	0.974	0.964	0.963	0.974	0.965	0.964	0.974	0.964	0.963	0.974
	μ5	S1	0.974	0.975	0.970	0.975	0.975	0.970	0.973	0.973	0.969	0.974	0.974	0.969
		S2	0.989	0.989	0.987	0.989	0.989	0.987	0.988	0.989	0.988	0.987	0.987	0.986
		S3	0.962	0.965	0.973	0.963	0.965	0.973	0.964	0.966	0.973	0.963	0.965	0.972
	μ6	S1	0.976	0.977	0.973	0.976	0.977	0.974	0.974	0.976	0.972	0.975	0.976	0.974
		S2	0.989	0.989	0.988	0.989	0.989	0.988	0.988	0.989	0.989	0.987	0.987	0.987
		S3	0.964	0.965	0.975	0.965	0.966	0.975	0.966	0.967	0.975	0.965	0.965	0.975
	μ7	S1	0.975	0.976	0.972	0.976	0.977	0.973	0.974	0.974	0.971	0.975	0.976	0.973
		S2	0.989	0.989	0.988	0.989	0.989	0.988	0.988	0.989	0.988	0.987	0.987	0.987
		S3	0.964	0.965	0.974	0.965	0.966	0.975	0.966	0.967	0.974	0.964	0.965	0.974
a	μ1	S1	0.892	0.897	0.786	0.890	0.894	0.798	0.888	0.890	0.817	0.889	0.895	0.809
		S2	0.851	0.857	0.827	0.854	0.858	0.822	0.850	0.858	0.832	0.849	0.855	0.825
		S3	0.889	0.892	0.831	0.875	0.882	0.841	0.887	0.892	0.839	0.885	0.890	0.831
	μ2	S1	0.892	0.899	0.787	0.890	0.896	0.796	0.888	0.891	0.824	0.891	0.895	0.810
		S2	0.851	0.859	0.826	0.854	0.859	0.821	0.849	0.855	0.835	0.850	0.855	0.829
		S3	0.889	0.893	0.833	0.875	0.880	0.842	0.887	0.891	0.840	0.886	0.891	0.832
	μ3	S1	0.893	0.899	0.790	0.889	0.896	0.792	0.887	0.891	0.818	0.890	0.892	0.810
		S2	0.851	0.858	0.827	0.855	0.858	0.821	0.849	0.855	0.834	0.849	0.855	0.832
		S3	0.888	0.892	0.834	0.876	0.883	0.840	0.888	0.891	0.843	0.885	0.890	0.831
	μ4	S1	0.888	0.892	0.768	0.887	0.886	0.778	0.882	0.880	0.800	0.886	0.885	0.789
		S2	0.846	0.852	0.810	0.848	0.851	0.803	0.844	0.846	0.814	0.842	0.844	0.817
		S3	0.889	0.888	0.832	0.876	0.877	0.839	0.887	0.887	0.841	0.886	0.889	0.830
	μ5	S1	0.892	0.900	0.793	0.890	0.898	0.799	0.887	0.893	0.825	0.891	0.895	0.811
		S2	0.849	0.854	0.821	0.854	0.857	0.818	0.847	0.855	0.828	0.847	0.853	0.822
		S3	0.886	0.893	0.824	0.873	0.879	0.826	0.884	0.891	0.831	0.884	0.889	0.822
	μ6	S1	0.891	0.898	0.784	0.890	0.895	0.791	0.886	0.890	0.815	0.889	0.894	0.804
		S2	0.852	0.858	0.824	0.854	0.857	0.822	0.849	0.855	0.833	0.848	0.851	0.827
		S3	0.889	0.894	0.833	0.876	0.882	0.840	0.889	0.892	0.842	0.886	0.891	0.834
	μ7	S1	0.891	0.899	0.790	0.889	0.896	0.794	0.887	0.892	0.823	0.889	0.896	0.810
		S2	0.851	0.856	0.824	0.855	0.858	0.820	0.849	0.857	0.832	0.849	0.855	0.829
		S3	0.888	0.892	0.834	0.875	0.880	0.838	0.887	0.891	0.837	0.885	0.890	0.831

* Note: d refers to the location parameter and a is the slope parameter. $\mu1 = (0,0,0)$, $\mu2 = (0.1,0.1,0.1)$, $\mu3 = (0,0.1,-0.1)$, μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. S1 is subtest 1; S2 is subtest 2; and S3 is subtest 3.

Table 4.7 and Table 4.8 show item parameter recovery results from an MIRT model. Table 4.7 displays the results from separate calibration. The correlations of location parameters were all above 0.95, which was greater than unidimensional results in general. Unlike the results from a unidimensional model, an MIRT model produced three slope parameters as there were three dimensions. Compared to the separate calibration from a unidimensional model, the correlations of slope parameters did not decrease as correlations among dimensions dropped to 0.4. Correlations between true and estimated slope parameters were all exceeded 0.8 except for several cases under the positively skewed distribution with high skewness (μ4), especially in *a*1. The range of correlations in this case was between 0.676 and 0.736.

Table 4.8 and Table 4.9 present MIRT parameter recovery results from concurrent calibration for FormX and FormY, respectively. The location parameters showed very high correlations above 0.95 in FormX and above 0.97 in FormY. The slope parameters ranged from 0.84 to 0.91 in FormX except for *a*1 parameters under μ4 where the range fell between 0.70 and 0.78. In FormY, the correlations of slope parameters were from 0.79 to 0.91. Differences between simulation conditions were very small. Compared to other estimation methods, item parameter recovery results from MIRT concurrent calibration were consistent across all conditions.

Table 4.7. Correlations between True and Estimated Item Parameters from the Multidimensional Model based on Separate Calibration

ρ			0.9			0.8			0.6			0.4		
Item&Common			32&4	32&8	16&4	32&4	32&8	16&4	32&4	32&8	16&4	32&4	32&8	16&4
Y	μ1	d	0.981	0.981	0.979	0.980	0.980	0.978	0.979	0.979	0.977	0.977	0.977	0.976
		a1	0.895	0.895	0.793	0.888	0.888	0.800	0.885	0.885	0.807	0.882	0.882	0.792
		a2	0.861	0.861	0.835	0.858	0.858	0.829	0.846	0.846	0.825	0.846	0.846	0.822
		a3	0.893	0.893	0.858	0.879	0.879	0.850	0.887	0.887	0.834	0.881	0.881	0.822
X	μ1	d	0.977	0.975	0.975	0.977	0.974	0.974	0.976	0.973	0.973	0.976	0.974	0.972
		a1	0.866	0.855	0.873	0.862	0.858	0.866	0.856	0.855	0.856	0.858	0.859	0.859
		a2	0.868	0.870	0.876	0.845	0.862	0.869	0.843	0.853	0.841	0.845	0.853	0.842
		a3	0.862	0.866	0.847	0.857	0.865	0.836	0.859	0.861	0.836	0.855	0.861	0.830
	μ2	d	0.974	0.971	0.972	0.974	0.971	0.972	0.973	0.971	0.971	0.973	0.970	0.969
		a1	0.863	0.851	0.860	0.864	0.857	0.857	0.855	0.854	0.858	0.860	0.848	0.860
		a2	0.864	0.867	0.884	0.847	0.866	0.865	0.847	0.852	0.857	0.840	0.856	0.843
		a3	0.867	0.870	0.858	0.867	0.863	0.843	0.863	0.860	0.842	0.861	0.864	0.839
	μ3	d	0.968	0.965	0.967	0.968	0.966	0.966	0.966	0.964	0.964	0.966	0.964	0.964
		a1	0.864	0.854	0.855	0.863	0.860	0.854	0.862	0.854	0.855	0.851	0.847	0.844
		a2	0.873	0.870	0.879	0.853	0.859	0.861	0.849	0.860	0.872	0.850	0.857	0.855
		a3	0.862	0.863	0.841	0.861	0.863	0.837	0.859	0.857	0.834	0.843	0.849	0.819
	μ4	d	0.963	0.961	0.957	0.962	0.959	0.955	0.960	0.955	0.949	0.959	0.954	0.946
		a1	0.736	0.729	0.735	0.726	0.699	0.716	0.686	0.697	0.704	0.676	0.678	0.676
		a2	0.789	0.790	0.769	0.767	0.787	0.749	0.753	0.742	0.731	0.770	0.750	0.715
		a3	0.828	0.822	0.817	0.822	0.816	0.802	0.809	0.801	0.790	0.809	0.803	0.791
	μ5	d	0.977	0.974	0.975	0.976	0.974	0.973	0.974	0.972	0.971	0.975	0.972	0.970
		a1	0.887	0.880	0.898	0.877	0.880	0.878	0.877	0.876	0.871	0.878	0.871	0.869
		a2	0.850	0.860	0.832	0.831	0.846	0.831	0.810	0.836	0.783	0.808	0.823	0.773
		a3	0.839	0.852	0.819	0.825	0.834	0.793	0.811	0.821	0.804	0.811	0.827	0.787
	μ6	d	0.976	0.974	0.974	0.975	0.973	0.974	0.974	0.972	0.972	0.974	0.972	0.970
		a1	0.845	0.839	0.850	0.843	0.835	0.839	0.837	0.836	0.822	0.842	0.833	0.835
		a2	0.851	0.859	0.870	0.845	0.857	0.867	0.832	0.847	0.852	0.836	0.839	0.839
		a3	0.863	0.865	0.851	0.859	0.857	0.839	0.853	0.858	0.839	0.851	0.844	0.828
	μ7	d	0.977	0.975	0.975	0.978	0.975	0.975	0.976	0.974	0.973	0.975	0.974	0.973
		a1	0.868	0.870	0.872	0.873	0.864	0.876	0.870	0.857	0.859	0.865	0.864	0.875
		a2	0.868	0.877	0.865	0.852	0.867	0.854	0.842	0.861	0.844	0.835	0.855	0.839
		a3	0.862	0.862	0.850	0.865	0.856	0.848	0.854	0.861	0.830	0.849	0.858	0.826

* Note: d refers to the location parameter and a1, a2, and a3 are the slope parameters for dimensions 1, 2, and 3, respectively. $\mu1 = (0,0,0)$, $\mu2 = (0.1,0.1,0.1)$, $\mu3 = (0,0.1,-0.1)$, μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25.

Table 4.8. Correlations between True and Estimated Item Parameters from the Multidimensional Model based on Concurrent Calibration (Parameters from FormX)

ρ		0.9			0.8			0.6			0.4		
Item&Common		32&4	32&8	16&4	32&4	32&8	16&4	32&4	32&8	16&4	32&4	32&8	16&4
μ1	d	0.979	0.977	0.978	0.978	0.977	0.976	0.977	0.976	0.976	0.977	0.976	0.975
	a1	0.872	0.871	0.885	0.869	0.870	0.880	0.861	0.866	0.872	0.863	0.869	0.871
	a2	0.874	0.884	0.891	0.851	0.874	0.887	0.852	0.866	0.860	0.852	0.865	0.856
	a3	0.870	0.880	0.869	0.868	0.879	0.856	0.870	0.878	0.857	0.865	0.875	0.852
μ2	d	0.977	0.974	0.976	0.976	0.975	0.976	0.975	0.974	0.975	0.975	0.974	0.973
	a1	0.867	0.864	0.873	0.872	0.869	0.869	0.861	0.868	0.870	0.866	0.862	0.870
	a2	0.870	0.879	0.894	0.855	0.879	0.882	0.854	0.863	0.874	0.847	0.871	0.864
	a3	0.875	0.883	0.877	0.877	0.879	0.859	0.872	0.879	0.862	0.867	0.878	0.857
μ3	d	0.970	0.970	0.970	0.971	0.971	0.970	0.970	0.971	0.969	0.970	0.971	0.969
	a1	0.868	0.868	0.869	0.870	0.871	0.870	0.868	0.866	0.868	0.859	0.858	0.859
	a2	0.877	0.881	0.891	0.858	0.870	0.876	0.858	0.874	0.884	0.856	0.870	0.872
	a3	0.870	0.880	0.862	0.870	0.878	0.853	0.868	0.871	0.854	0.853	0.867	0.836
μ4	d	0.965	0.967	0.961	0.964	0.965	0.959	0.962	0.961	0.955	0.961	0.960	0.949
	a1	0.756	0.767	0.781	0.747	0.744	0.757	0.707	0.730	0.755	0.695	0.714	0.712
	a2	0.801	0.812	0.795	0.784	0.812	0.779	0.766	0.764	0.756	0.785	0.777	0.741
	a3	0.841	0.842	0.835	0.829	0.834	0.824	0.820	0.824	0.814	0.819	0.819	0.804
μ5	d	0.978	0.977	0.978	0.977	0.976	0.976	0.976	0.975	0.974	0.977	0.975	0.973
	a1	0.893	0.891	0.909	0.884	0.890	0.892	0.883	0.889	0.887	0.885	0.882	0.883
	a2	0.857	0.872	0.853	0.839	0.861	0.854	0.819	0.851	0.814	0.819	0.838	0.795
	a3	0.851	0.868	0.852	0.839	0.853	0.825	0.827	0.840	0.833	0.824	0.845	0.812
μ6	d	0.977	0.976	0.976	0.976	0.975	0.976	0.976	0.974	0.974	0.976	0.975	0.973
	a1	0.853	0.854	0.866	0.851	0.852	0.853	0.845	0.850	0.841	0.850	0.848	0.850
	a2	0.858	0.871	0.883	0.852	0.871	0.879	0.842	0.860	0.870	0.847	0.859	0.854
	a3	0.871	0.877	0.867	0.870	0.873	0.859	0.860	0.874	0.859	0.858	0.860	0.844
μ7	d	0.978	0.977	0.977	0.979	0.977	0.977	0.977	0.977	0.975	0.977	0.976	0.976
	a1	0.875	0.881	0.882	0.879	0.879	0.890	0.875	0.870	0.870	0.870	0.877	0.888
	a2	0.874	0.886	0.877	0.857	0.880	0.869	0.849	0.875	0.859	0.842	0.868	0.852
	a3	0.872	0.878	0.868	0.877	0.873	0.865	0.867	0.877	0.854	0.860	0.871	0.848

* Note: d refers to the location parameter and a1, a2, and a3 are the slope parameters for dimensions 1, 2, and 3, respectively. $\mu 1 = (0,0,0)$, $\mu 2 = (0.1,0.1,0.1)$, $\mu 3 = (0,0.1,-0.1)$, μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25.

Table 4.9. Correlations between True and Estimated Item Parameters from the Multidimensional Model based on Concurrent Calibration (Parameters from FormY)

ρ		0.9			0.8			0.6			0.4		
Item&Common		32&4	32&8	16&4	32&4	32&8	16&4	32&4	32&8	16&4	32&4	32&8	16&4
μ1	d	0.982	0.982	0.981	0.981	0.982	0.980	0.980	0.981	0.979	0.978	0.979	0.978
	a1	0.900	0.904	0.808	0.893	0.899	0.819	0.892	0.894	0.827	0.889	0.894	0.813
	a2	0.866	0.873	0.850	0.863	0.868	0.848	0.853	0.860	0.842	0.851	0.857	0.830
	a3	0.902	0.905	0.870	0.885	0.891	0.867	0.892	0.896	0.851	0.886	0.892	0.836
μ2	d	0.981	0.981	0.979	0.980	0.981	0.980	0.979	0.980	0.978	0.978	0.978	0.977
	a1	0.899	0.906	0.810	0.893	0.899	0.816	0.892	0.896	0.833	0.890	0.894	0.814
	a2	0.866	0.873	0.850	0.864	0.869	0.847	0.853	0.857	0.843	0.852	0.857	0.832
	a3	0.901	0.906	0.872	0.885	0.890	0.868	0.891	0.896	0.853	0.888	0.893	0.838
μ3	d	0.982	0.982	0.981	0.981	0.982	0.980	0.980	0.980	0.979	0.978	0.978	0.978
	a1	0.900	0.906	0.809	0.893	0.900	0.814	0.891	0.896	0.828	0.889	0.892	0.814
	a2	0.866	0.872	0.853	0.865	0.869	0.846	0.852	0.858	0.846	0.851	0.858	0.835
	a3	0.901	0.905	0.871	0.886	0.893	0.868	0.893	0.896	0.856	0.886	0.892	0.838
μ4	d	0.982	0.982	0.979	0.981	0.981	0.979	0.980	0.980	0.978	0.978	0.978	0.977
	a1	0.898	0.901	0.791	0.891	0.892	0.801	0.886	0.885	0.814	0.885	0.885	0.794
	a2	0.862	0.868	0.836	0.859	0.862	0.828	0.848	0.849	0.825	0.845	0.846	0.823
	a3	0.901	0.902	0.868	0.886	0.886	0.864	0.891	0.891	0.853	0.888	0.890	0.835
μ5	d	0.981	0.982	0.979	0.981	0.981	0.979	0.979	0.980	0.977	0.978	0.978	0.976
	a1	0.901	0.909	0.815	0.894	0.902	0.820	0.891	0.897	0.834	0.891	0.895	0.815
	a2	0.865	0.872	0.849	0.864	0.868	0.846	0.851	0.857	0.840	0.850	0.855	0.827
	a3	0.901	0.907	0.866	0.884	0.890	0.860	0.889	0.895	0.844	0.886	0.891	0.828
μ6	d	0.982	0.982	0.980	0.981	0.982	0.980	0.980	0.981	0.979	0.978	0.979	0.978
	a1	0.899	0.904	0.809	0.892	0.899	0.810	0.889	0.895	0.827	0.889	0.893	0.809
	a2	0.865	0.872	0.848	0.864	0.867	0.844	0.852	0.857	0.842	0.851	0.854	0.832
	a3	0.901	0.905	0.872	0.886	0.891	0.866	0.893	0.897	0.853	0.888	0.893	0.840
μ7	d	0.982	0.982	0.980	0.981	0.982	0.980	0.980	0.981	0.979	0.978	0.979	0.978
	a1	0.900	0.907	0.813	0.893	0.900	0.817	0.891	0.896	0.831	0.889	0.896	0.814
	a2	0.865	0.872	0.848	0.864	0.869	0.844	0.852	0.859	0.843	0.851	0.858	0.834
	a3	0.901	0.905	0.873	0.885	0.889	0.866	0.892	0.895	0.848	0.886	0.891	0.837

* Note: d refers to the location parameter and a1, a2, and a3 are the slope parameters for dimensions 1, 2, and 3, respectively. $\mu 1 = (0,0,0)$, $\mu 2 = (0.1,0.1,0.1)$, $\mu 3 = (0,0.1,-0.1)$, μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25.

4.4.2 Bias, absolute difference, and RMSE

4.4.2.1 Bias

Table 4.10, Table 4.11, and Table 4.12 present bias from subtest 1, subtest 2, and subtest 3, respectively. Among five unidimensional methods, u1 and u2 are results from separate calibration at the total test level with one and three linking constants, respectively; u3 is from separate calibration at the subtest level; u4 and u5 are from concurrent calibration at the total and subtest levels, respectively. Multidimensional methods are m1 and m2 where m1 is from separate calibration and m2 is from concurrent calibration. When separate calibration was applied, two different linking methods—Stocking and Lord and Haebara methods—were implemented.

When a unidimensional model was used, the Haebara linking method slightly outperformed the Stocking and Lord method within the same calibration and linking design in subtests 1 and 3. On the other hand, the Stocking and Lord method produced slightly better results in subtest 2. In a multidimensional model, however, the difference between these two linking methods proved negligible.

Table 4.10 displays bias values from subtest 1. In subtest 1 under the separate calibration at the total test level, using three different linking constants (u2) showed slightly smaller bias than using one linking constant (u1) in many cases except for several cases such as the positively skewed distribution (μ4) with correlations of 0.9, 0.8, and 0.6 and the number of items equal to 32. Results from separate calibration at the subtest level (u3) produced slightly larger bias than results at the total test level (u2); however, the differences were small.

When concurrent calibration based on the unidimensional model was used, bias values from the total test level tended to be slightly smaller than those from the subtest level except for several cases from shifted or skewed distributions. Compared to the concurrent calibration methods at the subtest level (u5), separate calibration at the total test level using three linking constants (u2) yielded smaller bias values. However, when the correlation was high (0.9), this was not always the case. When the distributions were negatively skewed (μ5 and μ7), results from separate calibration produced smaller bias than those from conccurrent calibration.

Under the multidimensional methods, with the correlation of 0.9 and the correlation of 0.8 with smaller number of items, concurrent calibration yielded slightly better results than separate calibration. On the other hand, the separate calibration method outperformed the separate calibration method as correlation decreased.

In general, multidimensional approaches did not perform as well as unidimensional approaches in many cases, especially with high correlation (0.9). When the distributions were positively skewed, particularly with a skewness of 0.75 (μ4), and the multidimensional IRT approach was applied, bias values were the largest. With a skewness of 0.25 (μ6), bias values were still larger than those from unidimensional approaches regardless of correlations among dimensions and the number of items.

Table 4.10. Bias compared with True Equating Function (Subtest 1)

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4
μ1	u1 HB	-0.248	-0.145	-0.045	-0.286	-0.250	-0.116	-0.235	-0.182	-0.111	-0.359	-0.366	-0.230
	u1 SL	-0.300	-0.187	-0.072	-0.322	-0.289	-0.144	-0.293	-0.246	-0.143	-0.458	-0.468	-0.280
	u2 HB	-0.240	-0.131	-0.039	-0.149	-0.180	-0.083	-0.122	-0.118	-0.039	-0.115	-0.048	0.001
	u2 SL	-0.264	-0.175	-0.056	-0.169	-0.210	-0.102	-0.156	-0.167	-0.056	-0.164	-0.091	-0.022
	u3 HB	-0.258	-0.085	-0.069	-0.177	-0.210	-0.102	-0.199	-0.195	-0.072	-0.191	-0.119	-0.075
	u3 SL	-0.281	-0.127	-0.084	-0.199	-0.241	-0.120	-0.213	-0.233	-0.085	-0.232	-0.148	-0.081
	u4	-0.254	-0.164	-0.067	-0.324	-0.282	-0.139	-0.274	-0.217	-0.126	-0.350	-0.298	-0.186
	u5	-0.217	-0.169	-0.067	-0.262	-0.252	-0.115	-0.205	-0.195	-0.078	-0.206	-0.170	-0.072
	m1 HB	0.393	0.402	0.342	0.233	0.233	0.294	0.333	0.354	0.283	0.225	0.206	0.227
	m1 SL	0.390	0.400	0.342	0.222	0.228	0.295	0.334	0.350	0.280	0.216	0.193	0.224
	m2	0.359	0.376	0.332	0.306	0.239	0.281	0.367	0.354	0.288	0.366	0.214	0.232
μ2	u1 HB	-0.278	-0.225	-0.087	-0.288	-0.226	-0.089	-0.286	-0.230	-0.135	-0.356	-0.294	-0.221
	u1 SL	-0.320	-0.273	-0.106	-0.329	-0.282	-0.115	-0.337	-0.298	-0.173	-0.464	-0.398	-0.282
	u2 HB	-0.222	-0.152	-0.046	-0.215	-0.218	-0.047	-0.135	-0.102	-0.041	0.009	-0.034	0.018
	u2 SL	-0.258	-0.199	-0.061	-0.244	-0.265	-0.067	-0.165	-0.143	-0.062	-0.035	-0.100	-0.010
	u3 HB	-0.249	-0.105	-0.064	-0.247	-0.266	-0.064	-0.157	-0.172	-0.090	-0.105	-0.143	-0.074
	u3 SL	-0.293	-0.150	-0.074	-0.275	-0.317	-0.087	-0.180	-0.199	-0.103	-0.128	-0.190	-0.079
	u4	-0.030	-0.107	0.008	-0.114	-0.171	-0.028	-0.110	-0.141	-0.059	-0.089	-0.123	-0.084
	u5	0.207	0.053	0.102	0.079	-0.077	0.052	0.119	0.003	0.043	0.222	0.055	0.081
	m1 HB	0.429	0.471	0.374	0.269	0.274	0.308	0.265	0.280	0.261	0.176	0.219	0.203
	m1 SL	0.422	0.469	0.374	0.268	0.271	0.307	0.259	0.278	0.261	0.166	0.203	0.197
	m2	0.414	0.441	0.364	0.355	0.260	0.296	0.377	0.279	0.255	0.382	0.193	0.194
μ3	u1 HB	-0.251	-0.171	-0.051	-0.269	-0.180	-0.112	-0.238	-0.183	-0.119	-0.349	-0.336	-0.215
	u1 SL	-0.277	-0.201	-0.071	-0.291	-0.227	-0.140	-0.293	-0.261	-0.157	-0.477	-0.464	-0.283
	u2 HB	-0.284	-0.171	-0.021	-0.203	-0.178	-0.045	-0.096	-0.085	-0.028	-0.014	-0.079	-0.004
	u2 SL	-0.309	-0.210	-0.033	-0.238	-0.218	-0.064	-0.124	-0.134	-0.045	-0.058	-0.135	-0.028
	u3 HB	-0.302	-0.128	-0.035	-0.232	-0.219	-0.075	-0.169	-0.180	-0.076	-0.151	-0.190	-0.059
	u3 SL	-0.332	-0.164	-0.039	-0.269	-0.260	-0.097	-0.181	-0.222	-0.085	-0.183	-0.216	-0.072
	u4	-0.240	-0.168	-0.055	-0.332	-0.229	-0.133	-0.290	-0.196	-0.127	-0.320	-0.269	-0.184
	u5	-0.222	-0.182	-0.043	-0.301	-0.270	-0.119	-0.240	-0.210	-0.089	-0.215	-0.212	-0.068
	m1 HB	0.406	0.416	0.359	0.205	0.230	0.272	0.284	0.305	0.271	0.189	0.207	0.231
	m1 SL	0.399	0.410	0.360	0.200	0.224	0.270	0.276	0.302	0.270	0.182	0.196	0.229
	m2	0.366	0.398	0.356	0.313	0.219	0.262	0.356	0.302	0.269	0.360	0.202	0.226
μ4	u1 HB	-0.194	-0.174	0.087	-0.238	-0.209	0.056	-0.080	-0.013	0.082	0.248	0.271	0.114
	u1 SL	-0.230	-0.254	0.067	-0.265	-0.278	0.040	-0.108	-0.077	0.061	0.182	0.161	0.073
	u2 HB	-0.306	-0.253	0.076	-0.308	-0.261	0.004	-0.208	-0.111	0.012	0.025	-0.077	0.050
	u2 SL	-0.332	-0.324	0.055	-0.331	-0.318	-0.005	-0.224	-0.160	-0.002	-0.020	-0.171	0.031
	u3 HB	-0.211	-0.103	0.109	-0.202	-0.186	0.098	-0.171	-0.141	0.095	-0.058	-0.160	0.078
	u3 SL	-0.280	-0.248	0.074	-0.305	-0.329	0.065	-0.257	-0.273	0.025	-0.163	-0.288	0.058
	u4	0.158	0.080	0.192	0.087	0.022	0.155	0.242	0.180	0.161	0.561	0.418	0.191
	u5	0.200	0.114	0.223	0.139	0.053	0.208	0.203	0.115	0.214	0.241	0.094	0.214
	m1 HB	1.626	1.672	1.121	1.596	1.641	1.114	1.821	1.883	1.147	1.816	1.895	1.121
	m1 SL	1.616	1.657	1.119	1.582	1.637	1.114	1.805	1.861	1.139	1.804	1.870	1.122
	m2	1.327	1.274	0.911	1.371	1.240	0.909	1.502	1.467	0.938	1.554	1.463	0.931
μ5	u1 HB	-0.151	-0.097	-0.098	-0.245	-0.229	-0.166	-0.309	-0.275	-0.203	-0.608	-0.591	-0.400
	u1 SL	-0.226	-0.137	-0.137	-0.318	-0.301	-0.208	-0.410	-0.358	-0.257	-0.765	-0.715	-0.482
	u2 HB	-0.054	0.004	-0.038	-0.004	-0.071	-0.047	0.044	-0.008	-0.004	0.044	0.002	-0.019
	u2 SL	-0.099	-0.033	-0.061	-0.054	-0.123	-0.076	-0.013	-0.064	-0.036	-0.021	-0.043	-0.040
	u3 HB	-0.040	0.109	-0.058	0.016	-0.068	-0.077	0.019	-0.019	-0.045	0.050	0.039	-0.048
	u3 SL	-0.093	0.072	-0.079	-0.039	-0.117	-0.106	-0.028	-0.064	-0.080	0.011	-0.003	-0.068
	u4	-0.349	-0.242	-0.200	-0.452	-0.387	-0.252	-0.490	-0.381	-0.268	-0.745	-0.575	-0.392

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4
	u5	-0.363	-0.245	-0.235	-0.400	-0.377	-0.269	-0.343	-0.292	-0.227	-0.351	-0.273	-0.229
	m1 HB	0.081	0.091	0.127	-0.038	-0.024	0.095	0.019	0.035	0.079	-0.135	-0.108	-0.008
	m1 SL	0.080	0.090	0.126	-0.051	-0.029	0.091	0.017	0.033	0.075	-0.150	-0.134	-0.022
	m2	0.087	0.174	0.189	0.039	0.083	0.157	0.065	0.151	0.146	0.034	0.027	0.069
μ6	u1 HB	-0.258	-0.185	-0.035	-0.212	-0.226	-0.067	-0.210	-0.147	-0.071	-0.271	-0.215	-0.138
	u1 SL	-0.291	-0.225	-0.051	-0.235	-0.265	-0.084	-0.242	-0.195	-0.108	-0.338	-0.306	-0.177
	u2 HB	-0.293	-0.184	-0.036	-0.197	-0.218	-0.045	-0.183	-0.145	-0.015	-0.012	-0.028	0.037
	u2 SL	-0.325	-0.223	-0.045	-0.202	-0.247	-0.057	-0.203	-0.192	-0.033	-0.064	-0.082	0.019
	u3 HB	-0.340	-0.158	-0.050	-0.266	-0.296	-0.064	-0.275	-0.278	-0.067	-0.165	-0.153	-0.009
	u3 SL	-0.380	-0.198	-0.054	-0.261	-0.322	-0.078	-0.299	-0.318	-0.086	-0.210	-0.193	-0.016
	u4	-0.171	-0.151	-0.021	-0.210	-0.210	-0.062	-0.189	-0.135	-0.061	-0.189	-0.148	-0.095
	u5	-0.176	-0.162	-0.012	-0.218	-0.224	-0.042	-0.193	-0.172	-0.025	-0.114	-0.128	0.005
	m1 HB	0.591	0.622	0.468	0.466	0.460	0.411	0.534	0.574	0.412	0.477	0.486	0.381
	m1 SL	0.590	0.620	0.469	0.464	0.457	0.413	0.533	0.568	0.411	0.478	0.484	0.380
	m2	0.503	0.542	0.430	0.503	0.398	0.366	0.502	0.497	0.381	0.542	0.420	0.339
μ7	u1 HB	-0.181	-0.117	-0.060	-0.222	-0.187	-0.127	-0.242	-0.214	-0.151	-0.423	-0.414	-0.277
	u1 SL	-0.239	-0.164	-0.081	-0.284	-0.236	-0.152	-0.311	-0.267	-0.191	-0.554	-0.530	-0.329
	u2 HB	-0.147	-0.128	-0.011	-0.092	-0.141	-0.089	-0.071	-0.064	-0.015	0.006	-0.027	-0.009
	u2 SL	-0.176	-0.169	-0.025	-0.133	-0.180	-0.112	-0.107	-0.108	-0.030	-0.038	-0.080	-0.021
	u3 HB	-0.156	-0.063	-0.027	-0.102	-0.151	-0.110	-0.106	-0.113	-0.057	-0.093	-0.068	-0.070
	u3 SL	-0.170	-0.099	-0.038	-0.147	-0.190	-0.131	-0.128	-0.150	-0.069	-0.126	-0.113	-0.072
	u4	-0.284	-0.214	-0.099	-0.341	-0.287	-0.175	-0.337	-0.268	-0.166	-0.481	-0.373	-0.246
	u5	-0.268	-0.226	-0.093	-0.288	-0.275	-0.167	-0.245	-0.217	-0.113	-0.229	-0.199	-0.113
	m1 HB	0.254	0.262	0.256	0.136	0.128	0.212	0.180	0.193	0.193	0.051	0.090	0.127
	m1 SL	0.251	0.259	0.256	0.126	0.122	0.208	0.178	0.189	0.190	0.035	0.077	0.119
	m2	0.246	0.291	0.283	0.227	0.166	0.225	0.246	0.240	0.225	0.237	0.134	0.149

* Note: $\mu1 = (0,0,0)$, $\mu2 = (0.1,0.1,0.1)$, $\mu3 = (0,0.1,-0.1)$, μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. u1 is unidimensional method 1 (separate calibration at the total test level using one linking constant); u2 is unidimensional method 2 (separate calibration at the total test level using three linking constants); u3 is unidimensional method 3 (separate calibration at the subtest level); u4 is unidimensional method 4 (concurrent calibration at the total test level); u5 is unidimensional method 5 (concurrent calibration at the subtest level); m1 is multidimensional method 1(separate calibration); and m2 is multidimensional method 2 (concurrent calibration). SL refers to the Stocking & Lord method, and HB stands for the Haebara method.

In subtest 2, bias values were small when the second and third unidimensional methods using separate calibration at the total test level and at the subtest level (u2 and u3) were applied. As the mean vector of the proficiency distributions was shifted to $\mu 3 = (0, 0.1, -0.1)$, the first unidimensional method using separate calibration at the total test level with one linking constant (u1) produced the largest bias values, following the unidimensional methods using concurrent calibration at the subtest level (u5) and at the total test level (u4). Bias became larger as orrelations decreased to 0.4. When distributions were skewed ($\mu 4$ and $\mu 5$) with high skewness, multidimensional methods (m1 and m2) yielded larger bias than any unidimensional methods across all conditions. Under the distributions with mean shifts, multidimensional methods (m1 and m2) showed better results than unidimensional methods using concurrent calibration. When the distribution was $\mu 1 = (0, 0, 0)$ and the correlation was low (0.4), multidimensional methods produced results comparable to unidimensional methods.

Table 4.11. Bias compared with True Equating Function (Subtest 2)

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4
μ1	u1 HB	0.077	0.068	0.099	0.170	0.139	0.110	0.140	0.113	0.123	0.272	0.228	0.179
	u1 SL	0.044	0.041	0.081	0.138	0.105	0.087	0.094	0.062	0.092	0.196	0.153	0.131
	u2 HB	0.097	0.067	0.095	0.011	0.072	0.075	0.020	0.058	0.067	0.020	0.032	0.052
	u2 SL	0.076	0.055	0.085	-0.011	0.052	0.056	0.002	0.035	0.051	-0.008	0.003	0.033
	u3 HB	0.119	0.078	0.104	0.015	0.046	0.077	0.092	0.020	0.118	0.042	0.056	0.115
	u3 SL	0.085	0.050	0.097	-0.001	0.029	0.060	0.061	-0.008	0.109	0.033	0.034	0.096
	u4	0.068	0.049	0.083	0.111	0.091	0.082	0.084	0.064	0.086	0.211	0.139	0.116
	u5	0.102	0.060	0.089	0.070	0.071	0.077	0.072	0.053	0.099	0.077	0.050	0.099
	m1 HB	-0.202	-0.078	-0.131	-0.210	-0.093	-0.125	-0.185	-0.080	-0.112	-0.100	0.015	-0.071
	m1 SL	-0.200	-0.077	-0.132	-0.212	-0.094	-0.128	-0.189	-0.082	-0.112	-0.105	0.011	-0.074
	m2	-0.219	-0.092	-0.134	-0.201	-0.109	-0.135	-0.253	-0.087	-0.116	-0.231	-0.001	-0.080
μ2	u1 HB	0.039	0.090	0.085	0.128	0.098	0.116	0.216	0.154	0.154	0.318	0.232	0.211
	u1 SL	0.016	0.058	0.072	0.095	0.055	0.095	0.178	0.105	0.124	0.234	0.154	0.159
	u2 HB	-0.007	0.086	0.070	0.054	0.090	0.068	0.075	0.070	0.076	-0.019	0.023	0.026
	u2 SL	-0.012	0.063	0.061	0.019	0.056	0.048	0.064	0.046	0.050	-0.068	-0.008	-0.012
	u3 HB	-0.003	0.101	0.083	0.086	0.068	0.071	0.109	0.025	0.116	0.031	0.053	0.095
	u3 SL	-0.006	0.061	0.070	0.045	0.030	0.059	0.111	0.006	0.102	-0.012	0.017	0.059
	u4	0.225	0.161	0.155	0.259	0.147	0.155	0.336	0.184	0.184	0.450	0.249	0.218
	u5	0.379	0.265	0.219	0.369	0.239	0.196	0.391	0.227	0.220	0.413	0.234	0.230
	m1 HB	-0.183	-0.044	-0.119	-0.203	-0.096	-0.136	-0.164	-0.064	-0.088	-0.087	0.001	-0.048
	m1 SL	-0.184	-0.044	-0.119	-0.209	-0.096	-0.137	-0.164	-0.065	-0.089	-0.094	-0.007	-0.050
	m2	-0.190	-0.068	-0.122	-0.178	-0.127	-0.146	-0.210	-0.070	-0.095	-0.193	-0.017	-0.074
μ3	u1 HB	0.784	0.802	0.437	0.800	0.836	0.405	0.874	0.851	0.433	1.011	0.932	0.489
	u1 SL	0.777	0.775	0.431	0.787	0.793	0.390	0.838	0.782	0.405	0.912	0.825	0.437
	u2 HB	0.079	0.055	0.100	0.050	0.082	0.082	0.129	0.018	0.054	0.012	0.064	0.059
	u2 SL	0.063	0.044	0.087	0.024	0.048	0.062	0.128	-0.013	0.032	-0.005	0.018	0.028
	u3 HB	0.083	0.065	0.110	0.090	0.050	0.091	0.181	-0.020	0.112	0.046	0.060	0.139
	u3 SL	0.054	0.040	0.092	0.048	0.019	0.074	0.184	-0.059	0.094	0.045	0.025	0.123
	u4	0.711	0.600	0.350	0.666	0.593	0.309	0.735	0.596	0.320	0.878	0.685	0.363
	u5	0.443	0.250	0.234	0.365	0.223	0.202	0.418	0.202	0.219	0.418	0.240	0.250
	m1 HB	-0.150	-0.024	-0.089	-0.228	-0.110	-0.133	-0.183	-0.037	-0.101	-0.094	0.025	-0.051
	m1 SL	-0.146	-0.024	-0.089	-0.228	-0.110	-0.135	-0.186	-0.043	-0.101	-0.103	0.017	-0.054
	m2	-0.182	-0.055	-0.097	-0.175	-0.128	-0.148	-0.218	-0.052	-0.110	-0.202	0.011	-0.060
μ4	u1 HB	0.028	-0.005	0.143	0.070	0.076	0.158	0.115	0.136	0.172	0.335	0.248	0.237
	u1 SL	0.032	-0.030	0.153	0.066	0.051	0.163	0.109	0.112	0.165	0.297	0.191	0.211
	u2 HB	0.021	-0.035	0.111	0.007	-0.006	0.119	0.021	0.053	0.103	0.149	0.112	0.090
	u2 SL	0.042	-0.037	0.137	0.029	-0.007	0.132	0.061	0.077	0.111	0.189	0.128	0.093
	u3 HB	0.077	0.025	0.119	0.038	-0.001	0.134	0.036	-0.012	0.185	0.075	0.056	0.174
	u3 SL	0.113	-0.002	0.159	0.062	-0.023	0.155	0.045	-0.033	0.177	0.112	0.024	0.191
	u4	0.186	0.106	0.158	0.199	0.157	0.163	0.230	0.202	0.159	0.450	0.311	0.182
	u5	0.182	0.126	0.147	0.166	0.125	0.153	0.149	0.124	0.160	0.177	0.132	0.168
	m1 HB	0.665	0.787	0.424	0.706	0.849	0.406	0.852	1.018	0.457	0.985	1.121	0.490
	m1 SL	0.670	0.792	0.423	0.711	0.850	0.408	0.860	1.018	0.458	1.001	1.122	0.493
	m2	0.449	0.457	0.259	0.519	0.513	0.246	0.559	0.684	0.306	0.642	0.740	0.345
μ5	u1 HB	0.172	0.137	0.128	0.268	0.208	0.122	0.224	0.172	0.152	0.360	0.241	0.213
	u1 SL	0.110	0.100	0.091	0.198	0.144	0.080	0.140	0.102	0.099	0.237	0.144	0.143
	u2 HB	0.166	0.134	0.119	0.130	0.141	0.063	0.047	0.063	0.049	-0.025	0.008	0.002
	u2 SL	0.123	0.104	0.084	0.054	0.072	0.028	-0.025	0.010	-0.002	-0.112	-0.055	-0.052
	u3 HB	0.221	0.193	0.142	0.224	0.176	0.090	0.218	0.147	0.137	0.199	0.180	0.125
	u3 SL	0.160	0.137	0.094	0.142	0.110	0.043	0.140	0.095	0.078	0.111	0.136	0.068

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4
	u4	0.063	0.040	0.083	0.126	0.077	0.068	0.084	0.049	0.081	0.192	0.079	0.104
	u5	0.070	0.032	0.074	0.099	0.061	0.051	0.070	0.057	0.075	0.080	0.036	0.072
	m1 HB	-0.522	-0.419	-0.342	-0.534	-0.425	-0.341	-0.498	-0.380	-0.298	-0.381	-0.297	-0.241
	m1 SL	-0.524	-0.417	-0.343	-0.536	-0.428	-0.343	-0.508	-0.387	-0.303	-0.400	-0.314	-0.251
	m2	-0.471	-0.320	-0.267	-0.503	-0.327	-0.271	-0.560	-0.271	-0.230	-0.562	-0.183	-0.185
μ6	u1 HB	0.048	0.037	0.119	0.139	0.121	0.124	0.149	0.117	0.131	0.263	0.223	0.188
	u1 SL	0.034	0.014	0.110	0.120	0.091	0.111	0.123	0.081	0.102	0.209	0.158	0.151
	u2 HB	0.052	0.028	0.094	0.041	0.050	0.084	0.061	0.055	0.055	0.058	0.056	0.050
	u2 SL	0.055	0.015	0.092	0.047	0.032	0.081	0.073	0.049	0.036	0.057	0.035	0.043
	u3 HB	0.054	0.027	0.104	0.029	0.007	0.100	0.078	0.005	0.103	0.044	0.020	0.126
	u3 SL	0.064	0.003	0.110	0.042	-0.016	0.091	0.081	-0.014	0.091	0.028	0.001	0.120
	u4	0.084	0.039	0.109	0.113	0.097	0.107	0.128	0.095	0.093	0.237	0.158	0.123
	u5	0.102	0.043	0.111	0.064	0.071	0.098	0.090	0.067	0.097	0.087	0.058	0.115
	m1 HB	-0.019	0.086	-0.019	-0.072	0.077	-0.040	0.020	0.131	0.006	0.113	0.194	0.044
	m1 SL	-0.019	0.086	-0.021	-0.072	0.075	-0.041	0.019	0.130	0.005	0.114	0.184	0.043
	m2	-0.078	0.023	-0.052	-0.084	0.005	-0.068	-0.084	0.065	-0.035	-0.056	0.110	-0.003
μ7	u1 HB	0.130	0.112	0.101	0.193	0.149	0.117	0.196	0.141	0.133	0.302	0.216	0.184
	u1 SL	0.089	0.074	0.082	0.141	0.104	0.093	0.138	0.093	0.096	0.200	0.128	0.134
	u2 HB	0.087	0.120	0.092	0.082	0.116	0.082	0.066	0.033	0.049	0.035	-0.012	0.048
	u2 SL	0.069	0.093	0.071	0.036	0.069	0.065	0.050	0.011	0.017	-0.025	-0.046	0.015
	u3 HB	0.117	0.153	0.105	0.145	0.109	0.094	0.155	0.045	0.106	0.141	0.082	0.134
	u3 SL	0.083	0.106	0.082	0.100	0.063	0.078	0.129	0.014	0.073	0.095	0.035	0.104
	u4	0.068	0.058	0.075	0.098	0.070	0.087	0.113	0.059	0.087	0.187	0.100	0.112
	u5	0.077	0.070	0.077	0.081	0.067	0.083	0.102	0.049	0.088	0.095	0.033	0.101
	m1 HB	-0.293	-0.195	-0.209	-0.361	-0.223	-0.211	-0.291	-0.202	-0.167	-0.230	-0.106	-0.143
	m1 SL	-0.293	-0.195	-0.210	-0.364	-0.224	-0.211	-0.297	-0.208	-0.168	-0.236	-0.114	-0.143
	m2	-0.293	-0.166	-0.182	-0.335	-0.198	-0.190	-0.340	-0.173	-0.159	-0.348	-0.090	-0.133

* Note: μ1 = (0,0,0), μ2 = (0.1,0.1,0.1), μ3 = (0,0.1, −0.1), μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. u1 is unidimensional method 1 (separate calibration at the total test level using one linking onstant); u2 is unidimensional method 2 (separate calibration at the total test level using three linking constats); u3 is unidimensional method 3 (separate calibration at the subtest level); u4 is unidimensional method 4 (concurrent calibration at the total test level); u5 is unidimensional method 5 (concurrent calibration at the subtest level); m1 is multidimensional method 1(separate calibration); and m2 is multidimensional method 2 (concurrent calibration). SL refers to the Stocking & Lord method, and HB stands for the Haebara method.

As the proficiency distribution was shifted to μ3 = (0,0.1, −0.1) in subtest 3, bias values were very large when the first unidimensional method using the separate calibration at the total test level with one linking constant (u1) was used. Two multidimensional methods (m1 and m2) and two unidimensional concurrent calibration methods (u4 and u5) also produced large bias. Under the positively skewed distribution with skewness of 0.75 (μ4), two multidimensional methods (m1 and m2) showed the largest bias among all the methods although unidimensional methods also produced large bias around -.5.

Table 4.12. Bias compared with True Equating Function (Subtest 3)

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4
μ1	u1 HB	-0.195	-0.236	-0.036	-0.196	-0.262	-0.064	-0.156	-0.215	-0.003	-0.199	-0.224	0.044
	u1 SL	-0.233	-0.269	-0.058	-0.238	-0.308	-0.091	-0.217	-0.284	-0.042	-0.307	-0.323	-0.019
	u2 HB	-0.223	-0.234	-0.046	-0.130	-0.234	-0.050	-0.131	-0.207	-0.013	-0.175	-0.264	-0.003
	u2 SL	-0.277	-0.267	-0.071	-0.156	-0.267	-0.068	-0.173	-0.255	-0.032	-0.211	-0.301	-0.018
	u3 HB	-0.226	-0.205	-0.047	-0.119	-0.265	-0.064	-0.119	-0.283	-0.061	-0.111	-0.201	-0.074
	u3 SL	-0.278	-0.223	-0.073	-0.130	-0.294	-0.081	-0.165	-0.315	-0.084	-0.138	-0.221	-0.091
	u4	-0.209	-0.247	-0.050	-0.241	-0.288	-0.075	-0.227	-0.245	-0.025	-0.298	-0.243	0.014
	u5	-0.239	-0.259	-0.076	-0.213	-0.252	-0.097	-0.205	-0.243	-0.082	-0.205	-0.246	-0.092
	m1 HB	0.535	0.646	0.285	0.530	0.617	0.249	0.554	0.655	0.302	0.547	0.652	0.302
	m1 SL	0.534	0.644	0.285	0.536	0.616	0.248	0.551	0.654	0.300	0.542	0.647	0.299
	m2	-0.769	0.631	0.280	0.526	0.612	0.266	0.520	0.642	0.301	0.505	0.642	0.304
μ2	u1 HB	-0.206	-0.182	-0.048	-0.171	-0.235	-0.050	-0.094	-0.196	0.019	-0.169	-0.240	0.017
	u1 SL	-0.229	-0.219	-0.064	-0.214	-0.289	-0.075	-0.142	-0.259	-0.018	-0.286	-0.345	-0.051
	u2 HB	-0.198	-0.230	-0.069	-0.147	-0.224	-0.038	-0.095	-0.217	0.018	-0.167	-0.213	0.009
	u2 SL	-0.244	-0.271	-0.082	-0.171	-0.262	-0.046	-0.124	-0.253	0.011	-0.201	-0.256	-0.004
	u3 HB	-0.200	-0.210	-0.058	-0.115	-0.264	-0.049	-0.114	-0.303	-0.031	-0.150	-0.191	-0.069
	u3 SL	-0.246	-0.235	-0.081	-0.144	-0.299	-0.054	-0.133	-0.322	-0.033	-0.175	-0.220	-0.084
	u4	0.042	-0.099	0.038	0.011	-0.163	0.022	0.081	-0.116	0.092	0.085	-0.090	0.113
	u5	0.209	-0.032	0.096	0.195	-0.059	0.060	0.199	-0.061	0.087	0.240	-0.021	0.075
	m1 HB	0.594	0.673	0.310	0.546	0.635	0.256	0.578	0.672	0.311	0.577	0.664	0.319
	m1 SL	0.595	0.672	0.310	0.550	0.634	0.254	0.576	0.669	0.308	0.571	0.662	0.317
	m2	-0.704	0.664	0.318	0.556	0.632	0.270	0.567	0.670	0.321	0.550	0.660	0.329
μ3	u1 HB	-1.223	-1.141	-0.449	-1.161	-1.126	-0.443	-1.125	-1.139	-0.389	-1.094	-1.110	-0.297
	u1 SL	-1.213	-1.179	-0.453	-1.168	-1.187	-0.459	-1.159	-1.237	-0.421	-1.227	-1.269	-0.362
	u2 HB	-0.173	-0.205	-0.028	-0.172	-0.203	-0.083	-0.231	-0.253	0.007	-0.261	-0.370	0.021
	u2 SL	-0.223	-0.243	-0.051	-0.194	-0.227	-0.101	-0.271	-0.289	-0.010	-0.302	-0.394	0.003
	u3 HB	-0.163	-0.180	-0.030	-0.149	-0.227	-0.093	-0.162	-0.286	-0.036	-0.095	-0.230	-0.056
	u3 SL	-0.216	-0.207	-0.054	-0.167	-0.245	-0.114	-0.201	-0.300	-0.052	-0.151	-0.250	-0.075
	u4	-1.065	-0.897	-0.334	-1.064	-0.904	-0.350	-1.064	-0.919	-0.297	-0.999	-0.834	-0.205
	u5	-0.640	-0.454	-0.213	-0.666	-0.444	-0.265	-0.684	-0.473	-0.235	-0.589	-0.473	-0.231
	m1 HB	0.524	0.589	0.286	0.473	0.557	0.251	0.532	0.603	0.279	0.500	0.583	0.296
	m1 SL	0.524	0.591	0.286	0.476	0.559	0.252	0.527	0.602	0.277	0.495	0.581	0.294
	m2	-1.862	0.571	0.267	0.477	0.547	0.244	0.473	0.574	0.257	0.443	0.568	0.281
μ4	u1 HB	-0.599	-0.624	-0.108	-0.630	-0.613	-0.133	-0.498	-0.530	-0.098	-0.474	-0.555	-0.048
	u1 SL	-0.572	-0.634	-0.098	-0.621	-0.627	-0.129	-0.495	-0.544	-0.108	-0.511	-0.599	-0.088
	u2 HB	-0.404	-0.497	-0.047	-0.472	-0.468	-0.031	-0.348	-0.397	0.012	-0.338	-0.334	0.043
	u2 SL	-0.423	-0.528	-0.051	-0.495	-0.494	-0.034	-0.376	-0.429	0.008	-0.347	-0.355	0.042
	u3 HB	-0.521	-0.544	-0.003	-0.595	-0.609	-0.003	-0.533	-0.631	0.017	-0.465	-0.504	0.024
	u3 SL	-0.556	-0.580	-0.027	-0.645	-0.653	-0.023	-0.582	-0.682	-0.006	-0.525	-0.558	0.002
	u4	-0.427	-0.423	-0.015	-0.492	-0.427	-0.031	-0.394	-0.346	0.001	-0.418	-0.295	0.058
	u5	-0.699	-0.564	-0.079	-0.768	-0.574	-0.093	-0.705	-0.567	-0.080	-0.663	-0.545	-0.075
	m1 HB	1.449	1.582	0.878	1.544	1.628	0.845	1.610	1.722	0.876	1.609	1.729	0.879
	m1 SL	1.448	1.586	0.880	1.548	1.629	0.843	1.612	1.728	0.879	1.610	1.733	0.876
	m2	-0.184	1.242	0.707	1.384	1.295	0.679	1.426	1.360	0.721	1.415	1.387	0.743

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4
μ5	u1 HB	0.071	-0.036	0.048	0.009	-0.049	0.011	0.063	-0.020	0.069	0.035	-0.003	0.122
	u1 SL	-0.007	-0.086	0.005	-0.089	-0.134	-0.040	-0.054	-0.115	0.003	-0.136	-0.135	0.030
	u2 HB	-0.040	-0.116	-0.005	-0.099	-0.125	-0.039	-0.095	-0.120	0.009	-0.111	-0.206	0.034
	u2 SL	-0.118	-0.158	-0.044	-0.137	-0.170	-0.070	-0.151	-0.173	-0.011	-0.163	-0.252	0.016
	u3 HB	0.023	0.001	-0.004	-0.001	-0.071	-0.054	-0.003	-0.078	-0.034	0.085	-0.024	-0.039
	u3 SL	-0.046	-0.029	-0.043	-0.045	-0.104	-0.085	-0.057	-0.108	-0.053	0.031	-0.050	-0.063
	u4	-0.059	-0.164	-0.031	-0.143	-0.203	-0.061	-0.097	-0.166	-0.008	-0.173	-0.182	0.047
	u5	0.011	-0.134	-0.052	-0.025	-0.146	-0.090	-0.021	-0.102	-0.067	0.034	-0.128	-0.083
	m1 HB	0.261	0.320	0.069	0.220	0.307	0.063	0.275	0.357	0.089	0.289	0.362	0.114
	m1 SL	0.260	0.317	0.069	0.222	0.305	0.063	0.267	0.351	0.088	0.278	0.351	0.113
	m2	-0.940	0.416	0.135	0.232	0.418	0.138	0.228	0.451	0.157	0.200	0.463	0.191
μ6	u1 HB	-0.298	-0.283	-0.064	-0.273	-0.341	-0.075	-0.239	-0.286	-0.024	-0.250	-0.287	0.007
	u1 SL	-0.307	-0.308	-0.074	-0.298	-0.380	-0.091	-0.274	-0.334	-0.060	-0.328	-0.374	-0.043
	u2 HB	-0.229	-0.262	-0.031	-0.173	-0.260	-0.048	-0.178	-0.236	0.004	-0.319	-0.302	-0.005
	u2 SL	-0.262	-0.287	-0.042	-0.203	-0.293	-0.060	-0.206	-0.262	-0.008	-0.334	-0.335	-0.016
	u3 HB	-0.253	-0.266	-0.026	-0.196	-0.335	-0.055	-0.210	-0.361	-0.043	-0.304	-0.301	-0.065
	u3 SL	-0.291	-0.278	-0.044	-0.235	-0.371	-0.071	-0.240	-0.372	-0.052	-0.335	-0.328	-0.076
	u4	-0.259	-0.256	-0.045	-0.292	-0.321	-0.065	-0.273	-0.275	-0.020	-0.331	-0.271	0.009
	u5	-0.313	-0.300	-0.064	-0.328	-0.325	-0.093	-0.311	-0.302	-0.076	-0.352	-0.323	-0.095
	m1 HB	0.675	0.766	0.356	0.686	0.768	0.376	0.708	0.797	0.385	0.717	0.804	0.404
	m1 SL	0.671	0.766	0.356	0.683	0.770	0.373	0.705	0.797	0.384	0.707	0.800	0.400
	m2	-0.665	0.708	0.337	0.656	0.719	0.347	0.654	0.741	0.353	0.654	0.742	0.379
μ7	u1 HB	-0.096	-0.167	0.002	-0.097	-0.154	-0.046	-0.077	-0.149	0.013	-0.101	-0.166	0.075
	u1 SL	-0.147	-0.217	-0.021	-0.165	-0.216	-0.076	-0.157	-0.217	-0.033	-0.242	-0.283	0.009
	u2 HB	-0.079	-0.171	-0.033	-0.108	-0.153	-0.044	-0.085	-0.155	-0.013	-0.197	-0.230	0.020
	u2 SL	-0.151	-0.217	-0.048	-0.148	-0.182	-0.056	-0.147	-0.186	-0.031	-0.242	-0.266	0.004
	u3 HB	-0.054	-0.128	-0.037	-0.041	-0.166	-0.061	-0.016	-0.175	-0.049	-0.106	-0.104	-0.048
	u3 SL	-0.125	-0.159	-0.050	-0.093	-0.184	-0.070	-0.079	-0.199	-0.069	-0.137	-0.131	-0.070
	u4	-0.159	-0.225	-0.034	-0.189	-0.232	-0.070	-0.173	-0.219	-0.022	-0.249	-0.221	0.035
	u5	-0.123	-0.221	-0.054	-0.136	-0.193	-0.092	-0.104	-0.175	-0.067	-0.149	-0.180	-0.074
	m1 HB	0.454	0.525	0.216	0.429	0.490	0.175	0.461	0.516	0.224	0.445	0.530	0.241
	m1 SL	0.453	0.523	0.216	0.434	0.490	0.174	0.458	0.512	0.223	0.439	0.523	0.235
	m2	-0.810	0.553	0.241	0.431	0.541	0.206	0.429	0.557	0.241	0.404	0.555	0.257

* Note: $\mu 1 = (0,0,0)$, $\mu 2 = (0.1,0.1,0.1)$, $\mu 3 = (0,0.1,-0.1)$, μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. u1 is unidimensional method 1 (separate calibration at the total test level using one linking onstant); u2 is unidimensional method 2 (separate calibration at the total test level using three linking constats); u3 is unidimensional method 3 (separate calibration at the subtest level); u4 is unidimensional method 4 (concurrent calibration at the total test level); u5 is unidimensional method 5 (concurrent calibration at the subtest level); m1 is multidimensional method 1(separate calibration); and m2 is multidimensional method 2 (concurrent calibration). SL refers to the Stocking & Lord method, and HB stands for the Haebara method.

4.4.2.2 Absolute difference

Table 4.13, Table 4.14, and Table 4.15 present absolute differences between true and estimated equating results for subtests 1, 2, and 3, respectively. In subtest 1, Haebara slightly outperformed Stocking and Lord. When the unidimensional method using the separate calibration at the subtest level using one linking constant (u1) was employed, values of absolute differences increased as the correlation dropped to 0.4. Under the positively skewed distribution with a skewness of .75 (μ4), the pattern was unclear; values were all elevated across the methods and conditions. The unidimensional method at the total test level using three linking constants (u2) yielded smaller values than the one at the subtest level (u3) throughout the conditions. The first unidimensional method (u1) showed better results than the third unidimensional method (u3) when the correlations were high. With low correlations, the method at the subtest level had smaller absolute differences. When correlations were high, the unidimensional method based on concurrent calibration produced smaller errors than the methods from separate calibrations. The concurrent calibration at the total test level (u4) had smaller absolute differences than the separate calibration methods (u1, u2, and u3) in most cases except for the highly skewed distributions (μ4 and μ5). However, the concurrent calibration at the subtest level (u5) showed smaller errors than the other methods when the distribution was highly skewed (μ4). When the correlation among dimensions was low (0.4), the methods based on separate calibrations (u2 and u3) outperformed the one based on concurrent calibration at the subtest level (u5). In general, multidimensional methods did not perform well compared with unidimensional methods.

Table 4.13. Absolute Difference compared with True Equating Function (Subtest 1)

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4
μ1	u1 HB	0.435	0.375	0.248	0.474	0.443	0.285	0.504	0.445	0.266	0.710	0.680	0.398
	u1 SL	0.468	0.384	0.253	0.496	0.463	0.303	0.538	0.469	0.290	0.753	0.736	0.434
	u2 HB	0.527	0.409	0.279	0.485	0.389	0.288	0.512	0.351	0.259	0.534	0.346	0.244
	u2 SL	0.531	0.415	0.280	0.491	0.392	0.296	0.517	0.368	0.263	0.534	0.357	0.252
	u3 HB	0.553	0.612	0.300	0.516	0.579	0.308	0.520	0.641	0.272	0.546	0.397	0.275
	u3 SL	0.559	0.621	0.299	0.513	0.582	0.315	0.518	0.643	0.284	0.557	0.394	0.281
	u4	0.395	0.332	0.229	0.461	0.405	0.266	0.453	0.371	0.237	0.651	0.530	0.334
	u5	0.413	0.355	0.245	0.422	0.392	0.265	0.390	0.354	0.231	0.387	0.341	0.234
	m1 HB	0.614	0.661	0.414	0.523	0.565	0.374	0.565	0.629	0.354	0.606	0.654	0.357
	m1 SL	0.607	0.659	0.414	0.521	0.565	0.376	0.562	0.624	0.354	0.607	0.654	0.359
	m2	0.603	0.642	0.401	0.565	0.558	0.365	0.606	0.617	0.357	0.606	0.633	0.342
μ2	u1 HB	0.471	0.408	0.295	0.487	0.427	0.281	0.552	0.457	0.297	0.714	0.696	0.404
	u1 SL	0.505	0.441	0.305	0.499	0.462	0.298	0.574	0.488	0.320	0.749	0.741	0.444
	u2 HB	0.541	0.425	0.289	0.541	0.373	0.279	0.488	0.386	0.288	0.589	0.347	0.243
	u2 SL	0.557	0.437	0.290	0.548	0.384	0.284	0.497	0.395	0.292	0.594	0.367	0.248
	u3 HB	0.598	0.628	0.303	0.570	0.616	0.289	0.510	0.662	0.302	0.606	0.411	0.279
	u3 SL	0.610	0.631	0.304	0.588	0.630	0.300	0.525	0.669	0.306	0.583	0.428	0.279
	u4	0.344	0.337	0.232	0.384	0.354	0.225	0.429	0.363	0.230	0.585	0.512	0.281
	u5	0.405	0.352	0.235	0.371	0.324	0.223	0.358	0.331	0.203	0.424	0.336	0.210
	m1 HB	0.626	0.694	0.441	0.525	0.564	0.376	0.519	0.583	0.339	0.567	0.626	0.339
	m1 SL	0.624	0.693	0.441	0.527	0.561	0.374	0.523	0.585	0.339	0.573	0.630	0.341
	m2	0.641	0.677	0.429	0.599	0.567	0.369	0.611	0.582	0.334	0.616	0.605	0.325
μ3	u1 HB	0.440	0.369	0.259	0.463	0.399	0.272	0.492	0.432	0.275	0.668	0.668	0.383
	u1 SL	0.452	0.377	0.262	0.478	0.412	0.285	0.522	0.470	0.298	0.736	0.735	0.421
	u2 HB	0.570	0.421	0.274	0.493	0.374	0.255	0.479	0.348	0.265	0.450	0.360	0.248
	u2 SL	0.576	0.425	0.274	0.501	0.385	0.265	0.485	0.357	0.269	0.448	0.380	0.252
	u3 HB	0.602	0.610	0.299	0.508	0.581	0.279	0.525	0.638	0.293	0.501	0.431	0.287
	u3 SL	0.615	0.606	0.296	0.524	0.590	0.291	0.536	0.649	0.290	0.516	0.434	0.293
	u4	0.405	0.342	0.226	0.455	0.374	0.253	0.462	0.374	0.239	0.601	0.529	0.304
	u5	0.420	0.367	0.238	0.429	0.400	0.256	0.430	0.381	0.243	0.389	0.374	0.254
	m1 HB	0.614	0.653	0.430	0.510	0.576	0.354	0.522	0.591	0.355	0.556	0.617	0.346
	m1 SL	0.610	0.651	0.429	0.510	0.578	0.355	0.525	0.589	0.356	0.557	0.618	0.346
	m2	0.605	0.636	0.423	0.569	0.559	0.342	0.598	0.582	0.347	0.600	0.584	0.330
μ4	u1 HB	0.855	0.695	0.700	0.757	0.611	0.618	0.678	0.534	0.500	0.810	0.710	0.453
	u1 SL	0.839	0.712	0.692	0.762	0.634	0.614	0.681	0.536	0.507	0.809	0.697	0.462
	u2 HB	0.880	0.712	0.683	0.754	0.599	0.627	0.690	0.495	0.519	0.728	0.470	0.416
	u2 SL	0.878	0.736	0.677	0.761	0.621	0.625	0.689	0.503	0.521	0.730	0.486	0.414
	u3 HB	1.311	1.169	0.829	1.269	1.175	0.868	1.255	1.184	0.848	1.319	1.092	0.831
	u3 SL	1.301	1.199	0.820	1.273	1.210	0.863	1.260	1.216	0.855	1.304	1.121	0.824
	u4	0.819	0.533	0.579	0.728	0.472	0.528	0.714	0.480	0.445	0.943	0.704	0.433
	u5	1.327	0.942	0.759	1.293	0.914	0.800	1.274	0.921	0.768	1.310	0.922	0.764
	m1 HB	1.677	1.726	1.142	1.643	1.691	1.131	1.872	1.937	1.168	1.868	1.953	1.142
	m1 SL	1.672	1.715	1.140	1.633	1.689	1.132	1.860	1.921	1.162	1.863	1.937	1.145
	m2	1.400	1.354	0.938	1.440	1.311	0.931	1.561	1.531	0.964	1.608	1.526	0.956
μ5	u1 HB	0.416	0.410	0.236	0.497	0.485	0.267	0.574	0.565	0.312	0.865	0.907	0.547
	u1 SL	0.447	0.406	0.249	0.521	0.507	0.281	0.620	0.601	0.342	0.961	0.984	0.602
	u2 HB	0.537	0.398	0.250	0.489	0.374	0.267	0.451	0.331	0.232	0.533	0.347	0.226
	u2 SL	0.540	0.385	0.249	0.480	0.373	0.263	0.446	0.321	0.231	0.532	0.350	0.226
	u3 HB	0.565	0.614	0.275	0.503	0.603	0.299	0.477	0.693	0.263	0.558	0.418	0.259
	u3 SL	0.561	0.604	0.272	0.496	0.597	0.290	0.472	0.680	0.258	0.550	0.410	0.259

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4
	u4	0.472	0.422	0.252	0.585	0.521	0.295	0.624	0.549	0.314	0.943	0.783	0.476
	u5	0.584	0.533	0.307	0.621	0.618	0.336	0.570	0.553	0.297	0.575	0.556	0.317
	m1 HB	0.478	0.521	0.292	0.470	0.512	0.279	0.483	0.526	0.258	0.624	0.647	0.312
	m1 SL	0.485	0.527	0.295	0.481	0.521	0.282	0.493	0.532	0.263	0.645	0.670	0.323
	m2	0.430	0.531	0.304	0.405	0.524	0.293	0.417	0.539	0.276	0.403	0.631	0.295
μ6	u1 HB	0.481	0.405	0.298	0.448	0.404	0.303	0.504	0.411	0.292	0.672	0.615	0.397
	u1 SL	0.483	0.424	0.306	0.467	0.432	0.315	0.520	0.442	0.316	0.723	0.661	0.415
	u2 HB	0.584	0.413	0.331	0.508	0.395	0.318	0.530	0.359	0.286	0.497	0.365	0.293
	u2 SL	0.600	0.423	0.336	0.503	0.401	0.320	0.534	0.367	0.290	0.503	0.360	0.290
	u3 HB	0.632	0.651	0.355	0.569	0.594	0.363	0.592	0.666	0.347	0.576	0.398	0.338
	u3 SL	0.646	0.664	0.360	0.560	0.600	0.369	0.608	0.682	0.355	0.590	0.414	0.344
	u4	0.425	0.334	0.265	0.433	0.354	0.267	0.432	0.333	0.249	0.607	0.470	0.331
	u5	0.472	0.357	0.297	0.474	0.380	0.307	0.442	0.344	0.288	0.438	0.312	0.305
	m1 HB	0.746	0.807	0.524	0.641	0.684	0.470	0.697	0.765	0.468	0.711	0.752	0.456
	m1 SL	0.741	0.805	0.525	0.635	0.681	0.471	0.692	0.758	0.466	0.705	0.744	0.455
	m2	0.706	0.747	0.489	0.706	0.640	0.431	0.708	0.703	0.439	0.736	0.696	0.411
μ7	u1 HB	0.421	0.383	0.232	0.435	0.425	0.272	0.513	0.479	0.277	0.709	0.760	0.424
	u1 SL	0.450	0.392	0.238	0.468	0.451	0.281	0.545	0.501	0.304	0.770	0.829	0.461
	u2 HB	0.505	0.389	0.252	0.433	0.351	0.285	0.480	0.354	0.265	0.522	0.359	0.243
	u2 SL	0.515	0.390	0.256	0.441	0.360	0.292	0.472	0.353	0.270	0.527	0.357	0.246
	u3 HB	0.537	0.584	0.257	0.448	0.554	0.286	0.498	0.613	0.260	0.543	0.375	0.232
	u3 SL	0.531	0.583	0.255	0.446	0.556	0.296	0.502	0.608	0.262	0.555	0.367	0.230
	u4	0.410	0.372	0.217	0.465	0.408	0.258	0.505	0.419	0.251	0.697	0.611	0.356
	u5	0.432	0.411	0.217	0.421	0.430	0.257	0.427	0.399	0.223	0.405	0.399	0.223
	m1 HB	0.538	0.579	0.355	0.480	0.531	0.313	0.504	0.561	0.308	0.606	0.638	0.319
	m1 SL	0.539	0.579	0.354	0.483	0.534	0.314	0.504	0.562	0.309	0.617	0.642	0.323
	m2	0.524	0.587	0.365	0.513	0.541	0.317	0.523	0.572	0.321	0.519	0.628	0.304

* Note: μ1 = (0,0,0), μ2 = (0.1,0.1,0.1), μ3 = (0,0.1, −0.1), μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. u1 is unidimensional method 1 (separate calibration at the total test level using one linking onstant); u2 is unidimensional method 2 (separate calibration at the total test level using three linking constats); u3 is unidimensional method 3 (separate calibration at the subtest level); u4 is unidimensional method 4 (concurrent calibration at the total test level); u5 is unidimensional method 5 (concurrent calibration at the subtest level); m1 is multidimensional method 1(separate calibration); and m2 is multidimensional method 2 (concurrent calibration). SL refers to the Stocking & Lord method, and HB stands for the Haebara method.

Table 4.14 presents absolute differences between true and estimated results from subtest 2. The unidimensional method using concurrent calibration at the total test level (u4) outperformed the other unidimensional methods when the distribution was μ1=(0,0,0). The unidimensional method using concurrent calibration at the subtest level (u5) produced smaller absolute differences than unidimensional methods through separate calibrations (u1, u2, and u3) except for several cases especially when the correlation was 0.4. Multidimensional methods produced smaller absolute differences regardless of correlations especially when the number of items was 16. When the distributions were skewed, however, this was not always true. Under the positively skewed distribution (μ4), multidimensional methods using separate calibration yielded the largest absolute differences.

Table 4.14. Absolute Difference compared with True Equating Function (Subtest 2)

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4
μ1	u1 HB	0.452	0.425	0.256	0.448	0.412	0.252	0.508	0.434	0.293	0.545	0.494	0.359
	u1 SL	0.455	0.428	0.249	0.448	0.406	0.240	0.499	0.426	0.274	0.508	0.472	0.334
	u2 HB	0.554	0.438	0.273	0.534	0.431	0.263	0.546	0.437	0.304	0.521	0.457	0.330
	u2 SL	0.550	0.429	0.268	0.525	0.427	0.256	0.554	0.424	0.302	0.540	0.453	0.325
	u3 HB	0.539	0.624	0.283	0.569	0.551	0.284	0.580	0.546	0.290	0.540	0.442	0.280
	u3 SL	0.544	0.614	0.265	0.540	0.550	0.273	0.582	0.550	0.278	0.531	0.436	0.272
	u4	0.411	0.384	0.232	0.414	0.372	0.226	0.455	0.381	0.229	0.467	0.391	0.249
	u5	0.428	0.392	0.229	0.436	0.384	0.245	0.459	0.389	0.221	0.427	0.397	0.227
	m1 HB	0.423	0.412	0.220	0.436	0.427	0.230	0.453	0.446	0.212	0.530	0.535	0.230
	m1 SL	0.426	0.411	0.220	0.438	0.429	0.231	0.458	0.446	0.213	0.534	0.535	0.231
	m2	0.254	0.402	0.213	0.250	0.409	0.232	0.285	0.443	0.210	0.270	0.524	0.227
μ2	u1 HB	0.448	0.417	0.248	0.469	0.404	0.266	0.524	0.452	0.308	0.612	0.536	0.394
	u1 SL	0.464	0.419	0.244	0.458	0.399	0.256	0.504	0.431	0.290	0.561	0.513	0.365
	u2 HB	0.525	0.438	0.256	0.570	0.400	0.274	0.548	0.425	0.311	0.562	0.432	0.352
	u2 SL	0.547	0.432	0.254	0.568	0.413	0.272	0.548	0.411	0.301	0.577	0.432	0.350
	u3 HB	0.549	0.617	0.261	0.607	0.549	0.276	0.551	0.540	0.282	0.566	0.437	0.286
	u3 SL	0.563	0.613	0.261	0.593	0.573	0.271	0.565	0.550	0.276	0.588	0.440	0.288
	u4	0.452	0.386	0.270	0.483	0.387	0.269	0.538	0.403	0.292	0.629	0.445	0.324
	u5	0.541	0.434	0.309	0.551	0.427	0.297	0.545	0.425	0.301	0.568	0.418	0.306
	m1 HB	0.407	0.408	0.221	0.373	0.386	0.229	0.429	0.426	0.217	0.484	0.478	0.218
	m1 SL	0.409	0.408	0.221	0.377	0.387	0.230	0.432	0.426	0.219	0.485	0.479	0.220
	m2	0.243	0.409	0.220	0.228	0.393	0.229	0.268	0.420	0.219	0.259	0.476	0.216
μ3	u1 HB	0.845	0.836	0.491	0.868	0.870	0.470	0.962	0.922	0.535	1.096	1.005	0.620
	u1 SL	0.842	0.809	0.488	0.862	0.831	0.460	0.936	0.861	0.514	1.014	0.914	0.576
	u2 HB	0.533	0.405	0.283	0.537	0.419	0.289	0.563	0.466	0.317	0.600	0.441	0.362
	u2 SL	0.535	0.414	0.280	0.543	0.421	0.278	0.557	0.462	0.313	0.601	0.435	0.357
	u3 HB	0.526	0.597	0.283	0.564	0.554	0.291	0.551	0.548	0.282	0.541	0.442	0.298
	u3 SL	0.545	0.604	0.277	0.578	0.567	0.283	0.548	0.556	0.275	0.552	0.440	0.282
	u4	0.769	0.642	0.405	0.741	0.641	0.376	0.831	0.675	0.405	0.960	0.739	0.446
	u5	0.570	0.418	0.313	0.527	0.424	0.291	0.553	0.408	0.299	0.569	0.420	0.314
	m1 HB	0.419	0.391	0.204	0.425	0.390	0.234	0.432	0.425	0.206	0.483	0.494	0.215
	m1 SL	0.421	0.394	0.204	0.424	0.390	0.234	0.437	0.426	0.209	0.486	0.495	0.216
	m2	0.251	0.391	0.201	0.240	0.394	0.234	0.271	0.415	0.205	0.245	0.482	0.214
μ4	u1 HB	0.515	0.515	0.341	0.479	0.463	0.313	0.470	0.478	0.285	0.563	0.493	0.387
	u1 SL	0.523	0.538	0.342	0.493	0.478	0.315	0.468	0.478	0.281	0.542	0.470	0.373
	u2 HB	0.585	0.525	0.366	0.579	0.484	0.335	0.497	0.458	0.294	0.535	0.424	0.323

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4
	u2 SL	0.580	0.529	0.357	0.577	0.489	0.336	0.508	0.445	0.294	0.531	0.423	0.320
	u3 HB	0.800	0.856	0.420	0.795	0.805	0.409	0.757	0.822	0.428	0.819	0.785	0.441
	u3 SL	0.779	0.863	0.407	0.791	0.816	0.406	0.766	0.822	0.422	0.798	0.783	0.425
	u4	0.418	0.340	0.261	0.427	0.329	0.242	0.419	0.374	0.237	0.589	0.434	0.281
	u5	0.651	0.529	0.311	0.650	0.503	0.303	0.610	0.542	0.306	0.655	0.529	0.327
	m1 HB	0.787	0.896	0.435	0.809	0.938	0.419	0.992	1.134	0.472	1.148	1.265	0.515
	m1 SL	0.791	0.899	0.435	0.813	0.938	0.422	0.996	1.134	0.474	1.155	1.263	0.518
	m2	0.508	0.626	0.295	0.565	0.652	0.287	0.601	0.847	0.336	0.676	0.940	0.389
μ5	u1 HB	0.573	0.496	0.376	0.606	0.530	0.371	0.613	0.527	0.421	0.641	0.549	0.443
	u1 SL	0.557	0.483	0.354	0.570	0.502	0.341	0.577	0.500	0.388	0.572	0.511	0.399
	u2 HB	0.656	0.528	0.384	0.677	0.566	0.401	0.676	0.542	0.444	0.623	0.511	0.408
	u2 SL	0.649	0.503	0.369	0.660	0.547	0.380	0.668	0.533	0.425	0.645	0.522	0.405
	u3 HB	0.729	0.713	0.426	0.731	0.702	0.442	0.705	0.680	0.449	0.680	0.581	0.413
	u3 SL	0.701	0.676	0.396	0.688	0.678	0.411	0.674	0.658	0.415	0.660	0.545	0.382
	u4	0.581	0.491	0.327	0.585	0.503	0.316	0.591	0.479	0.336	0.575	0.473	0.308
	u5	0.708	0.598	0.382	0.700	0.606	0.390	0.694	0.584	0.386	0.685	0.593	0.373
	m1 HB	0.582	0.515	0.357	0.608	0.545	0.360	0.597	0.532	0.320	0.660	0.624	0.307
	m1 SL	0.583	0.512	0.358	0.607	0.544	0.362	0.601	0.532	0.324	0.661	0.624	0.310
	m2	0.475	0.455	0.293	0.506	0.497	0.300	0.564	0.491	0.272	0.565	0.586	0.276
μ6	u1 HB	0.449	0.391	0.232	0.424	0.385	0.241	0.456	0.412	0.273	0.533	0.486	0.353
	u1 SL	0.451	0.400	0.229	0.416	0.377	0.232	0.445	0.403	0.257	0.503	0.464	0.334
	u2 HB	0.515	0.412	0.248	0.492	0.394	0.251	0.522	0.418	0.284	0.521	0.419	0.294
	u2 SL	0.515	0.414	0.246	0.475	0.386	0.242	0.525	0.417	0.283	0.523	0.422	0.288
	u3 HB	0.527	0.585	0.249	0.526	0.541	0.249	0.541	0.527	0.255	0.534	0.413	0.250
	u3 SL	0.522	0.593	0.247	0.498	0.547	0.239	0.556	0.536	0.265	0.527	0.430	0.244
	u4	0.383	0.344	0.207	0.365	0.322	0.205	0.405	0.351	0.222	0.468	0.366	0.249
	u5	0.381	0.348	0.204	0.394	0.339	0.202	0.374	0.351	0.192	0.376	0.334	0.197
	m1 HB	0.404	0.428	0.193	0.402	0.438	0.213	0.442	0.478	0.204	0.530	0.564	0.234
	m1 SL	0.409	0.429	0.195	0.406	0.441	0.214	0.449	0.478	0.207	0.533	0.565	0.235
	m2	0.211	0.401	0.191	0.204	0.407	0.210	0.226	0.451	0.200	0.213	0.534	0.223
μ7	u1 HB	0.455	0.425	0.300	0.500	0.428	0.289	0.500	0.483	0.329	0.597	0.497	0.380
	u1 SL	0.449	0.414	0.287	0.481	0.411	0.274	0.486	0.463	0.306	0.547	0.469	0.350
	u2 HB	0.553	0.449	0.310	0.596	0.452	0.312	0.575	0.468	0.328	0.575	0.452	0.345
	u2 SL	0.564	0.431	0.298	0.577	0.446	0.302	0.580	0.464	0.320	0.581	0.463	0.338
	u3 HB	0.576	0.619	0.322	0.634	0.585	0.321	0.601	0.610	0.314	0.581	0.482	0.317
	u3 SL	0.575	0.615	0.305	0.598	0.583	0.306	0.599	0.604	0.299	0.564	0.476	0.309
	u4	0.438	0.395	0.262	0.482	0.397	0.252	0.463	0.426	0.258	0.511	0.407	0.276
	u5	0.492	0.433	0.272	0.519	0.448	0.279	0.500	0.459	0.269	0.491	0.441	0.277
	m1 HB	0.441	0.426	0.260	0.500	0.454	0.274	0.488	0.456	0.246	0.564	0.544	0.251
	m1 SL	0.442	0.426	0.261	0.500	0.455	0.274	0.491	0.456	0.247	0.566	0.544	0.250
	m2	0.313	0.407	0.234	0.341	0.429	0.257	0.355	0.443	0.235	0.355	0.533	0.249

* Note: $\mu1 = (0,0,0)$, $\mu2 = (0.1,0.1,0.1)$, $\mu3 = (0,0.1,-0.1)$, μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. u1 is unidimensional method 1 (separate calibration at the total test level using one linking onstant); u2 is unidimensional method 2 (separate calibration at the total test level using three linking constats); u3 is unidimensional method 3 (separate calibration at the subtest level); u4 is unidimensional method 4 (concurrent calibration at the total test level); u5 is unidimensional method 5 (concurrent calibration at the subtest level); m1 is multidimensional method 1(separate calibration); and m2 is multidimensional method 2 (concurrent calibration). SL refers to the Stocking & Lord method, and HB stands for the Haebara method.

Table 4.15 displays absolute difference values from subtest 3. Subtest 3 showed a similar pattern with subtest 1. In addition, when the proficiency distributions were $\mu3=(0,0.1,-0.1)$ or positively skewed ($\mu4$), absolute differences were large. Under $\mu3$, the values were especially large when the first unidimensional method of using separate calibration at the total test level with one linking constant (u1) and the fifth unidimensional method of using concurrent calibration at the subtest level (u5) were implemented. In most cases, multidimensional methods did not perform well compared to unidimensional methods. In some cases in which $\mu3=(0,0.1,-0.1)$ or with negatively skewed distributions ($\mu5$ and $\mu7$), however, multidimensional methods had smaller error than unidimensional methods.

Table 4.15. Absolute Difference compared with True Equating Function (Subtest 3)

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4
μ1	u1 HB	0.357	0.330	0.238	0.396	0.381	0.253	0.443	0.443	0.309	0.700	0.678	0.446
	u1 SL	0.372	0.344	0.239	0.401	0.403	0.262	0.473	0.471	0.318	0.745	0.724	0.465
	u2 HB	0.483	0.362	0.256	0.445	0.341	0.275	0.452	0.379	0.302	0.610	0.531	0.403
	u2 SL	0.503	0.372	0.263	0.457	0.364	0.274	0.471	0.394	0.311	0.616	0.546	0.410
	u3 HB	0.503	0.582	0.256	0.479	0.540	0.276	0.473	0.540	0.266	0.509	0.341	0.271
	u3 SL	0.536	0.587	0.262	0.490	0.540	0.276	0.481	0.538	0.277	0.519	0.358	0.279
	u4	0.325	0.303	0.228	0.367	0.355	0.249	0.461	0.391	0.306	0.868	0.650	0.433
	u5	0.380	0.331	0.233	0.365	0.318	0.245	0.363	0.299	0.230	0.363	0.324	0.238
	m1 HB	0.557	0.656	0.306	0.546	0.632	0.276	0.577	0.668	0.323	0.591	0.681	0.333
	m1 SL	0.556	0.654	0.305	0.550	0.632	0.275	0.577	0.668	0.323	0.587	0.677	0.330
	m2	0.775	0.639	0.297	0.526	0.623	0.279	0.520	0.654	0.316	0.505	0.666	0.327
μ2	u1 HB	0.389	0.304	0.234	0.389	0.354	0.269	0.432	0.450	0.325	0.704	0.652	0.417
	u1 SL	0.392	0.322	0.235	0.397	0.383	0.270	0.448	0.474	0.334	0.740	0.701	0.447
	u2 HB	0.476	0.348	0.266	0.466	0.338	0.288	0.488	0.379	0.307	0.560	0.492	0.390
	u2 SL	0.508	0.365	0.268	0.496	0.364	0.290	0.490	0.391	0.308	0.559	0.502	0.392
	u3 HB	0.515	0.612	0.275	0.496	0.565	0.279	0.503	0.550	0.260	0.488	0.344	0.268
	u3 SL	0.546	0.608	0.277	0.519	0.572	0.278	0.521	0.553	0.260	0.496	0.357	0.270
	u4	0.326	0.252	0.207	0.321	0.290	0.227	0.421	0.362	0.291	0.705	0.552	0.375
	u5	0.389	0.272	0.218	0.374	0.249	0.216	0.366	0.247	0.219	0.384	0.241	0.207
	m1 HB	0.612	0.686	0.324	0.556	0.642	0.279	0.598	0.682	0.326	0.611	0.679	0.338
	m1 SL	0.613	0.685	0.324	0.559	0.640	0.277	0.595	0.679	0.324	0.607	0.675	0.335
	m2	0.710	0.674	0.327	0.556	0.638	0.280	0.567	0.679	0.330	0.550	0.676	0.344
μ3	u1 HB	1.225	1.143	0.524	1.166	1.129	0.548	1.156	1.166	0.585	1.291	1.290	0.681
	u1 SL	1.215	1.180	0.526	1.172	1.189	0.561	1.186	1.259	0.611	1.403	1.431	0.730
	u2 HB	0.464	0.350	0.247	0.497	0.334	0.293	0.480	0.394	0.317	0.644	0.629	0.444
	u2 SL	0.477	0.355	0.260	0.499	0.342	0.302	0.499	0.409	0.326	0.666	0.640	0.454
	u3 HB	0.483	0.589	0.255	0.524	0.569	0.273	0.450	0.528	0.254	0.488	0.359	0.243
	u3 SL	0.493	0.592	0.266	0.531	0.578	0.279	0.475	0.529	0.264	0.501	0.372	0.254
	u4	1.065	0.898	0.418	1.069	0.908	0.460	1.122	0.948	0.494	1.470	1.083	0.618
	u5	0.663	0.479	0.299	0.687	0.468	0.348	0.689	0.481	0.313	0.608	0.484	0.308
	m1 HB	0.541	0.597	0.309	0.507	0.584	0.272	0.553	0.610	0.303	0.530	0.605	0.329
	m1 SL	0.542	0.599	0.309	0.508	0.585	0.273	0.548	0.609	0.302	0.526	0.604	0.327
	m2	1.862	0.575	0.287	0.477	0.561	0.259	0.473	0.581	0.280	0.443	0.586	0.305
μ4	u1 HB	0.734	0.745	0.590	0.769	0.758	0.592	0.729	0.766	0.595	0.983	1.075	0.670
	u1 SL	0.703	0.750	0.581	0.758	0.764	0.588	0.731	0.775	0.599	1.018	1.113	0.691
	u2 HB	0.655	0.651	0.562	0.711	0.650	0.532	0.656	0.657	0.533	0.862	0.846	0.600
	u2 SL	0.668	0.673	0.561	0.735	0.668	0.534	0.663	0.673	0.535	0.874	0.858	0.602
	u3 HB	0.968	1.023	0.684	1.034	1.053	0.652	0.978	1.060	0.637	0.919	0.919	0.636
	u3 SL	0.989	1.042	0.694	1.074	1.080	0.662	1.012	1.084	0.651	0.970	0.951	0.645
	u4	0.742	0.625	0.523	0.837	0.654	0.525	0.824	0.672	0.585	1.489	0.980	0.709
	u5	1.057	0.826	0.602	1.129	0.829	0.584	1.056	0.813	0.568	1.026	0.813	0.583
	m1 HB	1.451	1.583	0.879	1.547	1.630	0.847	1.612	1.723	0.879	1.614	1.732	0.884
	m1 SL	1.450	1.586	0.881	1.552	1.630	0.845	1.614	1.729	0.881	1.614	1.736	0.882
	m2	0.396	1.243	0.709	1.384	1.297	0.683	1.426	1.362	0.726	1.415	1.392	0.751
μ5	u1 HB	0.401	0.282	0.207	0.384	0.303	0.213	0.480	0.431	0.271	0.718	0.681	0.403
	u1 SL	0.394	0.286	0.193	0.388	0.319	0.211	0.463	0.444	0.268	0.730	0.706	0.407
	u2 HB	0.495	0.299	0.226	0.470	0.303	0.221	0.510	0.325	0.249	0.582	0.460	0.342
	u2 SL	0.494	0.315	0.240	0.493	0.320	0.221	0.496	0.347	0.252	0.589	0.476	0.347
	u3 HB	0.546	0.587	0.248	0.524	0.530	0.238	0.530	0.511	0.238	0.520	0.349	0.238
	u3 SL	0.532	0.595	0.251	0.547	0.533	0.240	0.521	0.510	0.244	0.510	0.352	0.246

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4
	u4	0.340	0.266	0.197	0.355	0.295	0.211	0.425	0.363	0.248	0.851	0.562	0.381
	u5	0.437	0.316	0.261	0.433	0.324	0.277	0.419	0.318	0.268	0.398	0.322	0.265
	m1 HB	0.364	0.401	0.193	0.346	0.404	0.193	0.375	0.437	0.221	0.393	0.437	0.234
	m1 SL	0.362	0.397	0.194	0.346	0.401	0.195	0.368	0.432	0.221	0.388	0.429	0.235
	m2	0.942	0.456	0.204	0.237	0.467	0.202	0.236	0.501	0.232	0.214	0.509	0.260
μ6	u1 HB	0.412	0.353	0.264	0.437	0.415	0.316	0.519	0.489	0.348	0.675	0.703	0.451
	u1 SL	0.413	0.366	0.267	0.444	0.437	0.319	0.526	0.510	0.358	0.711	0.747	0.467
	u2 HB	0.499	0.369	0.270	0.486	0.379	0.310	0.477	0.393	0.328	0.625	0.582	0.403
	u2 SL	0.514	0.382	0.273	0.522	0.393	0.314	0.486	0.405	0.337	0.633	0.601	0.409
	u3 HB	0.529	0.620	0.296	0.512	0.588	0.315	0.458	0.570	0.293	0.516	0.417	0.310
	u3 SL	0.550	0.627	0.296	0.561	0.598	0.320	0.490	0.578	0.299	0.539	0.434	0.315
	u4	0.370	0.319	0.244	0.425	0.380	0.286	0.520	0.425	0.324	0.895	0.673	0.451
	u5	0.426	0.362	0.257	0.454	0.379	0.285	0.410	0.349	0.263	0.443	0.382	0.282
	m1 HB	0.681	0.771	0.362	0.692	0.772	0.385	0.718	0.802	0.395	0.740	0.815	0.419
	m1 SL	0.678	0.771	0.362	0.690	0.773	0.383	0.715	0.803	0.395	0.735	0.812	0.417
	m2	0.672	0.713	0.345	0.656	0.722	0.354	0.654	0.748	0.363	0.654	0.755	0.392
μ7	u1 HB	0.348	0.294	0.218	0.373	0.339	0.236	0.446	0.419	0.283	0.671	0.643	0.413
	u1 SL	0.352	0.315	0.211	0.381	0.359	0.243	0.468	0.443	0.294	0.717	0.684	0.421
	u2 HB	0.479	0.323	0.239	0.464	0.330	0.255	0.450	0.333	0.269	0.595	0.488	0.350
	u2 SL	0.505	0.338	0.235	0.474	0.340	0.258	0.462	0.344	0.271	0.604	0.507	0.357
	u3 HB	0.514	0.546	0.244	0.485	0.539	0.249	0.450	0.484	0.231	0.496	0.323	0.224
	u3 SL	0.533	0.555	0.233	0.494	0.539	0.245	0.458	0.490	0.233	0.506	0.343	0.230
	u4	0.312	0.282	0.206	0.358	0.320	0.222	0.445	0.369	0.278	0.850	0.586	0.380
	u5	0.345	0.302	0.210	0.356	0.296	0.221	0.346	0.276	0.215	0.341	0.288	0.212
	m1 HB	0.484	0.546	0.254	0.479	0.524	0.225	0.499	0.543	0.266	0.511	0.576	0.290
	m1 SL	0.481	0.544	0.254	0.482	0.523	0.225	0.497	0.541	0.266	0.505	0.571	0.288
	m2	0.814	0.565	0.268	0.431	0.562	0.236	0.429	0.575	0.271	0.404	0.590	0.293

* Note: μ1 = (0,0,0), μ2 = (0.1,0.1,0.1), μ3 = (0,0.1, −0.1), μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. u1 is unidimensional method 1 (separate calibration at the total test level using one linking onstant); u2 is unidimensional method 2 (separate calibration at the total test level using three linking constats); u3 is unidimensional method 3 (separate calibration at the subtest level); u4 is unidimensional method 4 (concurrent calibration at the total test level); u5 is unidimensional method 5 (concurrent calibration at the subtest level); m1 is multidimensional method 1(separate calibration); and m2 is multidimensional method 2 (concurrent calibration). SL refers to the Stocking & Lord method, and HB stands for the Haebara method.

4.4.2.3 RMSE

Table 4.16, Table 4.17, and Table 4.18 present RMSE values from each subtest. RMSE and absolute differences provide very similar results.

Table 4.16. RMSE compared with True Equating Function (Subtest 1)

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4
μ1	u1 HB	0.647	0.603	0.491	0.669	0.650	0.524	0.690	0.651	0.506	0.812	0.797	0.606
	u1 SL	0.668	0.608	0.496	0.682	0.661	0.539	0.710	0.665	0.526	0.832	0.823	0.632
	u2 HB	0.714	0.629	0.521	0.684	0.613	0.530	0.703	0.582	0.502	0.717	0.576	0.486
	u2 SL	0.717	0.632	0.523	0.687	0.614	0.537	0.705	0.595	0.507	0.718	0.583	0.495
	u3 HB	0.731	0.771	0.539	0.706	0.751	0.547	0.710	0.786	0.515	0.727	0.621	0.517
	u3 SL	0.734	0.776	0.538	0.702	0.751	0.553	0.709	0.786	0.525	0.733	0.617	0.522
	u4	0.614	0.566	0.472	0.658	0.619	0.507	0.654	0.594	0.478	0.775	0.702	0.560
	u5	0.632	0.586	0.488	0.634	0.613	0.507	0.612	0.584	0.474	0.609	0.575	0.477
	m1 HB	0.752	0.781	0.615	0.702	0.732	0.588	0.724	0.765	0.570	0.748	0.783	0.575
	m1 SL	0.748	0.780	0.615	0.702	0.732	0.590	0.722	0.762	0.571	0.749	0.784	0.577
	m2	0.741	0.772	0.603	0.722	0.728	0.581	0.743	0.757	0.571	0.743	0.770	0.563
μ2	u1 HB	0.672	0.625	0.535	0.682	0.639	0.522	0.721	0.661	0.535	0.814	0.804	0.613
	u1 SL	0.694	0.647	0.544	0.688	0.663	0.536	0.734	0.680	0.554	0.832	0.826	0.641
	u2 HB	0.723	0.642	0.530	0.720	0.600	0.520	0.684	0.611	0.528	0.751	0.575	0.487
	u2 SL	0.733	0.649	0.532	0.723	0.607	0.524	0.690	0.617	0.532	0.755	0.592	0.492
	u3 HB	0.759	0.782	0.541	0.738	0.770	0.529	0.702	0.797	0.540	0.764	0.632	0.519
	u3 SL	0.767	0.783	0.543	0.748	0.777	0.539	0.712	0.800	0.542	0.750	0.643	0.520
	u4	0.577	0.573	0.473	0.607	0.586	0.465	0.636	0.591	0.470	0.738	0.694	0.513
	u5	0.632	0.591	0.476	0.602	0.565	0.463	0.591	0.572	0.444	0.645	0.577	0.453
	m1 HB	0.756	0.795	0.634	0.701	0.727	0.587	0.695	0.739	0.560	0.726	0.763	0.562
	m1 SL	0.755	0.795	0.634	0.702	0.726	0.586	0.698	0.740	0.559	0.730	0.766	0.564
	m2	0.761	0.789	0.625	0.740	0.731	0.583	0.745	0.740	0.555	0.747	0.754	0.551
μ3	u1 HB	0.651	0.600	0.502	0.663	0.619	0.513	0.683	0.645	0.515	0.787	0.790	0.592
	u1 SL	0.657	0.604	0.503	0.673	0.626	0.524	0.700	0.669	0.534	0.823	0.823	0.620
	u2 HB	0.742	0.639	0.516	0.690	0.599	0.500	0.681	0.583	0.509	0.662	0.586	0.491
	u2 SL	0.746	0.641	0.517	0.694	0.605	0.509	0.684	0.589	0.512	0.661	0.601	0.495
	u3 HB	0.763	0.770	0.538	0.700	0.748	0.521	0.713	0.786	0.533	0.699	0.646	0.527
	u3 SL	0.771	0.766	0.536	0.709	0.753	0.531	0.720	0.792	0.530	0.709	0.647	0.532
	u4	0.622	0.576	0.469	0.656	0.597	0.495	0.659	0.599	0.481	0.744	0.703	0.534
	u5	0.637	0.598	0.481	0.641	0.617	0.498	0.640	0.607	0.485	0.611	0.602	0.495
	m1 HB	0.750	0.775	0.626	0.693	0.738	0.573	0.698	0.743	0.573	0.719	0.760	0.566
	m1 SL	0.748	0.774	0.626	0.694	0.739	0.573	0.700	0.742	0.574	0.720	0.761	0.567
	m2	0.742	0.766	0.621	0.724	0.727	0.562	0.739	0.737	0.565	0.739	0.740	0.552
μ4	u1 HB	0.889	0.797	0.808	0.844	0.753	0.760	0.803	0.708	0.687	0.868	0.807	0.658
	u1 SL	0.881	0.808	0.802	0.846	0.768	0.758	0.804	0.710	0.693	0.868	0.801	0.666
	u2 HB	0.911	0.815	0.798	0.847	0.748	0.768	0.815	0.684	0.704	0.831	0.669	0.634
	u2 SL	0.910	0.829	0.793	0.850	0.762	0.766	0.814	0.691	0.705	0.833	0.681	0.633
	u3 HB	1.096	1.031	0.874	1.073	1.037	0.895	1.072	1.043	0.885	1.098	0.988	0.873
	u3 SL	1.092	1.048	0.869	1.078	1.056	0.892	1.077	1.061	0.889	1.093	1.005	0.869
	u4	0.860	0.688	0.729	0.821	0.655	0.699	0.810	0.659	0.643	0.922	0.796	0.640
	u5	1.092	0.909	0.837	1.080	0.899	0.860	1.068	0.900	0.840	1.082	0.900	0.840
	m1 HB	1.191	1.198	1.005	1.178	1.186	1.000	1.261	1.274	1.020	1.265	1.283	1.007
	m1 SL	1.190	1.194	1.004	1.176	1.186	1.001	1.258	1.269	1.019	1.266	1.281	1.010

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4
	m2	1.083	1.064	0.909	1.098	1.044	0.907	1.142	1.130	0.926	1.160	1.132	0.923
μ5	u1 HB	0.631	0.633	0.477	0.685	0.679	0.506	0.733	0.732	0.542	0.888	0.915	0.701
	u1 SL	0.649	0.628	0.489	0.696	0.689	0.517	0.754	0.749	0.564	0.927	0.946	0.731
	u2 HB	0.722	0.626	0.494	0.689	0.603	0.508	0.663	0.567	0.475	0.712	0.580	0.468
	u2 SL	0.722	0.615	0.493	0.680	0.600	0.504	0.658	0.556	0.474	0.710	0.581	0.469
	u3 HB	0.741	0.778	0.519	0.701	0.767	0.537	0.684	0.822	0.508	0.735	0.642	0.503
	u3 SL	0.737	0.770	0.515	0.693	0.761	0.528	0.678	0.812	0.502	0.727	0.635	0.502
	u4	0.663	0.636	0.486	0.732	0.693	0.525	0.755	0.713	0.539	0.917	0.847	0.651
	u5	0.742	0.712	0.536	0.761	0.758	0.557	0.733	0.722	0.528	0.734	0.727	0.544
	m1 HB	0.681	0.712	0.528	0.679	0.709	0.517	0.685	0.716	0.498	0.770	0.788	0.545
	m1 SL	0.686	0.717	0.530	0.687	0.715	0.520	0.692	0.721	0.503	0.784	0.802	0.555
	m2	0.640	0.716	0.532	0.624	0.714	0.525	0.631	0.720	0.508	0.621	0.776	0.527
μ6	u1 HB	0.678	0.624	0.535	0.655	0.623	0.543	0.692	0.628	0.531	0.795	0.760	0.609
	u1 SL	0.679	0.636	0.542	0.666	0.641	0.553	0.702	0.648	0.552	0.820	0.783	0.622
	u2 HB	0.748	0.630	0.566	0.701	0.618	0.556	0.713	0.587	0.528	0.692	0.592	0.531
	u2 SL	0.759	0.637	0.571	0.696	0.620	0.558	0.716	0.593	0.532	0.697	0.588	0.529
	u3 HB	0.778	0.792	0.584	0.740	0.759	0.592	0.754	0.800	0.578	0.746	0.623	0.569
	u3 SL	0.787	0.799	0.589	0.732	0.761	0.596	0.762	0.808	0.584	0.755	0.634	0.574
	u4	0.638	0.566	0.506	0.642	0.582	0.511	0.643	0.564	0.493	0.754	0.663	0.560
	u5	0.675	0.586	0.535	0.675	0.602	0.545	0.652	0.575	0.526	0.648	0.550	0.540
	m1 HB	0.814	0.845	0.688	0.760	0.787	0.653	0.790	0.826	0.651	0.797	0.823	0.644
	m1 SL	0.811	0.844	0.688	0.756	0.785	0.654	0.786	0.822	0.649	0.793	0.819	0.643
	m2	0.793	0.818	0.664	0.793	0.766	0.627	0.795	0.796	0.630	0.807	0.795	0.611
μ7	u1 HB	0.635	0.612	0.476	0.644	0.639	0.511	0.694	0.677	0.514	0.809	0.838	0.623
	u1 SL	0.653	0.617	0.481	0.664	0.653	0.518	0.712	0.688	0.537	0.838	0.870	0.649
	u2 HB	0.697	0.615	0.495	0.644	0.583	0.525	0.682	0.584	0.508	0.708	0.584	0.486
	u2 SL	0.704	0.615	0.498	0.649	0.588	0.532	0.676	0.582	0.512	0.711	0.581	0.489
	u3 HB	0.718	0.753	0.499	0.658	0.733	0.526	0.695	0.772	0.501	0.723	0.607	0.476
	u3 SL	0.715	0.751	0.497	0.655	0.733	0.534	0.697	0.767	0.504	0.731	0.599	0.473
	u4	0.623	0.600	0.458	0.658	0.620	0.497	0.684	0.630	0.490	0.797	0.751	0.575
	u5	0.642	0.630	0.458	0.632	0.641	0.495	0.636	0.621	0.464	0.622	0.621	0.464
	m1 HB	0.713	0.741	0.573	0.678	0.715	0.542	0.693	0.732	0.537	0.754	0.777	0.546
	m1 SL	0.714	0.741	0.573	0.681	0.717	0.543	0.694	0.733	0.538	0.762	0.780	0.550
	m2	0.699	0.743	0.578	0.693	0.720	0.543	0.697	0.736	0.544	0.695	0.770	0.532

* Note: μ1 = (0,0,0), μ2 = (0.1,0.1,0.1), μ3 = (0,0.1, −0.1), μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. u1 is unidimensional method 1 (separate calibration at the total test level using one linking onstant); u2 is unidimensional method 2 (separate calibration at the total test level using three linking constats); u3 is unidimensional method 3 (separate calibration at the subtest level); u4 is unidimensional method 4 (concurrent calibration at the total test level); u5 is unidimensional method 5 (concurrent calibration at the subtest level); m1 is multidimensional method 1(separate calibration); and m2 is multidimensional method 2 (concurrent calibration). SL refers to the Stocking & Lord method, and HB stands for the Haebara method.

Table 4.17. RMSE compared with True Equating Function (Subtest 2)

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4
μ1	u1 HB	0.664	0.643	0.501	0.665	0.634	0.498	0.712	0.656	0.537	0.735	0.701	0.590
	u1 SL	0.665	0.645	0.495	0.665	0.630	0.487	0.705	0.650	0.519	0.709	0.683	0.572
	u2 HB	0.742	0.657	0.519	0.728	0.652	0.510	0.736	0.656	0.547	0.718	0.672	0.568
	u2 SL	0.738	0.649	0.514	0.720	0.649	0.504	0.740	0.645	0.545	0.730	0.668	0.563
	u3 HB	0.732	0.788	0.530	0.750	0.740	0.531	0.760	0.735	0.535	0.730	0.659	0.527
	u3 SL	0.734	0.782	0.513	0.730	0.739	0.521	0.759	0.737	0.524	0.722	0.653	0.520
	u4	0.632	0.609	0.478	0.637	0.601	0.472	0.672	0.612	0.474	0.680	0.622	0.494
	u5	0.648	0.616	0.475	0.654	0.610	0.492	0.672	0.614	0.466	0.648	0.621	0.473
	m1 HB	0.643	0.634	0.460	0.653	0.645	0.468	0.666	0.660	0.451	0.719	0.719	0.472
	m1 SL	0.646	0.634	0.460	0.655	0.646	0.469	0.670	0.660	0.452	0.722	0.719	0.472
	m2	0.495	0.627	0.452	0.492	0.632	0.469	0.526	0.658	0.449	0.511	0.712	0.470
μ2	u1 HB	0.662	0.637	0.494	0.680	0.628	0.511	0.722	0.670	0.549	0.778	0.728	0.619
	u1 SL	0.673	0.638	0.492	0.671	0.624	0.502	0.708	0.654	0.534	0.744	0.711	0.599
	u2 HB	0.720	0.657	0.504	0.751	0.626	0.521	0.738	0.648	0.553	0.745	0.653	0.587
	u2 SL	0.734	0.652	0.502	0.748	0.637	0.519	0.738	0.637	0.544	0.753	0.652	0.585
	u3 HB	0.736	0.783	0.509	0.775	0.738	0.524	0.739	0.731	0.528	0.749	0.655	0.533
	u3 SL	0.745	0.780	0.509	0.764	0.753	0.519	0.748	0.738	0.523	0.762	0.656	0.535
	u4	0.663	0.609	0.514	0.688	0.612	0.513	0.733	0.631	0.534	0.792	0.665	0.560
	u5	0.727	0.646	0.547	0.735	0.640	0.535	0.729	0.641	0.540	0.747	0.633	0.542
	m1 HB	0.632	0.630	0.459	0.605	0.614	0.466	0.648	0.644	0.457	0.687	0.681	0.459
	m1 SL	0.633	0.630	0.460	0.608	0.615	0.467	0.651	0.645	0.458	0.689	0.681	0.461
	m2	0.484	0.632	0.457	0.469	0.620	0.464	0.510	0.640	0.457	0.502	0.680	0.457
μ3	u1 HB	0.908	0.892	0.681	0.921	0.917	0.665	0.970	0.951	0.708	1.035	0.995	0.761
	u1 SL	0.904	0.877	0.678	0.917	0.896	0.658	0.957	0.920	0.695	0.997	0.951	0.736
	u2 HB	0.727	0.630	0.530	0.729	0.643	0.534	0.748	0.678	0.557	0.769	0.659	0.594
	u2 SL	0.728	0.636	0.527	0.733	0.643	0.525	0.744	0.674	0.553	0.770	0.652	0.590
	u3 HB	0.723	0.770	0.530	0.748	0.742	0.538	0.740	0.735	0.529	0.732	0.659	0.542
	u3 SL	0.734	0.774	0.525	0.756	0.750	0.530	0.738	0.740	0.523	0.739	0.656	0.529
	u4	0.865	0.780	0.619	0.850	0.785	0.595	0.903	0.815	0.618	0.971	0.856	0.647
	u5	0.747	0.633	0.550	0.715	0.638	0.529	0.733	0.625	0.536	0.745	0.634	0.549
	m1 HB	0.640	0.616	0.444	0.645	0.618	0.471	0.651	0.643	0.446	0.687	0.691	0.458
	m1 SL	0.642	0.618	0.444	0.645	0.617	0.472	0.655	0.644	0.448	0.689	0.692	0.459
	m2	0.494	0.617	0.438	0.482	0.621	0.469	0.513	0.635	0.443	0.486	0.684	0.455
μ4	u1 HB	0.709	0.703	0.569	0.685	0.668	0.546	0.681	0.685	0.528	0.740	0.694	0.613
	u1 SL	0.713	0.717	0.570	0.694	0.678	0.549	0.679	0.685	0.525	0.727	0.678	0.603
	u2 HB	0.760	0.713	0.590	0.755	0.686	0.566	0.701	0.670	0.538	0.724	0.646	0.564
	u2 SL	0.756	0.715	0.583	0.753	0.689	0.567	0.709	0.660	0.539	0.722	0.644	0.562
	u3 HB	0.879	0.912	0.634	0.877	0.881	0.623	0.854	0.890	0.636	0.889	0.860	0.648
	u3 SL	0.867	0.917	0.623	0.875	0.888	0.622	0.860	0.891	0.634	0.877	0.859	0.635
	u4	0.637	0.572	0.494	0.641	0.562	0.477	0.638	0.604	0.479	0.738	0.642	0.521
	u5	0.784	0.707	0.538	0.784	0.688	0.530	0.759	0.715	0.531	0.783	0.705	0.551
	m1 HB	0.831	0.874	0.629	0.838	0.893	0.618	0.929	0.982	0.652	1.006	1.045	0.688
	m1 SL	0.833	0.875	0.629	0.840	0.893	0.621	0.930	0.982	0.654	1.008	1.043	0.690
	m2	0.662	0.745	0.517	0.694	0.756	0.511	0.712	0.858	0.549	0.752	0.909	0.596
μ5	u1 HB	0.752	0.693	0.599	0.775	0.721	0.594	0.779	0.724	0.634	0.793	0.737	0.651
	u1 SL	0.741	0.685	0.582	0.752	0.702	0.570	0.756	0.704	0.611	0.750	0.710	0.622
	u2 HB	0.808	0.720	0.606	0.818	0.743	0.621	0.815	0.728	0.653	0.779	0.705	0.626
	u2 SL	0.803	0.702	0.596	0.807	0.730	0.604	0.808	0.720	0.639	0.789	0.710	0.622
	u3 HB	0.850	0.842	0.638	0.851	0.832	0.652	0.834	0.822	0.657	0.818	0.751	0.627
	u3 SL	0.834	0.820	0.617	0.824	0.818	0.629	0.815	0.808	0.633	0.806	0.729	0.604

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4
	u4	0.751	0.685	0.560	0.756	0.696	0.550	0.762	0.684	0.568	0.755	0.683	0.545
	u5	0.826	0.756	0.604	0.823	0.761	0.609	0.817	0.747	0.606	0.813	0.754	0.596
	m1 HB	0.755	0.710	0.580	0.772	0.731	0.583	0.766	0.721	0.555	0.805	0.782	0.548
	m1 SL	0.756	0.708	0.581	0.772	0.730	0.584	0.768	0.721	0.558	0.806	0.783	0.551
	m2	0.678	0.667	0.524	0.699	0.698	0.530	0.736	0.693	0.510	0.737	0.757	0.518
μ6	u1 HB	0.661	0.616	0.479	0.645	0.612	0.488	0.673	0.639	0.519	0.726	0.693	0.586
	u1 SL	0.662	0.623	0.476	0.639	0.605	0.479	0.665	0.632	0.505	0.705	0.677	0.572
	u2 HB	0.713	0.636	0.497	0.699	0.622	0.499	0.720	0.643	0.530	0.718	0.645	0.539
	u2 SL	0.713	0.637	0.495	0.685	0.615	0.490	0.721	0.642	0.528	0.718	0.646	0.532
	u3 HB	0.721	0.762	0.498	0.720	0.732	0.498	0.730	0.720	0.504	0.725	0.635	0.499
	u3 SL	0.718	0.767	0.496	0.700	0.735	0.488	0.740	0.726	0.513	0.719	0.648	0.492
	u4	0.611	0.576	0.453	0.598	0.557	0.450	0.634	0.588	0.468	0.678	0.601	0.495
	u5	0.612	0.581	0.450	0.622	0.572	0.448	0.606	0.584	0.436	0.608	0.569	0.442
	m1 HB	0.626	0.640	0.430	0.626	0.646	0.452	0.654	0.674	0.442	0.712	0.729	0.472
	m1 SL	0.630	0.641	0.433	0.629	0.648	0.453	0.659	0.675	0.446	0.715	0.731	0.473
	m2	0.452	0.622	0.428	0.444	0.627	0.448	0.470	0.658	0.439	0.454	0.713	0.464
μ7	u1 HB	0.669	0.643	0.539	0.704	0.649	0.530	0.706	0.693	0.566	0.768	0.703	0.607
	u1 SL	0.665	0.635	0.529	0.690	0.635	0.516	0.695	0.677	0.547	0.733	0.681	0.585
	u2 HB	0.741	0.665	0.549	0.769	0.668	0.552	0.755	0.676	0.567	0.753	0.667	0.579
	u2 SL	0.749	0.651	0.539	0.755	0.664	0.544	0.758	0.673	0.559	0.757	0.673	0.574
	u3 HB	0.757	0.786	0.560	0.794	0.763	0.559	0.773	0.777	0.553	0.760	0.688	0.556
	u3 SL	0.756	0.783	0.546	0.770	0.761	0.547	0.772	0.772	0.541	0.748	0.683	0.550
	u4	0.652	0.616	0.505	0.687	0.620	0.496	0.677	0.646	0.503	0.711	0.635	0.519
	u5	0.694	0.648	0.514	0.712	0.658	0.521	0.699	0.666	0.511	0.693	0.654	0.519
	m1 HB	0.657	0.646	0.496	0.700	0.667	0.508	0.692	0.668	0.485	0.744	0.728	0.492
	m1 SL	0.658	0.646	0.498	0.700	0.667	0.508	0.694	0.668	0.486	0.745	0.728	0.492
	m2	0.552	0.631	0.472	0.576	0.649	0.491	0.588	0.658	0.473	0.588	0.720	0.490

* Note: μ1 = (0,0,0), μ2 = (0.1,0.1,0.1), μ3 = (0,0.1,−0.1), μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. u1 is unidimensional method 1 (separate calibration at the total test level using one linking onstant); u2 is unidimensional method 2 (separate calibration at the total test level using three linking constats); u3 is unidimensional method 3 (separate calibration at the subtest level); u4 is unidimensional method 4 (concurrent calibration at the total test level); u5 is unidimensional method 5 (concurrent calibration at the subtest level); m1 is multidimensional method 1(separate calibration); and m2 is multidimensional method 2 (concurrent calibration). SL refers to the Stocking & Lord method, and HB stands for the Haebara method.

Table 4.18. RMSE compared with True Equating Function (Subtest 3)

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4
μ1	u1 HB	0.595	0.571	0.482	0.627	0.612	0.495	0.661	0.658	0.546	0.827	0.809	0.654
	u1 SL	0.607	0.583	0.481	0.630	0.629	0.504	0.682	0.678	0.553	0.853	0.836	0.668
	u2 HB	0.690	0.596	0.500	0.662	0.577	0.518	0.668	0.608	0.540	0.769	0.711	0.618
	u2 SL	0.704	0.603	0.507	0.670	0.596	0.517	0.681	0.619	0.548	0.772	0.719	0.624
	u3 HB	0.704	0.756	0.500	0.687	0.728	0.520	0.684	0.726	0.511	0.708	0.581	0.517
	u3 SL	0.726	0.759	0.506	0.695	0.728	0.519	0.689	0.725	0.521	0.715	0.594	0.524
	u4	0.568	0.548	0.470	0.603	0.592	0.489	0.672	0.619	0.540	0.908	0.786	0.640
	u5	0.614	0.571	0.475	0.602	0.560	0.485	0.599	0.544	0.470	0.601	0.566	0.479
	m1 HB	0.736	0.800	0.531	0.733	0.789	0.508	0.748	0.806	0.543	0.757	0.814	0.552
	m1 SL	0.735	0.799	0.531	0.735	0.788	0.508	0.748	0.806	0.544	0.754	0.812	0.551
	m2	0.834	0.790	0.522	0.706	0.783	0.510	0.701	0.798	0.537	0.691	0.805	0.546
μ2	u1 HB	0.621	0.548	0.477	0.619	0.590	0.510	0.652	0.665	0.558	0.826	0.795	0.630
	u1 SL	0.623	0.563	0.477	0.625	0.613	0.510	0.664	0.682	0.566	0.847	0.824	0.653
	u2 HB	0.686	0.583	0.509	0.676	0.575	0.530	0.693	0.608	0.545	0.740	0.688	0.609
	u2 SL	0.708	0.596	0.510	0.697	0.595	0.531	0.695	0.617	0.545	0.740	0.694	0.610
	u3 HB	0.712	0.773	0.517	0.699	0.742	0.522	0.706	0.735	0.506	0.695	0.582	0.511
	u3 SL	0.733	0.771	0.519	0.714	0.746	0.520	0.718	0.736	0.506	0.701	0.592	0.513
	u4	0.566	0.498	0.449	0.560	0.534	0.470	0.637	0.594	0.524	0.813	0.723	0.593
	u5	0.615	0.516	0.465	0.602	0.494	0.463	0.596	0.493	0.465	0.611	0.488	0.452
	m1 HB	0.770	0.817	0.546	0.739	0.795	0.511	0.762	0.815	0.546	0.770	0.814	0.557
	m1 SL	0.770	0.816	0.547	0.741	0.794	0.510	0.760	0.814	0.545	0.768	0.812	0.555
	m2	0.798	0.811	0.548	0.725	0.792	0.512	0.731	0.814	0.550	0.721	0.812	0.561
μ3	u1 HB	1.093	1.052	0.704	1.066	1.044	0.720	1.055	1.056	0.747	1.112	1.107	0.807
	u1 SL	1.088	1.069	0.705	1.069	1.071	0.728	1.068	1.096	0.763	1.157	1.164	0.833
	u2 HB	0.677	0.585	0.493	0.699	0.572	0.532	0.686	0.620	0.551	0.787	0.771	0.644
	u2 SL	0.686	0.590	0.506	0.700	0.578	0.540	0.698	0.631	0.560	0.799	0.777	0.652
	u3 HB	0.691	0.760	0.502	0.720	0.747	0.515	0.668	0.720	0.500	0.695	0.596	0.489
	u3 SL	0.698	0.761	0.513	0.724	0.753	0.520	0.685	0.721	0.510	0.704	0.606	0.499
	u4	1.020	0.936	0.628	1.020	0.939	0.657	1.036	0.953	0.686	1.176	1.015	0.764
	u5	0.809	0.686	0.531	0.823	0.678	0.569	0.824	0.687	0.543	0.775	0.691	0.537
	m1 HB	0.726	0.764	0.533	0.707	0.760	0.508	0.733	0.771	0.525	0.719	0.769	0.549
	m1 SL	0.726	0.765	0.534	0.708	0.760	0.509	0.729	0.771	0.524	0.716	0.768	0.548
	m2	1.321	0.751	0.514	0.674	0.745	0.495	0.670	0.753	0.505	0.649	0.757	0.529
μ4	u1 HB	0.830	0.826	0.739	0.848	0.835	0.742	0.832	0.844	0.748	0.967	1.002	0.797
	u1 SL	0.809	0.827	0.733	0.840	0.836	0.740	0.832	0.849	0.750	0.983	1.019	0.809
	u2 HB	0.794	0.776	0.725	0.826	0.777	0.704	0.790	0.778	0.704	0.898	0.875	0.746
	u2 SL	0.801	0.789	0.723	0.840	0.788	0.706	0.794	0.787	0.706	0.906	0.881	0.746
	u3 HB	0.952	0.976	0.798	0.985	0.987	0.780	0.954	0.990	0.770	0.927	0.912	0.770
	u3 SL	0.961	0.985	0.803	1.004	1.000	0.785	0.970	0.999	0.780	0.954	0.928	0.776
	u4	0.841	0.769	0.693	0.890	0.786	0.696	0.877	0.789	0.735	1.163	0.945	0.812
	u5	0.996	0.877	0.746	1.029	0.881	0.735	0.993	0.869	0.725	0.982	0.872	0.734
	m1 HB	1.154	1.209	0.892	1.195	1.231	0.878	1.214	1.259	0.891	1.217	1.265	0.894
	m1 SL	1.154	1.210	0.893	1.197	1.231	0.878	1.215	1.262	0.893	1.218	1.266	0.893
	m2	0.600	1.078	0.801	1.127	1.104	0.790	1.143	1.125	0.810	1.138	1.139	0.824
μ5	u1 HB	0.630	0.530	0.453	0.618	0.548	0.458	0.689	0.651	0.515	0.837	0.813	0.622
	u1 SL	0.626	0.534	0.437	0.622	0.563	0.454	0.678	0.662	0.511	0.845	0.830	0.625
	u2 HB	0.701	0.544	0.472	0.683	0.548	0.467	0.710	0.567	0.493	0.752	0.670	0.572
	u2 SL	0.699	0.557	0.485	0.699	0.563	0.466	0.700	0.584	0.496	0.755	0.680	0.576
	u3 HB	0.737	0.762	0.496	0.722	0.726	0.486	0.726	0.712	0.486	0.718	0.589	0.486
	u3 SL	0.728	0.767	0.498	0.737	0.729	0.487	0.720	0.711	0.492	0.712	0.591	0.493

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4	32 & 4	32 & 8	16 & 4
	u4	0.580	0.511	0.440	0.592	0.538	0.453	0.645	0.595	0.491	0.895	0.736	0.601
	u5	0.653	0.551	0.505	0.648	0.558	0.519	0.638	0.553	0.512	0.621	0.557	0.508
	m1 HB	0.599	0.630	0.425	0.584	0.633	0.426	0.604	0.654	0.450	0.617	0.652	0.467
	m1 SL	0.597	0.627	0.425	0.585	0.630	0.429	0.599	0.650	0.450	0.613	0.646	0.468
	m2	0.920	0.672	0.435	0.479	0.679	0.434	0.478	0.698	0.458	0.455	0.701	0.485
μ6	u1 HB	0.638	0.588	0.506	0.656	0.636	0.551	0.712	0.687	0.579	0.810	0.819	0.659
	u1 SL	0.637	0.598	0.507	0.662	0.651	0.553	0.717	0.701	0.586	0.829	0.844	0.670
	u2 HB	0.699	0.600	0.514	0.691	0.607	0.548	0.685	0.617	0.560	0.774	0.743	0.619
	u2 SL	0.709	0.609	0.517	0.716	0.618	0.551	0.691	0.626	0.568	0.781	0.755	0.623
	u3 HB	0.720	0.777	0.539	0.711	0.757	0.554	0.672	0.744	0.534	0.708	0.637	0.550
	u3 SL	0.733	0.782	0.538	0.744	0.763	0.559	0.695	0.749	0.540	0.724	0.649	0.554
	u4	0.606	0.562	0.483	0.649	0.611	0.522	0.712	0.643	0.555	0.917	0.794	0.653
	u5	0.649	0.597	0.497	0.670	0.610	0.523	0.636	0.587	0.502	0.661	0.614	0.521
	m1 HB	0.809	0.863	0.577	0.819	0.867	0.598	0.830	0.879	0.600	0.843	0.887	0.619
	m1 SL	0.808	0.863	0.577	0.818	0.867	0.597	0.828	0.879	0.601	0.840	0.885	0.618
	m2	0.777	0.832	0.563	0.785	0.840	0.574	0.784	0.850	0.575	0.784	0.855	0.599
μ7	u1 HB	0.588	0.541	0.463	0.608	0.580	0.479	0.663	0.640	0.525	0.810	0.789	0.629
	u1 SL	0.592	0.559	0.454	0.616	0.596	0.485	0.679	0.659	0.534	0.838	0.814	0.636
	u2 HB	0.689	0.565	0.485	0.678	0.570	0.500	0.666	0.572	0.512	0.761	0.684	0.576
	u2 SL	0.707	0.577	0.481	0.685	0.578	0.500	0.674	0.581	0.514	0.766	0.696	0.582
	u3 HB	0.715	0.734	0.490	0.694	0.730	0.493	0.668	0.691	0.476	0.701	0.566	0.468
	u3 SL	0.726	0.739	0.479	0.700	0.730	0.488	0.674	0.695	0.478	0.708	0.582	0.474
	u4	0.557	0.528	0.448	0.596	0.563	0.462	0.660	0.600	0.518	0.897	0.748	0.600
	u5	0.585	0.544	0.452	0.594	0.539	0.460	0.584	0.520	0.456	0.582	0.532	0.452
	m1 HB	0.689	0.733	0.484	0.687	0.720	0.460	0.698	0.730	0.493	0.705	0.751	0.517
	m1 SL	0.687	0.732	0.485	0.689	0.719	0.461	0.697	0.729	0.493	0.702	0.747	0.515
	m2	0.852	0.745	0.497	0.642	0.745	0.469	0.640	0.751	0.496	0.621	0.759	0.517

* Note: μ1 = (0,0,0), μ2 = (0.1,0.1,0.1), μ3 = (0,0.1, −0.1), μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. u1 is unidimensional method 1 (separate calibration at the total test level using one linking onstant); u2 is unidimensional method 2 (separate calibration at the total test level using three linking constats); u3 is unidimensional method 3 (separate calibration at the subtest level); u4 is unidimensional method 4 (concurrent calibration at the total test level); u5 is unidimensional method 5 (concurrent calibration at the subtest level); m1 is multidimensional method 1(separate calibration); and m2 is multidimensional method 2 (concurrent calibration). SL refers to the Stocking & Lord method, and HB stands for the Haebara method.

4.4.3 Number of score points equal between True and Estimated Conversion and score difference that matters (DTM)

4.4.3.1 Score points equal between true and estimated equating output

Percentages of the number of score points equal between the conversion tables from true and estimated equating results are displayed in Table 4.19, Table 4.20, and Table 4.21. Table 4.19 shows results from subtest 1. When correlations were high, the unidimensional method at the total test level using one linking constant (u1) produced slightly better results than the one using three linking constants (u2). Under the low correlation, u2 outperformed u1. The second unidimensional method (u2) showed higher percentages implying better results than the unidimensional method at the subtest level (u3). In general, unidimensional methods from concurrent calibration tend to outperform unidimensional methods from separate calibration except for the cases under the skewed distributions with high skewness (μ4 and μ5).

With moderate or low correlations (0.8, 0.6, and 0.4) and 32 items including 8 common items, the multidimensional method using separate calibration (m1) produced the highest percentages compared to the other methods, with the exception of the positively skewed distributions (μ4 and μ6). In general, unidimensional methods showed better results than multidimensional methods in most cases.

Table 4.19. Average Percentages of the Number of Score Points Equal between True and Estimated Conversion (Subtest 1)

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
µ1	u1 HB	57.45	55.70	70.29	52.79	53.30	69.00	52.73	49.73	70.06	45.52	41.30	60.41
	u1 SL	56.55	54.33	70.53	52.36	52.36	68.06	52.55	49.30	68.06	44.39	41.76	58.88
	u2 HB	54.30	48.67	66.88	55.88	52.15	68.00	59.30	51.42	68.59	59.64	49.61	69.94
	u2 SL	53.70	49.67	67.29	55.48	52.45	67.53	57.82	51.24	67.94	59.06	50.21	70.18
	u3 HB	47.39	46.45	65.76	50.03	49.39	66.88	46.52	50.12	67.88	55.79	48.88	67.59
	u3 SL	46.82	47.45	66.18	49.76	50.33	66.29	46.70	51.33	67.00	55.79	49.30	66.88
	u4	60.21	58.82	71.06	56.30	53.52	69.65	58.85	54.09	72.00	51.24	45.27	66.06
	u5	59.39	55.58	69.76	57.15	55.48	69.76	58.52	56.70	71.24	60.03	57.79	70.94
	m1 HB	48.67	53.33	57.18	53.82	56.39	60.53	50.97	55.70	60.76	50.48	54.76	62.59
	m1 SL	48.55	53.45	57.24	53.97	56.36	60.12	51.15	55.82	60.59	50.70	54.76	62.35
	m2	49.76	48.94	57.41	54.67	49.30	60.59	51.18	48.79	59.76	50.58	48.39	62.71
µ2	u1 HB	54.39	53.52	66.65	53.55	52.06	67.71	52.52	47.73	67.18	45.24	42.12	61.82
	u1 SL	52.45	51.82	66.12	52.03	52.79	67.29	51.42	47.21	65.59	44.85	42.42	60.18
	u2 HB	52.82	47.55	66.53	58.48	49.09	67.24	55.30	51.61	66.71	59.42	45.55	67.29
	u2 SL	52.30	48.00	66.94	57.67	49.64	66.47	55.12	51.06	66.06	58.30	45.76	67.59
	u3 HB	44.76	44.79	64.94	48.27	46.64	67.00	46.03	49.42	65.59	53.88	45.58	67.59
	u3 SL	44.55	44.85	65.47	47.97	46.88	66.35	45.67	49.30	65.29	52.85	46.67	67.24
	u4	58.64	57.76	69.59	58.61	54.27	71.00	57.58	50.76	70.71	50.79	43.58	67.65
	u5	57.85	50.09	68.59	60.88	51.48	69.94	59.42	52.42	71.41	59.52	47.94	71.06
	m1 HB	45.67	52.30	55.65	52.58	56.61	59.12	51.64	57.76	61.47	49.91	55.94	65.18
	m1 SL	45.79	52.45	55.65	52.64	56.33	59.41	51.76	57.36	61.65	49.91	55.64	64.47
	m2	47.36	48.27	56.24	52.06	48.55	59.12	52.21	48.30	61.41	50.64	48.24	65.76
µ3	u1 HB	57.79	55.94	68.53	55.91	52.39	69.06	53.85	50.58	68.18	45.45	44.55	63.00
	u1 SL	57.06	55.06	68.94	55.85	51.79	68.29	51.39	49.94	66.82	43.79	43.88	61.41
	u2 HB	54.52	47.61	67.71	57.39	50.94	68.88	59.64	52.06	67.12	57.30	53.79	69.12
	u2 SL	54.24	47.70	66.82	57.15	51.03	69.12	59.27	52.09	67.00	57.21	53.85	69.18
	u3 HB	47.88	45.58	65.41	49.33	49.52	67.94	48.61	48.97	66.35	54.12	51.03	67.94
	u3 SL	48.36	45.67	65.41	49.39	49.52	67.71	48.55	48.36	66.94	54.12	50.67	67.65
	u4	59.58	57.45	71.06	57.76	53.88	71.18	57.94	52.30	72.06	51.42	46.73	67.47
	u5	57.70	55.88	70.06	55.39	56.00	70.12	57.45	54.82	70.47	57.64	57.79	69.47
	m1 HB	48.52	53.36	56.59	51.30	56.27	61.71	51.55	57.61	60.65	50.36	55.45	62.24
	m1 SL	48.42	53.42	56.82	51.30	56.55	61.53	51.85	57.67	60.47	50.36	55.27	62.29
	m2	49.42	48.73	56.24	52.36	48.88	61.35	51.27	48.42	60.53	51.88	48.67	62.71
µ4	u1 HB	34.52	37.45	43.18	38.18	40.58	47.00	45.91	43.79	52.29	44.09	45.73	55.94
	u1 SL	33.64	37.24	43.53	35.79	40.21	47.35	45.52	43.24	51.59	43.91	44.76	55.65
	u2 HB	32.39	34.42	44.06	38.09	38.39	47.76	46.58	41.58	51.88	51.73	44.55	56.53
	u2 SL	31.42	34.61	44.53	37.06	38.70	47.65	46.15	41.70	51.53	50.33	44.18	56.59
	u3 HB	24.15	26.97	38.65	22.85	27.33	36.94	24.00	27.33	37.29	23.67	27.73	37.59
	u3 SL	22.73	26.21	39.12	21.73	26.27	36.82	22.39	26.21	37.12	21.67	26.91	38.00
	u4	43.94	40.42	47.53	47.91	43.79	49.65	51.12	46.03	53.06	46.42	47.03	55.47
	u5	30.00	29.88	38.41	30.30	30.27	36.24	30.97	31.09	38.76	30.48	30.42	38.06
	m1 HB	15.88	20.18	32.94	16.61	20.55	32.53	15.06	18.61	30.59	15.30	18.52	31.71
	m1 SL	15.91	20.18	32.65	16.12	20.30	32.06	14.97	18.64	30.82	15.00	18.45	31.53
	m2	20.00	25.33	41.41	20.03	24.42	41.12	15.91	22.27	39.59	17.09	21.58	40.18
µ5	u1 HB	51.15	47.88	74.65	45.55	44.45	72.12	43.73	40.85	68.71	35.00	32.85	52.35
	u1 SL	51.33	47.82	73.12	45.03	45.24	71.47	42.88	40.52	67.71	34.30	31.85	50.53
	u2 HB	53.06	42.88	72.06	51.30	43.79	70.12	55.33	48.15	71.53	54.88	44.85	72.06
	u2 SL	53.48	44.45	71.94	51.48	45.45	70.41	56.21	48.97	72.29	55.61	45.70	72.29
	u3 HB	41.24	39.00	70.00	41.39	41.91	67.82	37.91	42.85	69.71	48.64	40.52	70.94
	u3 SL	41.48	40.15	70.82	41.55	42.91	68.71	39.30	44.70	71.00	49.42	41.21	70.41

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
	u4	50.58	45.27	74.88	45.21	39.76	71.24	44.82	39.76	70.24	38.03	29.61	58.29
	u5	40.97	35.06	70.47	36.76	32.91	68.06	40.09	35.09	71.65	40.82	35.55	69.94
	m1 HB	53.06	56.82	66.12	52.45	56.39	68.12	52.39	56.06	68.06	45.00	47.30	66.59
	m1 SL	52.70	56.52	65.82	51.82	55.76	68.06	51.88	55.24	67.82	44.42	46.91	66.18
	m2	54.64	55.12	63.71	53.45	58.94	66.35	53.21	57.30	66.18	47.55	58.18	67.65
μ6	u1 HB	54.00	53.52	65.94	55.18	55.03	66.41	55.85	50.45	67.59	48.30	45.03	60.29
	u1 SL	52.82	53.67	65.24	54.52	54.61	65.53	53.88	50.33	65.76	48.03	43.21	60.12
	u2 HB	53.36	47.36	63.53	55.21	52.24	63.88	58.15	51.15	65.71	58.39	52.39	64.29
	u2 SL	53.12	46.61	63.94	55.30	52.06	63.35	57.94	51.24	65.94	58.52	52.06	65.29
	u3 HB	41.15	44.79	62.76	46.03	47.64	60.88	43.82	46.27	63.06	53.15	45.30	62.94
	u3 SL	40.85	43.76	62.88	45.73	47.88	61.29	42.76	46.06	63.06	52.79	44.15	63.00
	u4	59.45	56.52	67.12	59.70	57.03	67.88	61.24	57.27	69.24	54.88	49.79	64.00
	u5	57.70	54.33	64.94	56.21	53.21	64.41	58.39	55.55	65.00	60.91	55.15	64.24
	m1 HB	40.58	45.70	54.00	45.97	51.03	57.00	42.33	48.45	55.18	43.73	48.52	57.88
	m1 SL	40.52	46.09	53.76	46.33	51.18	56.94	42.70	48.36	55.29	43.79	48.36	57.47
	m2	43.52	47.27	54.41	48.64	47.09	57.71	46.42	46.70	56.06	47.30	46.33	59.12
μ7	u1 HB	56.70	53.94	72.24	53.58	52.73	70.24	50.64	47.21	70.35	41.18	40.73	61.47
	u1 SL	56.15	53.36	72.29	52.27	52.09	70.00	49.82	46.97	67.94	40.36	39.55	59.29
	u2 HB	56.88	50.61	69.53	57.79	51.85	68.06	57.88	50.67	67.53	56.70	49.76	70.47
	u2 SL	57.48	51.12	69.29	57.97	52.45	68.06	58.70	51.76	67.18	57.45	49.91	69.94
	u3 HB	47.30	48.94	69.00	49.70	49.12	67.82	47.88	49.88	69.29	57.36	46.73	71.47
	u3 SL	47.15	48.64	68.76	49.58	51.27	67.24	47.70	50.18	69.29	57.33	47.36	71.35
	u4	56.91	55.15	73.35	55.21	51.91	72.24	54.88	48.55	73.71	46.12	40.88	65.88
	u5	53.52	51.21	74.18	52.64	51.12	72.29	54.27	51.42	73.12	53.73	53.09	75.06
	m1 HB	52.73	56.06	60.76	54.58	59.03	64.12	52.91	58.76	64.59	50.15	52.24	65.65
	m1 SL	52.70	56.18	60.94	54.42	58.64	63.88	53.00	58.73	64.53	50.18	51.91	65.18
	m2	51.94	49.97	58.65	54.42	50.36	62.71	54.36	50.06	62.76	50.42	50.06	66.06

* Note: μ1 = (0,0,0), μ2 = (0.1,0.1,0.1), μ3 = (0,0.1, −0.1), μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. u1 is unidimensional method 1 (separate calibration at the total test level using one linking onstant); u2 is unidimensional method 2 (separate calibration at the total test level using three linking constats); u3 is unidimensional method 3 (separate calibration at the subtest level); u4 is unidimensional method 4 (concurrent calibration at the total test level); u5 is unidimensional method 5 (concurrent calibration at the subtest level); m1 is multidimensional method 1(separate calibration); and m2 is multidimensional method 2 (concurrent calibration). SL refers to the Stocking & Lord method, and HB stands for the Haebara method.

In subtest 2, multidimensional methods (m1 and m2) produced higher percentages than unidimensional methods except for the results based on the skewed distribution (μ4). In subtest 3, on the other hand, the opposite pattern was observed. The multidimensional methods (m1 and m2) did not perform as well as the unidimensional methods (u1, u2, u3, u4, and u5). There was no method that performed the best throughout all conditions.

The best method slightly varied depending on the conditions—proficiency distributions, correlations, and the test length with a different number of common items.

Table 4.20. Average Percentages of the Number of Score Points Equal between True and Estimated Conversion (Subtest 2)

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
μ1	u1 HB	48.39	51.55	65.76	48.15	49.88	64.94	47.00	47.45	60.24	44.06	46.33	53.94
	u1 SL	49.33	52.30	67.00	49.45	50.39	67.06	48.48	48.15	63.18	45.36	47.24	56.65
	u2 HB	47.73	48.06	65.06	48.55	47.97	66.47	47.39	47.61	63.71	47.48	47.85	61.41
	u2 SL	47.79	48.30	65.88	48.42	47.67	68.29	47.24	48.48	64.18	47.21	46.82	62.47
	u3 HB	41.58	48.42	64.35	43.36	48.18	65.35	45.30	48.85	62.41	49.21	49.33	64.59
	u3 SL	42.70	49.09	65.76	43.79	49.64	67.35	45.73	49.97	64.59	48.91	51.00	65.12
	u4	51.21	53.82	68.24	52.27	51.06	69.06	50.33	49.76	65.71	49.12	49.03	61.24
	u5	51.39	53.76	67.94	51.55	53.18	67.88	52.00	52.06	66.94	51.82	54.12	67.59
	m1 HB	70.79	50.42	84.71	67.42	49.33	84.00	68.79	49.76	83.18	62.00	44.48	78.35
	m1 SL	70.70	50.58	84.88	67.55	49.33	84.00	68.79	49.76	82.94	62.00	44.48	78.47
	m2	71.55	56.91	85.00	71.45	55.67	84.71	70.67	56.73	82.94	62.48	56.94	79.71
μ2	u1 HB	49.30	54.64	65.76	48.73	50.24	62.94	43.64	45.45	56.53	39.64	43.45	52.12
	u1 SL	49.39	54.24	67.12	49.36	51.24	64.88	45.91	45.67	59.00	42.03	44.21	54.59
	u2 HB	47.33	49.67	67.06	49.73	46.79	65.47	48.27	46.94	62.35	48.33	47.33	62.71
	u2 SL	47.45	49.39	67.76	49.03	46.76	67.12	49.15	47.15	64.47	48.48	47.64	65.71
	u3 HB	41.42	48.64	67.59	44.79	46.88	67.29	45.55	48.82	63.59	49.12	48.33	65.53
	u3 SL	41.88	49.30	69.00	44.21	46.55	67.94	43.73	47.97	65.29	48.30	48.94	68.18
	u4	50.33	50.27	59.59	50.48	47.73	59.29	47.48	44.67	56.47	44.33	43.45	52.65
	u5	47.15	45.94	54.24	48.36	45.94	55.94	48.00	45.18	54.82	49.42	44.00	52.82
	m1 HB	68.61	50.03	83.29	71.12	51.55	84.18	68.88	49.12	82.24	65.58	45.42	79.29
	m1 SL	68.58	49.67	83.59	70.88	51.55	84.18	68.82	49.09	82.47	65.55	45.61	79.24
	m2	70.06	55.30	83.88	71.33	55.12	85.06	70.39	54.64	82.18	66.58	54.94	80.00
μ3	u1 HB	31.85	31.94	39.65	28.48	31.85	41.76	27.09	30.52	40.41	25.70	32.30	38.71
	u1 SL	32.39	32.03	39.76	29.09	32.67	42.47	28.24	31.06	41.35	27.06	33.70	40.41
	u2 HB	50.00	47.85	64.18	48.15	47.70	64.41	46.03	45.52	63.29	46.12	46.06	58.35
	u2 SL	49.39	47.79	65.00	48.39	47.97	66.47	45.73	46.27	64.06	46.58	46.52	60.53
	u3 HB	43.15	48.85	64.35	43.21	47.76	65.35	45.24	47.82	65.12	48.48	50.12	62.12
	u3 SL	43.30	48.94	66.35	43.18	47.06	67.06	44.61	49.79	65.18	49.76	50.39	64.29
	u4	36.94	34.94	44.12	36.15	35.52	46.47	32.91	33.12	45.82	31.06	36.06	44.00
	u5	49.88	44.82	53.76	48.79	44.88	57.00	49.30	43.94	53.53	48.82	45.27	52.59
	m1 HB	68.36	48.55	83.41	70.27	48.70	84.65	67.36	49.85	83.06	64.18	46.00	79.41
	m1 SL	68.15	48.64	83.47	70.27	48.79	84.47	67.27	49.88	82.82	63.76	46.24	79.71
	m2	70.52	54.03	83.71	71.21	53.91	84.94	68.88	55.76	83.41	65.97	56.33	79.47
μ4	u1 HB	40.24	43.12	62.53	43.42	43.55	64.12	43.36	44.45	62.18	44.67	43.67	52.71
	u1 SL	38.67	44.15	62.65	43.12	43.45	64.12	44.58	45.42	62.71	47.15	44.79	53.71
	u2 HB	40.30	36.21	61.82	43.21	36.52	64.29	46.30	43.27	66.71	50.21	44.79	64.47
	u2 SL	38.61	36.76	61.29	42.27	35.42	63.82	47.45	42.36	65.53	47.97	44.97	63.76
	u3 HB	27.82	29.18	57.06	29.97	29.27	59.41	29.42	30.39	55.59	30.55	27.91	53.47
	u3 SL	27.70	30.97	57.41	29.03	28.79	60.18	29.15	30.73	56.12	29.85	29.48	55.06
	u4	49.09	43.64	62.29	48.39	41.48	63.94	49.27	45.55	61.76	48.21	42.39	57.06
	u5	41.36	32.09	59.35	42.45	31.97	61.35	40.67	34.36	59.76	41.12	33.30	56.65
	m1 HB	25.27	25.09	40.88	24.45	26.06	42.41	22.36	23.64	39.29	21.45	23.27	37.53

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
	m1 SL	25.18	25.12	40.88	24.58	26.03	42.12	22.48	24.03	39.12	21.39	23.61	37.76
	m2	34.21	29.52	53.76	32.48	28.79	56.41	28.09	28.21	49.24	27.21	28.24	45.82
μ5	u1 HB	38.58	43.21	52.29	37.18	43.12	53.00	38.15	42.52	49.35	40.76	44.36	49.18
	u1 SL	39.15	45.06	54.47	38.48	44.97	55.82	40.15	44.64	52.53	42.91	45.88	51.59
	u2 HB	37.09	40.91	52.47	36.58	42.91	54.24	38.64	41.45	52.88	41.76	45.18	56.59
	u2 SL	37.67	42.09	55.41	37.12	44.39	56.00	39.15	42.94	55.35	41.03	45.39	59.06
	u3 HB	29.24	35.06	50.47	30.30	35.67	51.18	29.97	35.45	50.47	32.64	35.73	51.82
	u3 SL	30.76	37.55	53.18	31.12	38.91	53.65	30.55	37.97	53.18	33.91	39.42	55.06
	u4	40.61	46.30	58.35	39.91	45.33	59.71	42.24	45.00	56.47	43.06	46.73	58.06
	u5	32.85	37.67	55.06	32.94	38.21	55.35	33.33	38.55	55.06	33.33	39.18	56.41
	m1 HB	69.70	52.33	83.06	66.97	51.18	83.24	66.70	51.42	84.18	59.33	45.82	80.47
	m1 SL	69.73	52.42	83.12	67.06	51.48	83.24	66.73	51.55	84.59	59.27	46.27	80.76
	m2	73.76	59.79	85.94	70.76	58.79	84.24	70.55	56.61	84.24	63.48	55.94	79.41
μ6	u1 HB	50.33	51.64	67.76	49.18	49.21	64.65	48.36	49.73	61.65	44.88	45.61	54.18
	u1 SL	50.82	52.88	68.94	51.67	50.03	65.88	50.03	50.03	64.41	48.27	46.91	56.29
	u2 HB	48.58	46.73	68.94	50.15	47.76	67.53	48.15	48.36	68.18	50.73	47.42	65.12
	u2 SL	49.64	47.27	69.12	51.36	48.67	68.18	48.03	48.36	70.06	49.64	47.27	66.06
	u3 HB	41.48	46.06	68.35	44.15	46.48	68.65	44.76	46.91	68.00	50.91	45.97	67.06
	u3 SL	42.12	44.97	68.35	43.79	47.85	70.35	43.85	46.64	68.47	49.03	46.15	68.71
	u4	54.88	53.73	69.76	55.18	52.61	68.59	52.06	50.82	66.65	52.58	48.94	61.06
	u5	54.91	51.91	69.59	54.61	51.55	69.94	53.33	52.88	71.06	56.55	51.88	69.41
	m1 HB	60.15	42.18	79.41	58.97	44.27	80.29	57.00	41.18	76.65	53.18	35.91	72.88
	m1 SL	60.06	42.18	79.24	58.82	44.27	80.41	56.82	41.39	76.65	53.21	36.30	72.59
	m2	64.91	48.64	81.12	64.85	49.73	82.94	62.18	49.76	79.06	57.00	47.58	76.12
μ7	u1 HB	46.64	49.48	59.41	46.18	47.76	59.59	43.21	47.97	56.88	43.27	45.55	52.59
	u1 SL	47.36	50.70	61.29	47.45	49.18	61.88	45.12	49.18	59.12	45.73	46.97	55.29
	u2 HB	45.88	47.76	60.82	46.36	46.30	61.24	44.73	46.85	62.94	46.85	47.15	60.41
	u2 SL	46.06	46.91	62.12	46.52	48.64	62.59	44.70	46.88	64.82	45.73	46.42	63.00
	u3 HB	39.18	46.24	60.18	41.52	46.36	60.82	39.48	47.27	61.24	44.91	48.06	59.88
	u3 SL	40.48	47.18	61.88	42.45	47.39	61.06	38.64	47.73	63.29	44.82	49.33	61.29
	u4	49.61	52.70	65.71	49.85	49.58	64.47	47.39	48.85	62.82	48.24	49.36	60.76
	u5	46.58	50.67	64.06	46.24	50.21	63.53	44.79	49.48	63.94	47.24	51.15	63.12
	m1 HB	73.42	53.58	85.06	70.64	53.30	84.18	71.82	52.03	82.82	65.00	48.39	80.53
	m1 SL	73.52	53.67	85.12	70.52	53.21	83.94	71.76	51.82	82.88	64.94	48.67	80.71
	m2	73.39	58.85	86.00	72.70	60.03	84.94	73.03	58.73	83.47	64.64	59.85	80.41

* Note: μ1 = (0,0,0), μ2 = (0.1,0.1,0.1), μ3 = (0,0.1, −0.1), μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. u1 is unidimensional method 1 (separate calibration at the total test level using one linking onstant); u2 is unidimensional method 2 (separate calibration at the total test level using three linking constats); u3 is unidimensional method 3 (separate calibration at the subtest level); u4 is unidimensional method 4 (concurrent calibration at the total test level); u5 is unidimensional method 5 (concurrent calibration at the subtest level); m1 is multidimensional method 1(separate calibration); and m2 is multidimensional method 2 (concurrent calibration). SL refers to the Stocking & Lord method, and HB stands for the Haebara method.

Table 4.21. Average Percentages of the Number of Score Points Equal between True and Estimated Conversion (Subtest 3)

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
μ1	u1 HB	60.64	58.12	74.65	56.36	54.00	72.18	50.82	47.79	68.82	40.85	38.82	59.41
	u1 SL	59.12	56.73	74.76	54.09	54.42	72.18	49.76	46.58	67.71	39.55	38.03	58.41
	u2 HB	57.42	49.48	73.00	58.36	49.21	69.12	54.94	49.15	68.06	48.03	42.33	61.71
	u2 SL	57.36	48.76	73.24	57.09	49.21	69.71	53.64	48.39	68.35	47.27	41.91	62.12
	u3 HB	46.58	47.94	73.06	47.15	46.70	69.76	47.55	48.00	73.00	59.70	47.91	71.53
	u3 SL	46.15	46.36	73.65	47.76	46.85	70.06	48.27	47.94	72.88	58.09	47.27	70.24
	u4	63.09	58.67	75.41	57.58	54.94	73.18	54.27	46.15	68.47	44.12	35.73	60.65
	u5	59.94	55.03	76.65	59.82	54.27	74.59	60.97	53.30	76.47	59.24	56.55	75.41
	m1 HB	33.03	38.03	56.12	35.48	39.00	57.18	33.21	36.79	54.29	34.79	37.52	53.06
	m1 SL	33.15	38.27	56.18	35.58	38.91	57.35	33.42	37.00	54.47	35.03	37.91	52.94
	m2	34.42	34.76	55.71	36.24	28.94	56.00	33.94	29.82	53.88	34.18	33.18	52.53
μ2	u1 HB	62.64	52.70	74.24	58.91	51.24	71.41	51.67	50.27	65.82	41.94	38.64	62.71
	u1 SL	61.97	52.70	74.53	57.06	51.21	71.76	51.03	50.36	65.41	40.09	38.12	60.59
	u2 HB	57.45	47.79	72.53	58.79	47.45	68.18	56.73	45.88	67.12	50.24	45.09	62.47
	u2 SL	56.39	46.36	72.47	57.09	46.42	68.65	54.76	46.15	67.41	49.64	46.42	62.47
	u3 HB	42.79	45.00	71.06	46.24	45.03	70.53	48.61	45.79	72.53	59.21	48.06	71.76
	u3 SL	42.88	44.27	71.35	46.36	43.85	69.65	47.94	44.30	72.29	58.30	47.70	71.18
	u4	64.82	51.58	74.76	62.33	52.18	73.94	57.09	47.76	68.00	47.52	40.79	64.76
	u5	61.27	46.94	74.12	62.21	48.52	75.41	64.27	49.85	75.00	64.42	47.55	75.82
	m1 HB	32.64	34.27	53.65	35.15	37.61	56.76	30.91	35.00	53.82	31.88	34.91	52.06
	m1 SL	32.82	34.12	53.53	35.15	37.55	56.65	31.09	35.18	53.76	32.03	35.09	52.00
	m2	33.06	36.58	53.06	34.55	27.76	56.29	31.36	28.12	53.18	31.42	28.97	52.06
μ3	u1 HB	21.73	18.39	53.18	22.76	20.45	51.35	21.21	20.30	49.53	22.55	20.52	45.59
	u1 SL	20.67	18.36	52.59	20.79	20.27	50.59	19.55	19.76	48.00	19.88	18.33	42.82
	u2 HB	57.79	48.15	73.24	58.88	47.21	70.76	54.79	45.18	67.12	43.45	40.88	61.76
	u2 SL	58.27	48.58	73.18	58.39	46.12	70.71	53.85	44.03	66.59	42.88	40.09	61.00
	u3 HB	44.82	47.55	72.29	46.18	46.91	71.59	50.15	49.03	73.53	59.27	49.15	74.00
	u3 SL	44.73	47.76	72.71	45.97	46.21	71.29	49.79	48.30	72.59	58.42	48.52	73.94
	u4	27.76	21.76	62.00	27.48	21.79	57.94	24.91	20.82	55.12	25.00	19.12	47.65
	u5	52.48	40.64	71.94	51.64	40.36	68.88	51.00	38.21	70.59	51.00	45.18	71.06
	m1 HB	38.52	39.58	54.53	41.06	44.18	56.82	35.27	38.24	54.06	38.97	42.21	53.65
	m1 SL	38.42	39.55	54.47	41.24	44.18	56.82	35.36	38.61	54.00	38.91	42.33	53.59
	m2	39.45	12.94	54.71	41.79	37.21	57.35	36.85	37.76	54.06	39.00	38.30	54.47
μ4	u1 HB	32.33	33.36	52.18	33.73	32.82	51.65	35.52	37.30	52.24	28.55	31.30	47.76
	u1 SL	33.18	34.76	52.71	34.52	33.24	52.41	34.97	37.33	52.12	27.67	31.06	45.76
	u2 HB	37.94	38.70	54.76	37.79	36.55	57.82	38.33	40.15	57.82	36.85	35.42	52.00
	u2 SL	36.64	38.24	55.71	36.85	36.30	57.18	37.88	40.30	57.35	35.85	35.18	51.88
	u3 HB	30.30	29.94	47.82	28.94	27.79	49.06	28.76	30.76	50.18	30.73	32.36	50.53
	u3 SL	29.45	29.67	46.53	28.58	26.58	48.41	28.70	30.24	48.41	30.30	30.58	50.47
	u4	37.03	32.97	57.06	36.45	27.36	56.76	37.76	32.61	52.47	33.24	25.30	46.00
	u5	28.27	20.67	50.53	27.94	18.61	52.29	28.27	20.33	52.76	27.91	20.85	52.71
	m1 HB	15.00	20.39	44.00	14.73	19.39	44.47	15.00	20.36	43.53	14.52	19.27	44.12
	m1 SL	15.03	20.42	43.82	14.76	19.39	44.29	15.06	20.42	43.24	14.61	19.52	44.06
	m2	15.06	51.97	46.65	14.85	21.15	46.76	15.06	21.21	46.71	14.76	21.21	46.59
μ5	u1 HB	65.58	51.91	72.59	62.79	53.91	72.29	52.15	45.52	67.82	40.30	37.15	60.65
	u1 SL	65.18	53.09	74.24	62.15	54.61	73.94	52.55	47.36	67.71	39.85	37.42	60.53
	u2 HB	61.61	50.85	72.29	62.91	50.33	73.24	58.58	44.97	70.24	51.24	41.76	63.41
	u2 SL	61.27	50.36	72.35	60.64	50.39	74.59	57.61	47.39	70.71	49.85	42.52	64.29
	u3 HB	46.27	46.42	70.12	47.45	44.73	73.12	49.24	46.18	72.76	54.73	46.03	71.24

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
	u3 SL	45.85	46.12	70.41	47.70	45.12	73.47	49.48	46.76	72.41	55.06	46.91	71.53
	u4	62.18	56.61	80.41	59.94	55.27	78.18	54.70	49.21	73.35	46.36	35.48	64.35
	u5	52.21	49.58	78.06	51.42	48.82	76.82	51.39	50.18	78.35	52.24	50.15	76.94
	m1 HB	52.36	56.30	65.18	51.73	56.88	66.65	49.27	53.36	61.41	49.18	52.97	60.71
	m1 SL	52.82	55.91	65.53	52.33	57.21	66.59	49.52	54.42	61.35	50.06	53.64	60.65
	m2	49.85	32.73	62.82	49.58	67.39	63.18	45.85	68.30	58.18	46.61	69.73	55.71
μ6	u1 HB	55.97	51.97	73.71	53.27	50.76	69.76	47.21	44.42	66.71	40.27	39.97	60.65
	u1 SL	54.39	51.70	74.06	51.82	49.79	69.76	46.76	44.15	66.59	39.67	39.00	59.76
	u2 HB	55.36	44.36	73.00	55.67	47.09	69.18	53.30	46.61	67.24	46.21	40.30	61.29
	u2 SL	54.36	43.91	72.41	55.15	46.03	68.82	52.36	46.61	66.76	45.30	40.64	61.24
	u3 HB	41.79	41.73	70.35	42.36	45.06	68.71	43.82	47.67	72.06	52.12	44.88	69.71
	u3 SL	41.24	41.67	71.06	43.09	44.00	68.71	43.85	46.12	72.00	50.39	43.79	69.00
	u4	60.18	54.61	73.47	54.88	51.64	70.35	51.82	43.88	67.47	41.97	35.12	59.65
	u5	56.00	50.24	73.41	55.45	48.61	71.12	56.64	51.73	73.82	54.55	49.18	70.76
	m1 HB	25.48	28.91	51.82	25.21	27.79	51.06	23.55	27.58	50.47	23.18	27.39	50.00
	m1 SL	25.36	29.00	51.76	25.24	27.73	51.24	23.55	27.52	50.41	23.55	27.67	49.94
	m2	27.45	37.39	52.29	26.52	22.61	50.88	24.45	22.76	50.53	25.48	22.79	49.59
μ7	u1 HB	64.55	56.73	73.71	61.76	54.21	73.82	53.45	48.15	68.65	43.88	38.64	60.53
	u1 SL	63.55	56.55	74.76	60.94	54.18	73.41	52.18	48.39	67.82	42.24	37.88	59.94
	u2 HB	61.18	50.45	74.06	61.09	48.09	72.18	58.79	48.45	70.94	51.18	44.12	64.35
	u2 SL	60.06	48.97	74.65	60.09	47.91	72.06	58.03	48.15	71.12	49.39	43.97	65.29
	u3 HB	50.64	47.82	73.35	49.97	48.82	72.59	52.33	49.85	74.24	62.12	49.03	74.82
	u3 SL	48.76	46.70	74.47	50.55	48.61	72.24	52.18	50.88	75.00	60.36	48.52	74.24
	u4	64.64	58.42	77.35	60.12	54.79	76.41	55.55	47.45	70.65	46.73	36.33	64.12
	u5	60.82	56.64	78.88	61.24	55.52	78.59	62.36	54.94	79.65	62.82	58.03	78.41
	m1 HB	43.61	45.27	59.41	45.61	47.18	61.18	43.24	44.27	56.65	42.64	44.73	55.24
	m1 SL	44.03	45.27	59.65	45.73	47.39	61.47	43.42	44.48	56.76	42.91	44.94	55.41
	m2	41.67	35.12	57.18	42.94	43.09	59.41	40.48	41.15	56.65	40.97	46.45	54.88

* Note: μ1 = (0,0,0), μ2 = (0.1,0.1,0.1), μ3 = (0,0.1, −0.1), μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. u1 is unidimensional method 1 (separate calibration at the total test level using one linking onstant); u2 is unidimensional method 2 (separate calibration at the total test level using three linking constats); u3 is unidimensional method 3 (separate calibration at the subtest level); u4 is unidimensional method 4 (concurrent calibration at the total test level); u5 is unidimensional method 5 (concurrent calibration at the subtest level); m1 is multidimensional method 1(separate calibration); and m2 is multidimensional method 2 (concurrent calibration). SL refers to the Stocking & Lord method, and HB stands for the Haebara method.

4.4.3.2 <u>Difference that matters (DTM)</u>

Table 4.22, Table 4.23, and Table 4.24 present DTM results, which also demonstrate the similar results described in the previous section. This means that the number of score points equal between true and estimated equating conversion tables were very similar to the number of score points where the difference between true and estimated conversions was less than 0.5.

Table 4.22. Percentages of DTM using Observed Scores (Subtest 1)

		$\rho = 0.9$			$\rho = 0.8$			$\rho = 0.6$			$\rho = 0.4$		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
μ1	u1 HB	65.27	58.91	78.41	60.15	58.45	75.00	59.27	53.39	77.18	51.30	48.70	65.94
	u1 SL	64.06	56.79	77.47	58.82	55.97	73.00	58.45	53.21	76.06	49.36	47.55	62.76
	u2 HB	61.03	51.97	75.47	63.06	57.21	74.94	65.79	55.91	78.18	67.52	52.64	79.24
	u2 SL	60.85	53.00	74.71	63.15	56.82	74.71	64.85	55.39	77.88	67.27	52.55	78.76
	u3 HB	49.52	50.12	73.18	54.15	55.82	72.59	48.79	53.58	76.24	62.58	51.91	76.24
	u3 SL	49.58	51.09	73.12	53.91	55.09	72.06	48.88	54.76	75.35	63.00	53.24	75.41
	u4	67.73	61.00	81.29	63.76	56.39	77.18	65.55	56.97	80.53	56.85	51.00	70.00
	u5	66.88	59.85	78.53	63.91	59.39	78.06	66.58	60.97	81.00	68.12	61.88	79.94
	m1 HB	52.39	53.70	62.29	56.91	57.61	66.47	54.94	55.94	67.76	53.73	55.06	69.29
	m1 SL	52.30	54.03	62.41	56.91	57.55	66.35	54.91	55.76	67.76	53.36	54.79	69.24
	m2	53.91	48.33	62.65	58.06	49.67	67.59	55.97	48.36	67.35	53.58	48.15	69.24
μ2	u1 HB	60.97	55.15	74.53	59.73	55.70	76.29	59.21	51.85	73.71	49.76	47.03	66.29
	u1 SL	58.67	52.97	73.47	58.03	55.24	75.12	58.03	51.00	72.65	49.09	46.21	64.29
	u2 HB	60.64	51.61	75.29	65.06	53.00	76.18	63.00	56.24	75.88	65.52	48.39	80.59
	u2 SL	59.39	50.97	74.94	64.27	53.30	74.65	62.15	56.06	75.18	63.30	48.39	80.71
	u3 HB	46.91	50.27	74.29	51.45	51.45	75.94	48.70	54.21	73.82	60.48	48.97	75.53
	u3 SL	47.03	49.52	74.35	51.45	50.58	74.41	48.30	52.91	72.88	58.06	49.61	75.12
	u4	66.64	64.67	80.65	66.21	62.12	81.18	66.61	58.79	80.71	58.27	52.64	74.29
	u5	64.79	59.67	80.41	70.61	62.55	80.59	67.73	64.30	82.88	68.82	58.52	83.12
	m1 HB	49.94	52.73	60.76	55.82	58.09	65.06	55.82	57.85	69.24	53.76	56.15	68.65
	m1 SL	49.88	52.21	60.59	55.61	57.76	65.00	55.88	57.97	69.06	53.15	55.76	68.35
	m2	51.55	47.12	61.59	56.18	48.85	66.59	56.30	47.88	69.59	55.42	47.33	70.35
μ3	u1 HB	66.61	57.15	77.41	63.94	55.52	76.71	61.27	55.73	75.65	50.79	50.36	67.06
	u1 SL	65.52	56.48	76.94	63.03	55.73	75.65	58.79	54.82	72.76	48.33	46.91	65.24
	u2 HB	61.58	50.88	77.47	64.82	55.79	77.65	67.48	56.24	78.59	64.58	58.27	78.53
	u2 SL	60.30	50.42	77.06	63.24	54.09	77.47	66.76	55.73	77.24	62.82	58.15	78.47
	u3 HB	52.33	48.42	74.12	53.39	54.21	76.12	50.94	53.24	74.71	59.70	54.67	74.53
	u3 SL	52.64	48.09	74.24	53.21	52.67	74.59	51.42	53.06	73.88	59.42	53.97	73.94
	u4	68.70	60.21	82.12	66.85	56.82	78.53	65.24	55.91	80.12	56.85	51.18	73.06
	u5	64.55	58.76	80.65	61.97	59.79	78.59	63.91	58.91	80.41	65.48	61.52	77.65
	m1 HB	52.00	53.48	60.29	55.12	56.58	68.18	54.76	58.27	68.53	53.24	55.30	68.94
	m1 SL	51.70	52.79	60.12	54.82	56.48	68.12	54.91	58.09	68.24	53.24	54.94	69.12
	m2	53.27	48.36	60.82	56.15	49.27	69.00	55.36	48.55	68.65	56.03	48.39	69.82
μ4	u1 HB	36.85	31.97	41.65	40.55	36.91	45.59	47.36	41.70	54.53	46.91	43.85	62.53
	u1 SL	35.21	32.70	41.00	39.18	36.88	45.29	47.03	41.61	53.41	47.58	43.06	61.53
	u2 HB	36.15	33.21	42.24	41.42	36.64	43.53	50.15	41.18	53.35	54.24	43.30	63.82
	u2 SL	35.06	33.67	41.71	40.00	36.52	43.12	50.06	41.30	52.88	52.88	42.27	64.59
	u3 HB	26.36	23.64	35.41	25.79	23.64	33.65	26.06	23.76	34.06	25.33	24.61	34.94
	u3 SL	24.94	23.55	35.65	24.36	23.70	34.29	24.85	23.58	34.12	24.67	24.03	34.88
	u4	48.45	35.52	52.65	51.61	39.76	55.18	55.88	43.79	62.35	48.00	41.52	64.47
	u5	29.70	22.58	38.29	29.48	22.91	35.53	30.48	23.00	37.59	29.39	22.67	37.47
	m1 HB	25.55	26.03	29.88	25.67	26.09	30.18	23.09	23.79	29.06	22.91	23.61	30.35
	m1 SL	25.73	25.85	29.88	25.42	25.97	29.76	22.73	23.79	29.00	22.36	22.82	29.94
	m2	31.21	29.94	33.29	31.70	29.27	33.41	28.09	27.76	31.76	27.94	26.91	33.06
μ5	u1 HB	59.39	57.70	80.29	54.58	54.55	78.00	51.82	50.24	71.94	39.61	40.85	56.65
	u1 SL	59.39	55.85	79.06	53.18	53.64	76.94	51.00	47.82	69.06	39.58	39.30	53.94
	u2 HB	61.39	49.39	79.24	60.33	52.52	76.82	66.15	56.39	80.53	64.67	51.70	81.06
	u2 SL	62.64	50.27	80.06	60.85	53.52	76.65	66.82	57.55	80.82	65.55	51.39	81.06
	u3 HB	44.91	46.45	76.59	45.73	51.45	74.47	42.88	52.06	78.41	57.33	48.18	78.41
	u3 SL	45.79	46.70	77.35	47.00	51.67	75.12	43.27	53.91	78.82	58.52	48.42	78.35
	u4	58.79	53.45	76.82	53.61	48.21	73.29	51.85	47.00	71.47	43.33	38.42	58.06

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
	u5	47.45	44.39	70.94	43.55	42.27	69.29	47.94	45.67	72.94	47.03	45.79	70.71
	m1 HB	57.18	57.91	73.88	54.91	57.15	75.88	55.58	57.27	77.82	46.33	46.30	74.06
	m1 SL	56.39	57.18	73.59	54.09	56.52	75.76	54.67	55.91	77.41	45.42	45.27	73.41
	m2	59.30	55.09	73.00	57.03	56.79	74.53	57.76	56.36	75.29	49.45	56.82	73.59
μ6	u1 HB	59.45	54.36	73.18	61.15	57.82	73.71	61.76	54.61	74.88	53.24	48.24	67.53
	u1 SL	58.94	55.06	72.24	59.45	55.67	72.71	60.30	53.97	71.94	52.09	46.24	65.12
	u2 HB	59.58	49.97	71.29	61.97	54.82	72.35	64.64	53.12	75.59	64.67	54.18	74.53
	u2 SL	58.88	49.45	70.71	61.30	54.76	72.29	63.73	52.85	75.29	64.79	53.76	75.47
	u3 HB	43.30	47.33	68.71	49.73	50.70	67.82	46.61	48.33	68.59	60.79	50.09	70.41
	u3 SL	43.30	46.30	68.71	49.94	50.39	67.41	46.03	48.09	67.53	59.76	48.64	69.06
	u4	65.12	57.36	77.94	65.64	57.09	78.18	68.27	58.82	79.59	58.91	53.39	71.53
	u5	62.64	54.09	75.94	60.61	53.00	73.76	64.21	55.12	75.53	66.70	56.09	73.47
	m1 HB	44.42	47.24	53.12	49.36	51.61	57.76	46.55	49.55	57.82	48.88	49.94	59.94
	m1 SL	44.33	47.06	53.18	49.45	51.36	57.65	46.52	49.45	57.53	48.55	49.36	59.94
	m2	46.42	44.76	56.12	51.58	44.73	60.47	49.58	45.03	59.53	51.42	44.03	63.47
μ7	u1 HB	64.64	59.76	80.00	61.27	60.06	76.47	56.67	55.30	75.29	47.03	48.15	64.00
	u1 SL	63.67	57.61	79.47	58.76	57.48	75.71	55.94	53.45	72.47	45.67	47.12	61.41
	u2 HB	64.33	55.45	78.82	66.18	59.00	75.82	65.73	57.06	77.76	65.64	52.06	80.00
	u2 SL	64.00	55.36	77.59	65.88	58.09	75.00	65.52	57.06	77.41	65.70	51.61	80.71
	u3 HB	51.82	53.12	77.06	54.82	57.82	75.29	51.97	56.42	76.94	64.91	53.15	81.06
	u3 SL	51.76	54.00	77.00	55.64	58.91	74.53	52.76	55.67	76.71	65.76	53.15	80.88
	u4	63.85	59.52	81.94	61.67	57.00	76.94	61.70	53.91	79.06	51.94	48.61	68.35
	u5	62.06	56.85	82.29	60.33	59.33	77.41	61.82	60.15	81.06	63.52	60.36	81.24
	m1 HB	55.52	57.88	69.18	57.15	60.94	71.00	56.73	59.79	72.76	52.18	51.70	72.35
	m1 SL	55.33	57.24	69.18	56.70	60.82	71.00	56.39	59.79	72.59	52.85	50.85	71.88
	m2	55.55	52.03	68.00	57.42	51.94	70.59	57.48	51.94	71.47	54.33	51.70	73.06

* Note: $\mu 1 = (0,0,0)$, $\mu 2 = (0.1,0.1,0.1)$, $\mu 3 = (0,0.1,-0.1)$, μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. u1 is unidimensional method 1 (separate calibration at the total test level using one linking onstant); u2 is unidimensional method 2 (separate calibration at the total test level using three linking constats); u3 is unidimensional method 3 (separate calibration at the subtest level); u4 is unidimensional method 4 (concurrent calibration at the total test level); u5 is unidimensional method 5 (concurrent calibration at the subtest level); m1 is multidimensional method 1(separate calibration); and m2 is multidimensional method 2 (concurrent calibration). SL refers to the Stocking & Lord method, and HB stands for the Haebara method.

Table 4.23. Percentages of DTM using Observed Scores (Subtest 2)

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
μ1	u1 HB	56.09	56.91	79.35	56.45	55.64	79.82	56.24	52.70	73.82	53.91	52.33	67.47
	u1 SL	55.55	55.73	80.06	56.48	56.33	80.94	56.82	53.30	75.65	55.12	54.91	70.18
	u2 HB	55.30	50.64	76.71	54.30	51.45	78.53	54.67	50.45	74.12	54.45	52.85	70.35
	u2 SL	55.70	52.09	77.53	55.64	52.55	79.24	56.55	50.12	73.06	55.55	52.73	72.00
	u3 HB	44.18	52.03	75.82	48.48	50.64	75.94	49.70	48.91	74.76	55.30	51.67	77.00
	u3 SL	45.27	52.33	78.65	48.52	51.79	77.53	49.24	48.12	75.71	56.06	52.67	77.94
	u4	58.73	60.42	82.12	60.33	61.06	82.94	60.67	57.48	81.76	61.61	58.67	79.53
	u5	58.94	60.09	82.41	58.36	57.73	81.82	59.58	54.79	83.65	59.18	57.64	84.12
	m1 HB	59.79	61.58	81.94	57.30	61.27	80.35	56.18	57.82	82.41	50.85	55.52	81.41
	m1 SL	59.85	61.36	82.06	57.18	61.33	80.29	56.58	57.58	82.47	50.88	54.88	81.53
	m2	60.94	81.79	82.59	60.33	81.24	79.82	57.36	79.58	83.18	50.52	80.67	81.88
μ2	u1 HB	55.55	56.55	80.94	56.85	55.33	78.29	53.70	52.33	74.24	49.67	48.42	65.29
	u1 SL	55.30	54.67	81.29	57.67	57.00	79.06	56.09	54.21	76.47	51.97	51.15	68.29
	u2 HB	53.73	54.06	79.12	58.21	52.42	77.82	56.97	49.79	73.94	57.18	50.52	68.82
	u2 SL	53.97	52.00	79.41	57.21	53.03	79.24	57.88	49.76	75.18	56.48	48.97	68.41
	u3 HB	42.85	52.45	78.35	46.94	48.27	76.71	49.55	50.36	76.94	54.85	51.21	77.12
	u3 SL	43.03	52.03	78.71	45.48	50.55	78.35	48.67	51.09	76.71	54.52	49.70	75.65
	u4	57.21	56.76	79.06	58.03	53.58	79.82	59.18	49.88	75.88	54.36	45.94	71.76
	u5	53.58	49.79	74.41	55.24	47.91	75.35	54.91	47.91	74.88	56.76	47.00	73.82
	m1 HB	58.85	64.30	82.41	60.97	66.94	81.06	56.79	60.67	82.35	54.94	57.88	81.88
	m1 SL	58.61	64.18	82.53	60.55	66.52	80.94	56.03	59.88	82.29	54.76	58.03	82.29
	m2	59.21	83.36	81.53	59.76	83.76	80.71	57.00	79.45	82.35	54.58	81.15	82.53
μ3	u1 HB	31.06	32.33	53.12	28.70	32.03	55.88	27.48	30.48	49.71	26.06	29.91	45.94
	u1 SL	31.85	32.88	53.41	30.12	32.24	56.18	29.33	30.97	50.71	28.52	30.91	48.41
	u2 HB	57.00	51.00	75.76	55.73	51.21	74.76	53.27	49.70	72.18	55.42	48.94	67.47
	u2 SL	56.76	51.30	76.71	55.30	50.18	76.82	54.55	49.15	72.41	56.88	48.79	68.24
	u3 HB	45.67	53.12	75.65	45.94	47.76	75.82	48.52	50.64	76.47	54.24	52.48	75.59
	u3 SL	45.09	50.27	76.53	45.88	47.06	76.47	47.85	49.45	77.53	55.24	50.97	77.00
	u4	38.82	35.00	62.53	40.18	35.70	64.35	36.70	32.48	60.24	32.64	32.67	56.47
	u5	56.55	46.76	74.18	55.64	50.48	75.65	58.06	48.36	74.71	56.15	47.03	73.29
	m1 HB	60.18	60.70	83.12	60.64	62.27	80.24	57.30	60.58	83.59	53.12	57.76	82.35
	m1 SL	60.06	60.67	83.12	60.88	62.70	80.00	56.79	60.18	83.53	53.00	57.30	82.47
	m2	60.61	80.30	83.47	61.48	81.09	79.82	58.39	79.91	83.88	53.94	82.97	82.59
μ4	u1 HB	40.76	42.12	68.82	46.58	45.18	72.12	47.52	50.21	76.47	51.55	52.67	66.47
	u1 SL	38.91	41.21	68.71	44.94	44.70	72.12	47.24	50.21	76.65	53.21	54.12	68.24
	u2 HB	40.39	38.30	65.59	44.18	39.24	69.88	48.85	49.67	75.24	56.79	53.45	72.41
	u2 SL	40.33	39.00	66.35	44.09	39.76	69.41	50.30	49.18	76.24	55.91	52.73	73.12
	u3 HB	27.24	30.42	60.06	29.00	29.39	62.35	28.24	30.58	59.06	28.52	27.91	58.12
	u3 SL	27.24	31.15	61.12	28.52	30.12	62.18	28.00	31.09	59.00	27.55	28.79	59.94
	u4	56.88	49.91	76.53	58.45	51.61	78.82	57.64	59.06	80.94	57.61	53.88	75.35
	u5	41.42	35.48	70.29	43.18	34.91	71.53	39.73	36.09	69.59	41.88	34.91	68.53
	m1 HB	34.03	35.88	57.94	34.03	35.33	60.12	29.70	30.73	54.53	26.52	27.39	51.76
	m1 SL	33.88	35.58	58.06	34.03	35.58	59.82	29.79	30.94	54.47	27.00	27.70	51.41
	m2	40.73	43.03	71.00	40.48	39.45	71.76	34.30	37.61	65.94	31.67	35.42	61.24
μ5	u1 HB	49.06	43.39	64.59	46.06	41.94	65.94	50.30	44.15	59.41	51.00	47.33	58.65
	u1 SL	51.15	44.82	67.41	49.85	45.18	69.12	53.06	46.48	62.47	52.79	50.58	63.29
	u2 HB	47.91	42.12	64.35	44.94	40.24	61.47	47.97	40.06	58.12	51.15	46.70	62.65
	u2 SL	49.91	42.45	66.59	46.39	41.64	64.29	49.67	42.15	60.41	50.82	46.52	63.76
	u3 HB	39.36	36.21	59.65	39.06	34.94	57.94	39.52	35.67	57.18	42.61	37.70	60.18
	u3 SL	41.33	37.97	63.06	41.09	37.52	60.12	40.24	37.55	62.12	45.06	38.85	63.82
	u4	48.88	43.33	71.06	48.24	42.73	73.18	52.61	45.64	69.82	53.88	49.36	73.18

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
	u5	40.06	35.21	64.41	39.94	34.55	63.47	40.82	34.55	63.47	40.15	35.45	65.41
	m1 HB	53.18	45.12	69.47	49.36	41.85	67.82	48.45	40.09	72.41	43.82	39.12	76.00
	m1 SL	53.58	45.48	69.12	49.33	41.97	67.71	48.85	40.33	72.24	44.00	39.24	76.29
	m2	57.12	51.00	74.94	53.73	48.33	73.00	53.33	38.09	77.65	48.12	39.70	78.94
μ6	u1 HB	55.42	55.58	82.12	57.09	58.91	80.29	56.55	55.85	77.00	54.21	54.15	68.71
	u1 SL	54.73	55.58	82.59	57.48	59.52	81.65	57.97	57.48	78.18	55.09	56.15	71.76
	u2 HB	54.85	52.39	80.24	55.79	55.55	80.00	54.15	54.91	76.24	57.42	55.36	75.18
	u2 SL	55.64	52.33	80.12	57.45	55.88	81.65	54.39	54.88	76.18	57.03	55.61	75.47
	u3 HB	42.36	50.24	79.53	44.48	51.45	81.35	47.64	50.27	79.06	53.76	46.97	80.35
	u3 SL	41.67	50.82	79.47	44.64	52.91	82.00	45.91	50.58	78.59	51.88	48.94	81.53
	u4	60.76	63.09	84.59	63.15	65.18	84.88	63.30	62.88	83.29	63.73	60.03	79.71
	u5	60.15	60.73	84.35	60.91	58.94	86.00	59.91	62.55	86.18	62.70	61.64	85.82
	m1 HB	53.79	60.58	84.00	54.39	61.67	82.06	50.58	57.67	83.88	46.42	52.55	80.29
	m1 SL	54.12	60.36	83.76	54.48	61.55	82.12	50.48	57.09	83.71	45.94	52.58	80.18
	m2	55.88	81.21	85.24	58.15	81.24	82.88	53.52	79.58	84.65	47.15	79.36	82.12
μ7	u1 HB	57.00	56.33	75.29	56.15	51.39	75.06	54.06	54.30	71.41	54.24	49.85	66.12
	u1 SL	57.76	57.12	76.41	58.00	53.73	76.76	55.45	56.09	74.18	56.58	53.42	69.53
	u2 HB	54.24	50.36	72.82	54.18	48.33	73.06	54.67	49.33	71.41	56.03	51.03	69.59
	u2 SL	56.06	49.73	74.24	54.36	48.36	73.82	54.82	48.91	72.12	54.76	51.27	70.82
	u3 HB	45.73	48.58	71.47	47.55	45.48	71.18	47.12	46.67	72.18	52.33	49.85	72.00
	u3 SL	46.36	49.27	73.82	48.21	46.91	72.82	47.52	46.67	73.65	53.42	50.82	72.29
	u4	58.45	57.76	79.24	58.88	54.52	80.35	59.42	58.30	79.18	59.85	54.64	76.76
	u5	55.27	52.33	78.06	54.33	49.79	77.76	53.88	51.15	79.24	55.12	52.64	77.88
	m1 HB	58.12	59.88	77.65	57.82	54.61	74.65	57.30	55.09	79.94	52.18	51.33	79.76
	m1 SL	58.55	60.12	77.76	57.94	54.64	74.82	57.82	55.00	79.88	52.15	50.88	79.65
	m2	59.52	77.94	80.65	60.73	75.12	76.35	58.79	71.33	81.06	51.33	73.18	79.76

* Note: $\mu1 = (0,0,0)$, $\mu2 = (0.1,0.1,0.1)$, $\mu3 = (0,0.1,-0.1)$, μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. u1 is unidimensional method 1 (separate calibration at the total test level using one linking onstant); u2 is unidimensional method 2 (separate calibration at the total test level using three linking constats); u3 is unidimensional method 3 (separate calibration at the subtest level); u4 is unidimensional method 4 (concurrent calibration at the total test level); u5 is unidimensional method 5 (concurrent calibration at the subtest level); m1 is multidimensional method 1(separate calibration); and m2 is multidimensional method 2 (concurrent calibration). SL refers to the Stocking & Lord method, and HB stands for the Haebara method.

Table 4.24. Percentages of DTM using Observed Scores (Subtest 3)

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
μ1	u1 HB	65.00	65.79	80.76	57.33	60.12	79.00	51.64	54.18	71.76	40.18	42.15	60.12
	u1 SL	62.88	64.15	80.88	56.70	59.67	77.88	49.48	52.48	70.41	38.06	39.55	58.18
	u2 HB	62.18	54.48	77.82	62.03	56.55	77.94	55.52	53.91	73.82	48.48	44.79	64.06
	u2 SL	61.18	54.91	77.24	60.48	56.76	77.65	55.27	52.85	73.12	48.42	45.15	62.59
	u3 HB	48.12	53.76	78.00	49.45	56.61	78.12	48.12	55.82	78.47	64.85	53.94	78.12
	u3 SL	48.94	51.58	77.71	49.94	55.64	78.00	47.88	55.85	77.24	63.79	53.88	77.65
	u4	66.52	68.52	81.65	59.79	61.39	79.24	54.85	51.33	72.88	43.94	38.91	60.24
	u5	65.48	64.00	80.53	67.30	64.39	79.94	69.76	64.82	81.06	66.55	66.24	81.18
	m1 HB	32.27	46.64	71.59	35.00	48.58	76.88	34.27	45.39	70.53	34.36	46.03	71.12
	m1 SL	32.27	46.79	71.65	35.03	48.30	76.88	34.70	45.36	70.88	34.55	46.39	71.41
	m2	33.18	36.79	73.00	35.33	45.36	76.94	33.00	46.42	71.76	32.85	50.27	71.88
μ2	u1 HB	65.88	60.09	81.24	61.61	58.82	76.94	51.30	55.12	71.24	41.97	41.42	63.65
	u1 SL	65.30	61.21	81.06	58.91	58.52	75.94	49.39	54.45	70.24	38.76	40.64	61.00
	u2 HB	62.55	53.45	77.47	61.36	54.82	76.12	56.42	52.24	73.29	48.61	48.58	65.82
	u2 SL	61.12	51.79	76.88	59.91	53.24	75.76	55.36	52.55	73.53	48.03	48.15	65.47
	u3 HB	43.82	51.82	76.82	49.36	53.42	76.53	48.55	53.30	78.35	62.12	54.79	77.06
	u3 SL	43.48	49.61	75.88	49.52	51.64	77.06	48.67	52.73	79.18	60.61	53.97	77.29
	u4	71.24	64.67	83.76	65.64	64.67	81.94	57.67	55.42	73.94	48.00	46.33	67.29
	u5	67.88	59.52	84.47	71.18	62.88	83.82	71.73	62.91	83.88	71.76	61.06	84.53
	m1 HB	31.30	40.67	70.59	32.09	44.18	76.35	30.42	42.36	69.59	32.00	41.39	69.65
	m1 SL	31.33	40.58	70.71	32.27	43.97	76.35	30.52	42.73	70.12	32.21	41.94	69.88
	m2	31.00	37.70	71.29	32.09	43.12	76.18	28.82	42.70	68.88	31.30	44.06	69.35
μ3	u1 HB	20.91	21.67	49.35	21.45	23.58	49.12	20.76	22.94	45.12	20.91	21.48	41.29
	u1 SL	19.85	21.67	48.88	20.12	23.52	47.82	19.42	22.15	42.82	19.64	20.18	38.06
	u2 HB	64.33	56.70	80.00	62.85	53.61	73.06	54.58	50.97	71.18	42.61	42.09	62.18
	u2 SL	63.21	57.03	78.94	61.79	53.15	73.12	53.79	49.76	70.35	41.88	41.42	61.29
	u3 HB	47.97	58.06	79.65	48.67	53.97	76.24	51.97	58.27	80.00	66.00	57.30	79.82
	u3 SL	46.64	57.61	78.59	48.76	53.85	76.35	51.21	57.64	79.29	64.24	56.61	78.82
	u4	24.27	23.58	59.41	24.00	24.52	55.65	21.97	21.79	52.18	22.30	21.27	44.65
	u5	54.00	44.21	73.24	55.06	41.76	67.53	54.18	40.06	71.35	53.39	46.48	72.88
	m1 HB	38.97	50.82	72.24	39.18	51.00	77.29	38.06	46.67	71.94	43.82	54.27	70.12
	m1 SL	38.64	50.76	72.29	39.15	51.27	77.35	38.24	47.67	72.06	43.88	54.73	70.41
	m2	40.82	18.91	75.53	40.52	55.15	79.12	37.64	54.45	74.71	44.67	59.12	72.06
μ4	u1 HB	30.00	31.18	45.18	31.85	31.79	45.88	34.09	36.52	46.35	27.39	29.15	42.82
	u1 SL	30.97	33.18	46.06	32.12	31.97	46.24	34.18	36.79	46.24	27.00	28.52	42.18
	u2 HB	35.73	38.03	49.53	35.55	35.12	51.24	36.82	39.73	51.35	34.21	35.58	49.82
	u2 SL	34.45	37.03	49.18	35.00	33.91	51.35	35.82	39.15	52.00	34.06	34.61	49.18
	u3 HB	29.33	28.12	42.76	27.88	25.73	43.59	27.58	28.55	43.47	27.94	30.39	44.71
	u3 SL	27.88	27.61	41.41	27.55	24.45	43.12	26.70	27.82	42.53	28.03	28.45	43.94
	u4	33.00	28.39	50.47	32.39	26.42	49.35	35.85	32.18	45.82	31.21	26.24	42.82
	u5	25.27	20.76	43.29	24.94	19.73	43.65	25.24	20.79	45.06	24.94	21.30	44.00
	m1 HB	13.52	15.06	33.41	12.52	13.85	34.59	13.00	14.21	33.94	13.00	14.36	33.88
	m1 SL	13.48	15.09	33.53	12.58	13.91	34.76	12.82	14.18	33.47	12.94	14.15	33.82
	m2	14.88	51.27	38.35	13.94	15.09	39.29	14.52	14.73	37.41	14.61	14.82	36.24
μ5	u1 HB	75.94	64.61	83.82	71.85	66.18	82.53	57.21	54.45	77.00	43.00	41.91	66.41
	u1 SL	75.06	65.42	84.65	70.30	65.06	82.82	55.21	57.18	76.76	41.48	42.18	65.35
	u2 HB	74.00	57.70	81.18	72.33	56.36	82.82	66.48	53.39	79.88	56.03	48.24	70.71
	u2 SL	72.06	56.03	79.65	70.03	55.76	81.88	66.30	54.21	79.47	54.79	47.82	70.00
	u3 HB	53.27	52.15	79.00	53.15	50.18	81.18	58.15	50.48	81.59	67.85	52.06	82.00
	u3 SL	52.48	51.85	78.88	52.58	49.45	81.29	57.82	51.42	81.47	66.73	53.03	80.71
	u4	76.61	70.42	84.12	71.06	66.24	82.29	60.67	59.52	79.29	49.64	40.94	68.18

		ρ = 0.9			ρ = 0.8			ρ = 0.6			ρ = 0.4		
		Items & Common items			Items & Common items			Items & Common items			Items & Common items		
		32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4	32 & 8	32 & 4	16 & 4
	u5	69.85	58.39	80.06	69.30	57.03	78.12	70.39	59.91	78.82	69.00	60.18	78.59
	m1 HB	66.03	69.18	82.65	64.97	68.48	82.35	62.21	66.82	79.76	62.70	65.76	78.76
	m1 SL	66.45	69.58	82.82	65.61	68.45	82.35	62.55	67.12	79.59	63.27	66.36	78.41
	m2	58.64	32.30	81.29	59.85	82.91	81.76	56.67	83.64	77.94	58.79	83.03	76.59
μ6	u1 HB	59.03	57.30	77.29	53.82	54.82	70.88	48.88	46.85	69.53	39.82	41.76	60.12
	u1 SL	57.36	57.42	76.94	52.30	54.09	70.71	47.55	47.15	67.82	38.42	40.94	58.18
	u2 HB	58.36	50.88	78.35	56.85	51.58	72.59	55.33	50.61	70.12	45.52	44.12	63.82
	u2 SL	56.55	50.03	77.59	56.36	50.09	73.12	53.48	50.88	69.82	44.70	43.21	63.35
	u3 HB	42.88	47.55	74.65	43.67	49.67	72.76	44.67	53.88	75.00	53.58	50.85	73.65
	u3 SL	41.97	46.76	74.94	43.39	47.03	72.53	45.33	52.03	74.76	51.73	48.73	72.82
	u4	62.91	59.73	80.71	56.73	54.79	75.76	51.09	47.30	70.29	40.36	38.30	60.76
	u5	57.79	56.06	79.35	58.82	52.82	76.00	59.82	55.39	77.76	55.39	53.79	76.06
	m1 HB	24.12	32.09	66.53	22.70	29.88	64.18	22.67	29.58	62.88	22.39	29.21	60.18
	m1 SL	24.18	32.36	66.47	22.76	29.82	64.41	23.00	29.64	63.00	22.67	30.12	60.24
	m2	27.42	37.45	68.76	25.06	32.09	68.76	22.97	31.48	66.94	24.09	32.58	64.06
μ7	u1 HB	70.27	67.79	82.29	65.52	63.27	79.24	54.91	54.73	74.94	44.03	42.00	63.71
	u1 SL	69.55	66.36	82.35	63.24	62.91	79.29	52.33	53.12	74.35	42.12	40.67	62.59
	u2 HB	66.70	57.91	80.24	65.97	55.85	78.18	61.36	54.09	76.53	52.24	47.55	69.94
	u2 SL	66.12	55.58	80.82	65.79	56.67	77.24	61.03	54.15	76.24	50.67	47.58	68.47
	u3 HB	53.73	55.21	79.82	54.48	58.24	79.82	56.58	60.85	81.35	70.70	56.39	81.76
	u3 SL	52.55	53.67	80.59	53.94	58.21	79.00	55.15	60.06	80.24	68.52	55.27	81.06
	u4	71.91	70.21	84.00	66.06	64.27	81.71	58.88	53.00	75.59	46.73	40.36	65.41
	u5	72.76	69.24	83.12	72.82	68.61	81.59	73.79	69.03	82.18	74.52	69.36	82.53
	m1 HB	46.30	55.61	76.47	48.24	57.64	81.18	49.06	55.70	75.41	47.52	56.64	74.12
	m1 SL	46.42	55.52	76.71	48.55	57.97	81.00	49.33	55.73	75.88	48.03	56.79	74.59
	m2	42.55	36.70	75.53	43.24	61.36	80.18	42.91	61.30	75.29	44.45	65.55	74.76

* Note: $\mu1 = (0,0,0)$, $\mu2 = (0.1,0.1,0.1)$, $\mu3 = (0,0.1,-0.1)$, μ4 has skewness of .75, μ5 with -.75, μ6 with .25, and μ7 with -.25. u1 is unidimensional method 1 (separate calibration at the total test level using one linking onstant); u2 is unidimensional method 2 (separate calibration at the total test level using three linking constats); u3 is unidimensional method 3 (separate calibration at the subtest level); u4 is unidimensional method 4 (concurrent calibration at the total test level); u5 is unidimensional method 5 (concurrent calibration at the subtest level); m1 is multidimensional method 1(separate calibration); and m2 is multidimensional method 2 (concurrent calibration). SL refers to the Stocking & Lord method, and HB stands for the Haebara method.

4.5 Discussion

Subtest score equating methods based on both unidimensional and multidimensional item response theory were applied and compared under several study conditions including proficiency distributions, a number of items and common items, and correlations among dimensions.

The findings from this study are as follows:

First, item parameter recovery results showed that when separate calibration was implemented, correlations between true and estimated slope parameters decreased as correlations among dimensions decreased. In the case of concurrent calibration, however, correlations of the slope parameters remained consistent regardless of correlations among dimensions. Correlations between the true and estimated location parameters were all above 0.9 across all conditions. When multidimensional item response theory was applied, the correlations between true and estimated parameters were slightly higher than those from unidimensional item response theory. In addition, under the skewed proficiency distribution (.75 skewness) from one group of examinees, correlations between true and estimated item parameters were the smallest among the seven conditions of different proficiency distributions.

Second, multidimensional IRT methods did not perform well compared to unidimensional IRT methods. Strictly speaking, however, it is not clear whether equating errors originated from the unidimensional approximation procedure or from equating itself. Based on the results of this study, using unidimensional IRT methods could offer a better choice considering the accuracy of equating results and efficiency of time. The results from subtest 1 and subtest 3 supported this finding. On the other hand, subtest 2 showed that multidimensional methods could produce better results in some cases. Given the computation process and time, however, it took about ten minutes to run one data set using a multidimensional IRT model whereas it took only a few seconds to run the same data set using a unidimensional IRT model. Note that this study had 1,000 examinees and either 96 or 48 items in total. Using a multidimensional IRT model may not be efficient if data sets are very large including many items and examinees.

Third, among the unidimensional methods, when the correlations among dimensions were high, separate calibration at the total test level with 1 linking constant performed better than separate calibration at the total test level with 3 different linking constants. Separate calibration at the subtest level usually did not perform as well as these two methods at the total test level. On the other hand, when the correlations were low, the second unidimensional method—separate calibration at the total test level but using three different linking constants—had better results than the first method which used just one linking constant. Of the two concurrent calibration methods, the method at the subtest level showed better results than the one at the total test level when correlations were low. There were some exceptions when the distributions were either shifted to different directions or skewed with relatively high skewness. This pattern was less consistent under those distributions.

Fourth, between separate calibration and concurrent calibration methods, in general, the concurrent calibration method at the subtest level, among the five unidimensional methods, tended to perform better than the three separate calibration methods except when the distributions were skewed with a relatively high skewness. Concurrent calibration at the total test level was more likely to have better results than separate calibration methods, but not for the low correlation condition. In that case, separate calibration at the total test level with three linking constants or at the subtest level showed slightly better results. In multidimensional cases, separate calibration generally showed better results than concurrent calibration. However, when the number of items and common items is small—16 items including 4 common items—the concurrent calibration method produced slightly better equating results than the separate calibration method.

Fifth, the skewed distribution with a skewness of .75 and the distribution with a mean shift to different directions for each dimension produced less accurate equating results compared to the other distributions. Lastly, there was no one linking method that performed better than others in all conditions. Under some conditions, the Haebara method produced slightly more accurate results whereas the Stocking and Lord method worked slightly better in other conditions.

In general, using multidimensional IRT parameters did not produce better equating results in terms of the accuracy of the results and efficiency of time and computation. Regarding accuracy, this could have resulted from this study not implementing a full multidimensional IRT equating method but rather the unidimensional approximation method using the multidimensional IRT parameters. Future research should compare the results of this study with full MIRT equating. If the results provide relatively similar or slightly improved results as compared to the results from unidimensional methods, applying multidimensional methods would not be worthwhile applying multidimensional methods in reality. Moreover, the current study only dealt with a simple structure model. If different models, such as compensatory or non-compensatory MIRT models, were applied, the item parameter recovery or equating results could have been different. Although the procedure is mathematically complicated and there is no software package currently available for MIRT equating, it is important to use more realistic models to investigate better methods for subtest score equating based on item response theory framework.

CHAPTER 5

DISCUSSION

5.1 Summary

Subtest score equating methods based on both classical test theory and item response theory were examined and compared under several conditions including correlations among dimensions, the number of items and common items, and the proficiency distributions. In addition to these conditions, three different input scores for equating—observed scores, weighted averages, and augmented scores—as well as three different anchor sets—anchor scores from each subtest, equated total scores, and anchor total scores—were used within the classical test theory framework. Under the item response theory, both unidimensional and multidimensional item response theory as well as concurrent and separate calibrations were adopted to estimate item parameters.

As preliminary studies, proportional reduction in mean squared error (PRMSE) and multidimensional scaling (MDS) analysis were conducted. First, PRMSE values were computed to examine whether subtest scores represent added value over the total test score. When the correlations among dimensions were moderate or low, subtest scores did provide additional information over the total score. With high correlations, however, subtest scores did not have added value. Second, MDS analyses demonstrated that each dimension or each subtest tended to group together on the stimulus space as correlations among dimensions decreased. Item parameter recovery studies via correlations between true and estimated parameters were also completed as preliminary analyses prior to IRT equating. Regardless of the IRT models, either a unidimensional or a multidimensional model, the correlations between true and estimated location parameters were greater than 0.9. For the slope parameters, correlations varied depending on the calibration methods and IRT models. It was more likely that the multidimensional IRT model and concurrent calibration produced higher and more consistent correlations across the simulation conditions, especially the correlation among dimensions. Among seven different proficiency distributions of FormX, the skewed distribution with skewness of .75 produced the worst item parameter recovery results.

Subtest score equating methods applied in the first simulation study were based on classical test theory. Among three input scores, augmented scores produced smaller equating error than the other two scores; observed scores outperformed weighted averages in most cases. As correlations among dimensions decreased, using the subtest anchor score tended to yield more accurate equating results than the other two methods. In some cases with more items or common items, the subtest anchor score method still provided better equating results despite the high correlations. In general, when the mean of the proficiency distribution was shifted to different directions for each dimension, $\mu = (0, 0.1, -0.1)$, equating results were less accurate compared to the distribution without mean shift or the one with mean shift to the same direction for all dimensions, $\mu = (0.1, 0.1, 0.1)$. Moreover, the skewed proficiency distribution with relatively large skewness (.75) produced the largest error while the distributions with relatively small skewness (.25) performed similar to the base distribution with a mean vectors of 0s, which was also observed in the second simulation study based on item response theory. The second simulation study showed that multidimensional IRT methods generally did not produce more accurate results than unidimensional IRT methods although there were some cases in which the multidimensional IRT methods performed better than the unidimensional methods in subtest 2. There was no obvious pattern as the correlation among dimensions decreased or increased.

5.2 Conclusions and Discussion

This study examined and compared subtest score equating methods using three important factors—proficiency distributions, correlations among dimensions, and test length. In both classical test theory and IRT based methods, when the proficiency distribution was skewed with a relatively large skewness (.75) or shifted to different directions for each dimension from $\mu = (0,0,0)$ to $\mu = (0,0.1,-0.1)$, equating errors were larger than the other distributions. Operational data used to obtain true simulation parameters showed a negatively skewed pattern. Generated data sets also shared this pattern; thus the distribution of FormY (old form) had this property. Adding more positive skewness to the distribution of FormX (new form) moved the distribution to the opposite direction from FormY, which caused a larger difference between the two forms. Compared to the other distributions used in the simulation, the skewed distribution with a skewness of .75 and the distribution with a mean shift $\mu = (0,0.1,-0.1)$ showed larger differences between FormY and FormX which could lead to large equating errors.

Under the classical test theory framework, equating results showed a clear pattern—when correlations were high, using the total or anchor total score as the anchor performed better than using the anchor score from each subtest in many cases. Preliminary analyses using PRMSE and MDS also demonstrated that when correlations were high, subtests did not confer added value over the total test or did not create clear and distinctive clusters. Equating results from item response theory based methods did not represent a consistent pattern in relation to correlations although item parameter recovery results showed that when separate calibration at the total test level from unidimensional methods was applied, correlations between the true and estimated slope parameters slightly decreased as correlations among dimensions decreased. When concurrent calibration and separate calibration at the subtest level were implemented, item parameter recovery results remained consistent across different conditions of correlations. Analyzing data, which has moderate or low correlations among dimensions, using separate calibration at the total test level via unidimensional IRT models would not be accurate. In this case, either choosing multidimensional IRT models or calibration at the subtest level using unidimensional IRT models would produce better results.

Having more items or more common items improved equating results in classical test theory based methods. In item response theory, however, including more items did not guarantee better equating results although it produced higher correlations between true and estimated slope parameters compared to having a smaller number of items. In addition to these three factors, different methods within each framework were examined and compared. When the methods based on classical test theory were employed, using augmented scores and having more items and common items produced slightly better equating results than the other conditions. Among IRT based methods, concurrent calibration using the unidimensional model slightly outperformed the other methods when the two proficiency distributions were normal with no mean shift or one of the distributions was skewed with a small skewness. Multidimensional IRT methods with the unidimensional approximation procedure did not perform as well as unidimensional methods in many cases. However, it was not obvious whether the major source of error came from equating or from the approximation procedure. Further research is required for this unsolved problem.

5.3 Limitations and Future Research

This study applied and compared possible methods for subtest score equating under some important factors. There are several limitations in this study which could be used as the basis future research. First, item responses simulated in this study were based on the simple structure model where each item is loaded on one dimension; however, this does not represent a realistic situation. In a future study, item responses can be generated using more complex models such as the bifactor model, compensatory model, or non-compensatory model. Moreover, this study applied only a three parameter logistic model. Different multidimensional IRT models can be considered in the IRT analysis.

Second, this study adopted three dimensions and the score differences between FormX and FormY in each subtest were relatively small. Although it may not be perfectly realistic, it is important to investigate the impact of mean score differences on the equating results using two forms with relatively large score differences. The current study manipulated mean vectors and skewness. In future research, other moments such as standard deviation and kurtosis can be used.

Third, multidimensional methods did not yield less equating error than unidimensional methods although multidimensional IRT models produced better item parameter recovery results. Given this finding, it would be worthwhile comparing the results directly from multidimensional IRT equating rather than having two steps of the unidimensional approximation procedure and unidimensional IRT equating. The approximation procedure could have created the major source of error.

Finally, due to the computation time in MIRT models, this study had one condition for the sample size—1,000 examinees. Larger or smaller sample sizes than this number could form another simulation condition. Although perhaps not a critical issue for classical test theory based methods, for IRT based methods, the computation time and mathematical complexity of the models could be an important to factor to consider when choosing the best method under a specific condition. It may be impractical to select a method which provides slightly more accurate results than the other ones but it takes very long to implement.

APPENDIX A

TRUE ITEM PARAMETERS

A.1. True Item Parameters (32 Items with 4 Common Items)

	FormX					FormY				
	a1	a2	a3	d	c	a1	a2	a3	d	c
1	1.55	0	0	1.25	0.15	1.69	0	0	1.56	0.12
2	1.86	0	0	1.46	0.2	1.39	0	0	2.82	0.01
3	1.07	0	0	1.68	0.03	0.9	0	0	1.05	0.01
4	1.13	0	0	1.77	0.03	1.55	0	0	2.5	0.08
5	1.87	0	0	2.41	0.14	1.13	0	0	1.63	0.02
6	1.36	0	0	1.07	0.14	2.39	0	0	0.25	0.2
7	1.03	0	0	0.87	0.12	0.95	0	0	0.52	0.16
8	1.65	0	0	2.25	0.19	1.54	0	0	3.51	0.02
9	1.32	0	0	0.72	0.18	1.56	0	0	1.71	0.2
10	0.96	0	0	1.2	0	1.7	0	0	-0.29	0.17
11	1.2	0	0	1.65	0.08	1.16	0	0	0.73	0.2
12	1.61	0	0	3.2	0.03	0.88	0	0	1.28	0.02
13	1.25	0	0	2.61	0.01	0.82	0	0	1.29	0.02
14	1.08	0	0	0.7	0.22	0.91	0	0	1.89	0
15	1.53	0	0	1.11	0.16	1.04	0	0	1.01	0.1
16	1.72	0	0	1.37	0.22	1.06	0	0	2.26	0.02
17	1.64	0	0	2.9	0.21	1.11	0	0	2.01	0.15
18	1.2	0	0	0.99	0.22	1.13	0	0	1.38	0.05
19	2	0	0	1.79	0.13	1.29	0	0	2.35	0.09
20	1.57	0	0	1.47	0.2	1.02	0	0	1.73	0.05
21	2.03	0	0	1.78	0.09	0.86	0	0	0.7	0.07
22	1.72	0	0	0.54	0.2	1.07	0	0	2.07	0.01
23	1.61	0	0	1.17	0.18	1.06	0	0	1.68	0.02
24	1.02	0	0	1.68	0	1.69	0	0	1.33	0.12
25	1.84	0	0	3.06	0.11	1.04	0	0	1.47	0.09
26	0.88	0	0	2.8	0.02	1.12	0	0	1.32	0.03
27	1.7	0	0	0.71	0.2	1.3	0	0	1.88	0.25
28	1.48	0	0	0.43	0.12	1.68	0	0	1.78	0.12
29	1.24	0	0	2.09	0.05	1.24	0	0	2.09	0.05
30	1.25	0	0	0.6	0.21	1.25	0	0	0.6	0.21
31	1.3	0	0	1.3	0.11	1.3	0	0	1.3	0.11
32	1.32	0	0	1.8	0.09	1.32	0	0	1.8	0.09
33	0	1.19	0	2.04	0.01	0	0.8	0	2.54	0.02
34	0	2	0	0.67	0.17	0	1.1	0	1.78	0.11
35	0	1.38	0	1.29	0.1	0	1.11	0	0.71	0.18
36	0	1.08	0	1.14	0.07	0	1.93	0	2.8	0.12
37	0	0.9	0	-0.48	0.21	0	1.2	0	1.66	0.11
38	0	1.11	0	0.65	0.19	0	1.79	0	-0.32	0.14
39	0	1.48	0	0.27	0.23	0	1.17	0	1.16	0.07
40	0	0.92	0	1.73	0	0	1.18	0	1.11	0.23
41	0	1.3	0	0.33	0.21	0	1.62	0	2.06	0.11
42	0	0.93	0	1.5	0.01	0	1.67	0	0.31	0.17
43	0	1.58	0	2.11	0.28	0	1.64	0	3.98	0.03
44	0	1.24	0	0.54	0.2	0	0.99	0	-0.04	0.23
45	0	0.99	0	1.19	0.03	0	0.9	0	3.07	0.02
46	0	1.2	0	2.19	0	0	1.92	0	2.46	0.2
47	0	1.43	0	2.99	0.07	0	1.29	0	0.78	0.21

	FormX					FormY				
	a1	a2	a3	d	c	a1	a2	a3	d	c
48	0	1.83	0	0.64	0.4	0	1.55	0	0.16	0.29
49	0	0.61	0	1.17	0.01	0	1.92	0	3.06	0.16
50	0	1.4	0	2.6	0.01	0	2.17	0	-2.98	0.13
51	0	1.39	0	0.12	0.3	0	1.75	0	0.06	0.22
52	0	1.03	0	0.72	0.16	0	1.95	0	-1.84	0.3
53	0	1.31	0	0.64	0.15	0	1.61	0	-0.08	0.25
54	0	1.09	0	0.32	0.15	0	1.4	0	1.56	0.05
55	0	1.46	0	2.3	0.02	0	1.55	0	1.82	0.05
56	0	2.18	0	-0.66	0.23	0	1.33	0	-0.09	0.11
57	0	1.1	0	1.87	0.09	0	1.03	0	1.32	0.02
58	0	1.01	0	1.41	0.2	0	1.87	0	1.77	0.19
59	0	1.64	0	-0.25	0.3	0	1.61	0	-0.12	0.16
60	0	1.61	0	0.72	0.24	0	1.11	0	1	0.11
61	0	1.23	0	0.96	0.16	0	1.23	0	0.96	0.16
62	0	1.61	0	0.09	0.15	0	1.61	0	0.09	0.15
63	0	1.44	0	-0.46	0.19	0	1.44	0	-0.46	0.19
64	0	1.42	0	2.07	0.07	0	1.42	0	2.07	0.07
65	0	0	1.69	1.69	0.12	0	0	1.83	2.37	0.27
66	0	0	0.96	1.1	0	0	0	0.76	1.07	0.01
67	0	0	2.19	0.45	0.14	0	0	1.24	1.78	0.01
68	0	0	1.08	1.86	0.09	0	0	2.2	2.18	0.15
69	0	0	1.37	0.11	0.24	0	0	1.64	2.12	0.03
70	0	0	1.84	0.7	0.06	0	0	1.08	1.16	0.06
71	0	0	1.11	1.3	0.04	0	0	1.35	1.04	0.38
72	0	0	1.32	1.1	0.14	0	0	0.94	2.11	0.01
73	0	0	1.92	-0.9	0.12	0	0	1.11	0.17	0.33
74	0	0	1.68	-0.68	0.2	0	0	1.94	0.26	0.3
75	0	0	1.47	0.72	0.21	0	0	1.96	0.15	0.33
76	0	0	1.45	-0.42	0.25	0	0	1.04	0.89	0.04
77	0	0	1.12	2.03	0	0	0	1.17	1.86	0.01
78	0	0	1.79	1.1	0.14	0	0	1.83	0.78	0.35
79	0	0	1.89	1.48	0	0	0	1.99	0.4	0.09
80	0	0	1.18	0.83	0.04	0	0	2.24	-0.4	0.16
81	0	0	1.39	0.67	0.32	0	0	1.81	1.77	0.13
82	0	0	2.06	-1.4	0.18	0	0	1.93	0.6	0.1
83	0	0	1.29	-0.39	0.15	0	0	2.43	-1.78	0.29
84	0	0	1.32	1.09	0.14	0	0	1.23	0.32	0.31
85	0	0	1.92	1.12	0.1	0	0	0.93	0.37	0.08
86	0	0	2.12	1.6	0.2	0	0	2.06	0.75	0.05
87	0	0	1.87	1.85	0.08	0	0	2.08	-0.38	0.2
88	0	0	1.99	0.76	0.08	0	0	2.11	1.55	0.02
89	0	0	2.06	-0.17	0.2	0	0	1.09	0.6	0
90	0	0	1.42	1.59	0.14	0	0	1.51	1.38	0.03
91	0	0	2.14	0.98	0.17	0	0	1.54	0.82	0.2
92	0	0	1.66	1.1	0.2	0	0	2.19	0.45	0.14
93	0	0	1.28	2.01	0.05	0	0	1.28	2.01	0.05
94	0	0	1.81	0.55	0.28	0	0	1.81	0.55	0.28
95	0	0	1.32	1.1	0.14	0	0	1.32	1.1	0.14
96	0	0	1.74	0.16	0.4	0	0	1.74	0.16	0.4

* Note: Highlighted cells indicate common items between FormX and FormY

A.2. True Item Parameters (32 Items with 8 Common Items)

	FormX					FormY				
	a1	a2	a3	d	c	a1	a2	a3	d	c
1	1.55	0	0	1.25	0.15	1.69	0	0	1.56	0.12
2	1.86	0	0	1.46	0.2	1.39	0	0	2.82	0.01
3	1.07	0	0	1.68	0.03	0.9	0	0	1.05	0.01
4	1.13	0	0	1.77	0.03	1.55	0	0	2.5	0.08
5	1.87	0	0	2.41	0.14	1.13	0	0	1.63	0.02
6	1.36	0	0	1.07	0.14	2.39	0	0	0.25	0.2
7	1.03	0	0	0.87	0.12	0.95	0	0	0.52	0.16
8	1.65	0	0	2.25	0.19	1.54	0	0	3.51	0.02
9	1.32	0	0	0.72	0.18	1.56	0	0	1.71	0.2
10	0.96	0	0	1.2	0	1.7	0	0	-0.29	0.17
11	1.2	0	0	1.65	0.08	1.16	0	0	0.73	0.2
12	1.61	0	0	3.2	0.03	0.88	0	0	1.28	0.02
13	1.25	0	0	2.61	0.01	0.82	0	0	1.29	0.02
14	1.08	0	0	0.7	0.22	0.91	0	0	1.89	0
15	1.53	0	0	1.11	0.16	1.04	0	0	1.01	0.1
16	1.72	0	0	1.37	0.22	1.06	0	0	2.26	0.02
17	1.64	0	0	2.9	0.21	1.11	0	0	2.01	0.15
18	1.2	0	0	0.99	0.22	1.13	0	0	1.38	0.05
19	2	0	0	1.79	0.13	1.29	0	0	2.35	0.09
20	1.57	0	0	1.47	0.2	1.02	0	0	1.73	0.05
21	2.03	0	0	1.78	0.09	0.86	0	0	0.7	0.07
22	1.72	0	0	0.54	0.2	1.07	0	0	2.07	0.01
23	1.61	0	0	1.17	0.18	1.06	0	0	1.68	0.02
24	1.02	0	0	1.68	0	1.69	0	0	1.33	0.12
25	1.04	0	0	1.47	0.09	1.04	0	0	1.47	0.09
26	1.12	0	0	1.32	0.03	1.12	0	0	1.32	0.03
27	1.3	0	0	1.88	0.25	1.3	0	0	1.88	0.25
28	1.68	0	0	1.78	0.12	1.68	0	0	1.78	0.12
29	1.24	0	0	2.09	0.05	1.24	0	0	2.09	0.05
30	1.25	0	0	0.6	0.21	1.25	0	0	0.6	0.21
31	1.3	0	0	1.3	0.11	1.3	0	0	1.3	0.11
32	1.32	0	0	1.8	0.09	1.32	0	0	1.8	0.09
33	0	1.19	0	2.04	0.01	0	0.8	0	2.54	0.02
34	0	2	0	0.67	0.17	0	1.1	0	1.78	0.11
35	0	1.38	0	1.29	0.1	0	1.11	0	0.71	0.18
36	0	1.08	0	1.14	0.07	0	1.93	0	2.8	0.12
37	0	0.9	0	-0.48	0.21	0	1.2	0	1.66	0.11
38	0	1.11	0	0.65	0.19	0	1.79	0	-0.32	0.14
39	0	1.48	0	0.27	0.23	0	1.17	0	1.16	0.07
40	0	0.92	0	1.73	0	0	1.18	0	1.11	0.23
41	0	1.3	0	0.33	0.21	0	1.62	0	2.06	0.11
42	0	0.93	0	1.5	0.01	0	1.67	0	0.31	0.17
43	0	1.58	0	2.11	0.28	0	1.64	0	3.98	0.03
44	0	1.24	0	0.54	0.2	0	0.99	0	-0.04	0.23
45	0	0.99	0	1.19	0.03	0	0.9	0	3.07	0.02
46	0	1.2	0	2.19	0	0	1.92	0	2.46	0.2
47	0	1.43	0	2.99	0.07	0	1.29	0	0.78	0.21
48	0	1.83	0	0.64	0.4	0	1.55	0	0.16	0.29
49	0	0.61	0	1.17	0.01	0	1.92	0	3.06	0.16
50	0	1.4	0	2.6	0.01	0	2.17	0	-2.98	0.13
51	0	1.39	0	0.12	0.3	0	1.75	0	0.06	0.22
52	0	1.03	0	0.72	0.16	0	1.95	0	-1.84	0.3

	FormX					FormY				
	a1	a2	a3	d	c	a1	a2	a3	d	c
53	0	1.31	0	0.64	0.15	0	1.61	0	-0.08	0.25
54	0	1.09	0	0.32	0.15	0	1.4	0	1.56	0.05
55	0	1.46	0	2.3	0.02	0	1.55	0	1.82	0.05
56	0	2.18	0	-0.66	0.23	0	1.33	0	-0.09	0.11
57	0	1.03	0	1.32	0.02	0	1.03	0	1.32	0.02
58	0	1.87	0	1.77	0.19	0	1.87	0	1.77	0.19
59	0	1.61	0	-0.12	0.16	0	1.61	0	-0.12	0.16
60	0	1.11	0	1	0.11	0	1.11	0	1	0.11
61	0	1.23	0	0.96	0.16	0	1.23	0	0.96	0.16
62	0	1.61	0	0.09	0.15	0	1.61	0	0.09	0.15
63	0	1.44	0	-0.46	0.19	0	1.44	0	-0.46	0.19
64	0	1.42	0	2.07	0.07	0	1.42	0	2.07	0.07
65	0	0	1.69	1.69	0.12	0	0	1.83	2.37	0.27
66	0	0	0.96	1.1	0	0	0	0.76	1.07	0.01
67	0	0	2.19	0.45	0.14	0	0	1.24	1.78	0.01
68	0	0	1.08	1.86	0.09	0	0	2.2	2.18	0.15
69	0	0	1.37	0.11	0.24	0	0	1.64	2.12	0.03
70	0	0	1.84	0.7	0.06	0	0	1.08	1.16	0.06
71	0	0	1.11	1.3	0.04	0	0	1.35	1.04	0.38
72	0	0	1.32	1.1	0.14	0	0	0.94	2.11	0.01
73	0	0	1.92	-0.9	0.12	0	0	1.11	0.17	0.33
74	0	0	1.68	-0.68	0.2	0	0	1.94	0.26	0.3
75	0	0	1.47	0.72	0.21	0	0	1.96	0.15	0.33
76	0	0	1.45	-0.42	0.25	0	0	1.04	0.89	0.04
77	0	0	1.12	2.03	0	0	0	1.17	1.86	0.01
78	0	0	1.79	1.1	0.14	0	0	1.83	0.78	0.35
79	0	0	1.89	1.48	0	0	0	1.99	0.4	0.09
80	0	0	1.18	0.83	0.04	0	0	2.24	-0.4	0.16
81	0	0	1.39	0.67	0.32	0	0	1.81	1.77	0.13
82	0	0	2.06	-1.4	0.18	0	0	1.93	0.6	0.1
83	0	0	1.29	-0.39	0.15	0	0	2.43	-1.78	0.29
84	0	0	1.32	1.09	0.14	0	0	1.23	0.32	0.31
85	0	0	1.92	1.12	0.1	0	0	0.93	0.37	0.08
86	0	0	2.12	1.6	0.2	0	0	2.06	0.75	0.05
87	0	0	1.87	1.85	0.08	0	0	2.08	-0.38	0.2
88	0	0	1.99	0.76	0.08	0	0	2.11	1.55	0.02
89	0	0	1.09	0.6	0	0	0	1.09	0.6	0
90	0	0	1.51	1.38	0.03	0	0	1.51	1.38	0.03
91	0	0	1.54	0.82	0.2	0	0	1.54	0.82	0.2
92	0	0	2.19	0.45	0.14	0	0	2.19	0.45	0.14
93	0	0	1.28	2.01	0.05	0	0	1.28	2.01	0.05
94	0	0	1.81	0.55	0.28	0	0	1.81	0.55	0.28
95	0	0	1.32	1.1	0.14	0	0	1.32	1.1	0.14
96	0	0	1.74	0.16	0.4	0	0	1.74	0.16	0.4

* Note: Highlighted cells indicate common items between FormX and FormY

A.3. True Item Parameters (16 Items with 4 Common Items)

	FormX					FormY				
	a1	a2	a3	d	c	a1	a2	a3	d	c
1	1.55	0	0	1.25	0.15	1.69	0	0	1.56	0.12
2	1.86	0	0	1.46	0.2	1.39	0	0	2.82	0.01
3	1.13	0	0	1.77	0.03	1.13	0	0	1.63	0.02
4	1.03	0	0	0.87	0.12	0.95	0	0	0.52	0.16
5	1.32	0	0	0.72	0.18	1.56	0	0	1.71	0.2
6	1.2	0	0	1.65	0.08	1.16	0	0	0.73	0.2
7	1.25	0	0	2.61	0.01	0.91	0	0	1.89	0
8	1.53	0	0	1.11	0.16	1.04	0	0	1.01	0.1
9	1.64	0	0	2.9	0.21	1.11	0	0	2.01	0.15
10	2	0	0	1.79	0.13	1.29	0	0	2.35	0.09
11	2.03	0	0	1.78	0.09	0.86	0	0	0.7	0.07
12	1.61	0	0	1.17	0.18	1.06	0	0	1.68	0.02
13	1.24	0	0	2.09	0.05	1.24	0	0	2.09	0.05
14	1.25	0	0	0.6	0.21	1.25	0	0	0.6	0.21
15	1.3	0	0	1.3	0.11	1.3	0	0	1.3	0.11
16	1.32	0	0	1.8	0.09	1.32	0	0	1.8	0.09
17	0	1.19	0	2.04	0.01	0	1.1	0	1.78	0.11
18	0	1.38	0	1.29	0.1	0	1.11	0	0.71	0.18
19	0	1.11	0	0.65	0.19	0	1.79	0	-0.32	0.14
20	0	1.48	0	0.27	0.23	0	1.17	0	1.16	0.07
21	0	1.3	0	0.33	0.21	0	1.62	0	2.06	0.11
22	0	1.58	0	2.11	0.28	0	1.67	0	0.31	0.17
23	0	0.99	0	1.19	0.03	0	0.9	0	3.07	0.02
24	0	1.43	0	2.99	0.07	0	1.29	0	0.78	0.21
25	0	0.61	0	1.17	0.01	0	1.92	0	3.06	0.16
26	0	1.03	0	0.72	0.16	0	1.75	0	0.06	0.22
27	0	1.31	0	0.64	0.15	0	1.61	0	-0.08	0.25
28	0	2.18	0	-0.66	0.23	0	1.33	0	-0.09	0.11
29	0	1.23	0	0.96	0.16	0	1.23	0	0.96	0.16
30	0	1.61	0	0.09	0.15	0	1.61	0	0.09	0.15
31	0	1.44	0	-0.46	0.19	0	1.44	0	-0.46	0.19
32	0	1.42	0	2.07	0.07	0	1.42	0	2.07	0.07
33	0	0	1.69	1.69	0.12	0	0	1.83	2.37	0.27
34	0	0	2.19	0.45	0.14	0	0	2.2	2.18	0.15
35	0	0	1.37	0.11	0.24	0	0	1.64	2.12	0.03
36	0	0	1.11	1.3	0.04	0	0	1.08	1.16	0.06
37	0	0	1.92	-0.9	0.12	0	0	1.11	0.17	0.33
38	0	0	1.47	0.72	0.21	0	0	1.94	0.26	0.3
39	0	0	1.12	2.03	0	0	0	1.17	1.86	0.01
40	0	0	1.89	1.48	0	0	0	1.99	0.4	0.09
41	0	0	1.39	0.67	0.32	0	0	1.81	1.77	0.13
42	0	0	1.29	-0.39	0.15	0	0	2.43	-1.78	0.29
43	0	0	1.92	1.12	0.1	0	0	0.93	0.37	0.08
44	0	0	1.99	0.76	0.08	0	0	2.08	-0.38	0.2
45	0	0	1.28	2.01	0.05	0	0	1.28	2.01	0.05
46	0	0	1.81	0.55	0.28	0	0	1.81	0.55	0.28
47	0	0	1.32	1.1	0.14	0	0	1.32	1.1	0.14
48	0	0	1.74	0.16	0.4	0	0	1.74	0.16	0.4

* Note: Highlighted cells indicate common items between FormX and FormY

APPENDIX B

STIMULUS COORDINATES

B.1. Stimulus Coordinates (96 Items)

		ρ =0.9			ρ =0.8			ρ =0.6			ρ =0.4		
		Dim1	Dim2	Dim3	Dim1	Dim2	Dim3	Dim1	Dim2	Dim3	Dim1	Dim2	Dim3
S1	I1	0.387	0.2391	-1.7089	-0.6402	-0.7028	0.9086	0.8605	-1.3543	-0.6034	1.0983	1.2957	-0.1477
	I2	1.2415	0.3965	-1.1599	-1.1128	-0.2199	0.6491	1.0811	-0.9543	0.1643	1.1431	0.9996	-0.4104
	I3	2.0326	-0.1886	0.9857	-1.089	-1.5827	0.0943	1.1508	-1.0822	1.6409	0.928	1.4657	0.2152
	I4	1.527	-1.44	-0.0095	-0.7759	-1.386	1.5497	0.7558	-0.6205	-1.7959	0.5608	1.546	0.6261
	I5	1.4825	-0.6502	-0.9912	0.878	-0.4169	1.5002	1.3743	-0.9264	-0.1128	0.9796	1.0812	0.2665
	I6	0.3064	-1.5407	-0.63	-1.1005	-0.6156	1.0855	0.7269	-1.657	-0.8391	0.9496	1.3696	-0.4747
	I7	2.1297	0.995	0.7986	-0.1567	-1.5632	0.8779	0.8325	-1.2995	-1.3147	1.1871	1.3283	-0.6913
	I8	-0.0072	-1.254	-1.2267	-1.3702	-1.3306	0.2886	1.2793	-1.1609	0.743	1.1803	1.0161	-0.3595
	I9	1.2153	0.531	-1.072	-1.3645	-1.3606	-0.0876	0.8217	-1.1171	0.8294	1.1605	1.104	0.0148
	I10	2.0201	0.1541	-1.0602	-1.3044	0.3014	1.1643	1.5347	-0.4507	0.1779	1.0984	1.0743	-1.0393
	I11	1.0939	0.7606	-1.394	-0.3887	-0.62	1.7186	0.7995	-1.5699	0.6842	0.9342	1.3681	-0.3369
	I12	1.5536	-1.0068	0.6559	-0.3608	0.2108	1.8251	0.8657	-1.3797	0.5682	1.2555	0.9379	-0.9288
	I13	-0.2348	-0.3604	-1.504	0.4167	-1.337	1.6095	0.7387	-1.455	0.8676	0.996	0.8352	-0.9932
	I14	1.7286	-1.0244	-0.8405	-1.8301	-0.4726	-0.6318	0.5965	-0.8912	-1.5052	0.712	1.0145	1.2978
	I15	0.2369	-1.3183	-0.9828	-0.6472	-1.1205	1.203	0.676	-1.4203	0.2522	1.1139	1.1795	-0.3391
	I16	0.3748	0.3078	-1.3285	-0.8082	-0.4655	1.2616	1.2377	-1.1576	-0.4281	0.9653	1.2257	-0.0291
	I17	-0.9065	-1.7529	-0.1688	-1.8942	-0.7307	0.1147	1.2664	-1.0216	-0.8238	0.6694	1.5547	0.2886
	I18	1.2131	-0.8995	-1.0279	-0.2958	0.554	1.8534	1.3369	-1.3759	0.2746	1.2938	1.2934	-0.2904
	I19	0.82	-0.3188	0.1133	-0.8466	-0.548	0.6998	0.6643	-0.8505	-0.2691	0.9601	1.2457	0.1347
	I20	-0.2706	0.6731	-1.5722	-0.4077	-0.6936	1.3817	1.015	-1.2743	0.3524	0.8879	1.3473	-0.131
	I21	0.6534	-0.8354	0.2223	-0.8463	-0.3117	0.0163	0.5722	-1.0122	-0.1078	1.0132	1.2192	-0.0861
	I22	1.0007	0.1078	0.2057	-0.9579	-0.6428	0.7394	1.0912	-0.8685	-0.3758	1.0904	1.4239	-0.1711
	I23	0.6403	0.2624	-1.0969	-0.8393	-0.4008	1.2414	0.9304	-1.056	0.2739	1.2058	1.1285	-0.2675
	I24	-0.0989	-2.0998	-0.407	-0.9484	0.637	1.8156	1.4419	-0.6394	-1.162	0.5614	1.3459	1.3769
	I25	0.6957	-1.6624	-0.2322	0.7698	-1.5459	1.3539	1.116	-1.0002	0.9811	1.1499	1.1555	-0.3016
	I26	-2.5132	-2.4177	0.4235	-1.7644	1.118	0.682	0.7897	-1.2381	-2.0594	0.4571	0.3331	2.3564
	I27	0.7001	-0.8239	-0.483	-1.1984	-0.7103	0.1105	0.7142	-0.7397	0.1629	1.0229	1.1626	-0.1955
	I28	0.2141	-0.7501	-0.843	-0.3725	-0.271	1.2793	0.9314	-1.1254	0.625	1.261	0.9412	0.5002

	I29	-0.6916	-1.9501	0.0326	-0.8315	0.2702	1.4622	1.2795	-1.5645	-0.4172	1.3667	1.0204	-0.1362
	I30	0.4498	-1.4826	1.2392	-0.9409	-1.9052	-0.5306	0.9823	-1.8736	0.6018	1.3527	1.2155	0.0404
	I31	0.9878	-0.6637	1.3403	0.1721	-0.8179	1.489	0.9191	-1.1168	-0.842	0.6621	1.0213	-1.1654
	I32	0.8047	-1.4522	-0.945	-1.1026	-1.8406	-0.434	1.1952	-1.1376	-1.1197	1.3513	1.281	0.1337
S2	I33	-1.8699	-0.5648	-0.2515	0.6539	1.6602	0.8526	0.7294	0.8447	-1.5142	0.5555	-1.319	1.4159
	I34	0.4876	0.333	-0.242	-0.3897	0.8487	-0.3564	0.2456	0.8893	0.1158	0.5476	-1.4473	-0.129
	I35	-0.5445	-0.8829	1.1366	-1.2097	0.7982	-0.6444	0.3526	1.4162	0.5186	0.7635	-1.4107	-0.1256
	I36	-0.1371	0.2467	1.6942	1.8543	-0.1408	0.5548	0.1388	1.2678	-0.8609	0.7612	-1.3723	-0.5926
	I37	2.1277	1.1289	0.1565	-1.9094	0.1684	-1.494	0.5701	1.1291	1.8048	0.4637	-0.8016	1.9707
	I38	-1.4139	-0.5972	-1.17	0.0565	1.6718	1.1071	0.6602	1.226	-1.2926	0.66	-1.5836	0.7432
	I39	0.6119	-0.4667	1.2779	-1.6096	0.7835	-0.4655	0.8543	0.9448	0.2294	0.4817	-1.5276	0.0373
	I40	0.0614	1.9424	-1.5794	-1.7321	0.4605	-1.7343	0.6613	1.5096	1.2651	0.2767	-1.1455	-1.6036
	I41	0.1472	-0.3325	1.0535	0.5424	1.6442	0.4175	0.2142	1.7281	-0.4593	0.5949	-1.4887	-0.7599
	I42	0.5565	-0.1483	1.5401	0.5696	1.3394	1.0451	0.4915	1.3689	-1.5316	0.708	-1.3651	1.1299
	I43	1.5217	0.9412	0.2828	-1.5137	0.9008	0.0601	0.8827	1.4305	-0.6949	0.5151	-1.1432	-1.1446
	I44	0.8407	1.8241	-0.2976	-0.7339	1.6238	-0.4446	0.4514	1.5441	0.7511	0.7343	-1.2602	0.9099
	I45	-1.2534	0.1073	-1.6701	-0.9225	0.8147	-1.0713	0.7391	1.6937	0.1246	0.5867	-1.5367	-0.7046
	I46	0.2431	-0.6875	-2.1065	-1.2382	0.281	-1.6985	0.1892	1.6666	-1.4086	0.3949	-1.3176	-1.4382
	I47	0.6888	-0.4279	1.597	0.2558	2.0466	0.2031	-0.3365	2.0169	-0.699	0.5906	-1.7431	0.5393
	I48	1.652	1.0375	-0.0643	0.2623	0.9812	1.1551	1.0235	0.6509	1.3343	0.5351	-1.7	0.2212
	I49	-0.029	-1.8336	1.7745	1.372	1.0172	1.5811	0.6114	0.9071	2.3171	0.3038	-0.2963	2.4109
	I50	0.6405	-1.1964	1.6208	-0.1224	1.862	-0.7809	0.4555	1.205	1.5891	0.1032	-1.0324	1.5225
	I51	-0.9489	-0.5673	-1.3209	1.1547	1.3162	0.8658	0.851	1.1118	0.8311	0.565	-1.5041	-0.5156
	I52	-2.2527	0.2449	-1.0749	-0.5025	1.9074	-0.338	0.5787	1.7716	0.5932	0.5449	-1.478	-0.8347
	I53	-0.4278	-1.1162	0.5724	0.9775	1.5859	-0.5055	0.2661	1.1642	-0.5985	0.4847	-1.4984	-0.5733
	I54	0.1439	-0.0748	1.9247	0.6916	1.7831	0.5691	-0.1426	1.4438	0.1311	0.419	-1.3868	-0.7903
	I55	1.2199	-0.0569	1.1484	-0.7152	0.7882	-1.6847	0.097	1.6674	-0.291	0.9363	-1.4664	0.643
	I56	-0.8467	-0.4454	-0.427	-0.3878	0.7191	-1.118	0.4785	1.1362	-0.448	0.8113	-1.5645	-0.0571
	I57	-1.1641	-1.24	-0.4643	-0.9971	1.4869	-0.7014	0.3923	1.1656	-1.7591	0.6452	-1.4502	-1.3344
	I58	0.5004	1.5319	1.5201	1.0344	1.3002	1.1462	1.2605	1.2721	1.2605	0.4776	-1.4506	1.2446
	I59	1.1566	0.7321	0.6843	-1.0535	-0.2988	-1.4325	0.4879	1.2507	0.5735	0.6854	-1.632	0.0323
	I60	-0.6158	-0.7741	-0.7265	0.4107	0.9927	0.0423	0.0515	1.2578	-0.449	0.5487	-1.3061	-0.6417
	I61	0.855	0.5518	-0.7754	0.044	1.7008	-0.4707	0.5089	1.393	-0.347	0.546	-1.3439	-0.6608
	I62	0.245	0.0906	0.9547	-0.6543	1.051	-0.4842	0.7575	1.023	0.0945	0.2611	-1.6482	0.1172
	I63	0.459	1.6009	0.1201	-0.3204	1.5759	-0.6586	0.9378	1.5808	-0.0417	0.4734	-1.5675	0.3279
	I64	-0.8081	-1.0452	-1.1922	-0.1458	1.7298	-0.8964	0.5213	1.0722	-0.3676	0.8559	-1.4656	0.5192
S3	I65	-1.0305	-0.2472	1.0553	1.0367	-0.859	-0.8602	-1.4352	0.1765	-0.2246	-1.3418	0.0716	-0.6718

I66	-0.1996	2.5984	-0.2272	1.6006	-0.9767	0.5138	-1.7253	-0.6548	1.0205	-1.7249	0.2731	-0.1282
I67	-0.1329	0.3303	-0.085	0.3788	-0.2165	-0.4407	-1.0341	-0.0724	0.359	-1.5645	0.1334	0.0982
I68	-1.197	-0.3832	1.9927	0.1393	-1.4826	-1.8362	-1.609	-0.4833	1.2898	-1.7502	0.1019	0.9519
I69	-0.5726	1.1382	0.7308	0.8634	-0.9042	-0.9297	-1.4689	-0.134	-1.0115	-1.7388	0.2339	0.4061
I70	-0.527	-0.015	0.2151	0.6664	-0.6966	-0.5445	-1.3671	-0.1572	0.1499	-1.5347	0.2923	0.0616
I71	-0.722	1.0928	-0.8958	1.6459	-0.7074	-0.4443	-1.9725	0.3962	-0.2592	-1.4718	0.3807	-0.1883
I72	-0.5826	1.4663	-0.3619	0.5693	-0.1118	-1.4282	-1.5336	-0.0298	0.0369	-1.5685	0.2345	-0.0458
I73	-0.5162	0.043	-0.1439	0.601	-0.1571	-0.6988	-1.3212	0.0558	0.0206	-1.4885	0.1375	-0.0417
I74	-0.2798	0.674	1.1245	0.8867	-0.3877	-1.2136	-1.5325	-0.2019	-0.0642	-1.6013	0.2248	-0.1495
I75	-1.2755	0.2425	0.3791	1.5666	-0.4107	-0.2549	-1.8187	-0.143	0.5505	-1.6293	0.1505	-0.1021
I76	-0.5887	0.3307	-1.2423	0.2591	-1.0108	-1.1731	-1.6918	-0.4398	0.8081	-1.6123	0.3427	0.1673
I77	0.3033	0.594	2.0615	1.5543	0.5483	0.4599	-1.5147	-0.5352	1.0304	-1.6051	0.1548	-0.8318
I78	-0.9137	-0.0603	0.0176	0.7269	-0.1675	-0.5315	-1.3691	0.0528	-0.0133	-1.5232	0.0084	-0.1665
I79	-0.1388	0.2565	0.3168	0.6228	-0.4148	-0.4244	-1.2637	-0.045	-0.0648	-1.4126	-0.0511	-0.1527
I80	-0.7028	0.465	1.0396	0.8625	-0.0892	-1.1981	-1.765	0.1849	0.0043	-1.6304	0.2397	0.049
I81	-1.2947	1.6022	-0.1469	1.9292	-0.0458	-0.2683	-1.1643	0.0616	-0.9872	-1.6741	0.2527	0.0953
I82	0.3374	1.7489	0.2977	1.1402	-0.3059	-0.6797	-1.4592	-0.152	-0.1933	-1.6804	0.2534	-0.2269
I83	-1.0323	1.3423	0.257	0.6881	-0.8831	-0.6355	-1.2035	-0.5559	0.4004	-1.7846	0.1696	-0.0927
I84	-1.0208	1.3194	0.3353	0.42	-0.5935	-1.2546	-1.4655	-0.4289	0.2436	-1.7233	0.363	0.1285
I85	-0.6852	0.4081	-0.1461	0.4386	-0.4011	-0.9695	-1.3098	0.0311	0.2717	-1.5286	0.3843	0.1576
I86	-0.463	0.5863	0.1061	1.1584	-0.184	-0.498	-1.4615	-0.2199	-0.1114	-1.6003	0.1825	0.2237
I87	-0.7759	0.7432	0.1147	1.3306	-0.5568	-0.3158	-1.1764	0.0782	-0.1454	-1.516	0.2758	0.1092
I88	-0.6696	0.7879	-0.1116	0.3503	-0.0911	-0.8401	-1.3175	-0.3266	-0.0843	-1.1675	0.2398	-0.1818
I89	-0.4978	0.6329	-0.4085	0.8386	-0.4609	-0.3527	-1.3596	-0.2687	-0.1567	-1.4995	0.1585	-0.2601
I90	-1.5702	0.455	-0.2786	0.4947	-0.971	-1.5863	-1.8342	-0.3561	-0.3708	-1.6097	0.3547	-0.3743
I91	-0.4406	0.4567	-0.0366	0.9176	-0.3594	-0.553	-1.3733	-0.1875	0.1431	-1.4702	-0.0959	0.1343
I92	-0.5565	-0.119	1.2138	1.0398	-0.4058	-0.419	-1.71	-0.3619	0.4745	-1.6	0.3406	0.303
I93	-1.569	1.4171	-0.0159	1.6362	-0.897	0.0099	-1.5806	-0.1781	0.8335	-1.7099	-0.1344	0.4173
I94	-0.4436	0.8103	-0.1208	1.1811	-0.6752	-0.7466	-1.4833	-0.1223	-0.2121	-1.5777	0.3571	0.0519
I95	-1.5741	0.3531	0.5805	1.4496	-0.8769	0.5716	-1.747	-0.1024	0.4235	-1.591	0.3311	0.3064
I96	-1.3011	0.5386	1.3541	1.2437	-0.1271	-1.369	-1.4898	-0.2391	-0.3559	-1.4685	0.3613	-0.6825

Note: S1 indicates subtest 1; S2 is subtest 2; and S3 refers to subtest 3. Dim 1 is dimension 1; Dim 2 is dimension 2; and Dim3 is dimension3. I1-I96 are items 1-96.

B.2. Stimulus Coordinates (48 Items)

		ρ =0.9			ρ =0.8			ρ =0.6			ρ =0.4		
		Dim1	Dim2	Dim3	Dim1	Dim2	Dim3	Dim1	Dim2	Dim3	Dim1	Dim2	Dim3
S1	I1	0.5482	-0.8794	-0.9652	0.367	-0.7435	-1.1539	0.646	-1.0983	1.3541	1.4305	0.7271	0.2149
	I2	-0.747	-1.3221	-0.4766	1.2931	-0.2172	-0.7143	1.1195	-0.8452	-0.4938	1.0316	0.7881	-0.2263
	I3	1.5627	1.8106	-0.2361	1.8221	0.4084	-0.157	0.2183	-1.0971	1.7521	1.5841	1.0828	0.2368
	I4	-2.2106	0.0819	-1.0947	-0.2641	-0.4356	-1.8191	0.6129	-1.147	-1.5516	1.4434	0.9439	-0.0045
	I5	0.2057	1.4718	-0.7875	0.6422	-0.3061	-1.4086	0.5987	-1.4404	-0.9355	1.493	0.9056	-0.407
	I6	1.7787	-0.1981	-1.1155	1.5586	0.2188	-1.7893	0.9393	-1.5384	0.1381	1.4216	0.8221	0.9912
	I7	0.6564	-1.7181	-2.0754	0.3076	-2.4558	0.7845	0.4949	-1.5625	-1.1016	1.5225	0.9428	-0.6641
	I8	0.4492	1.3724	-0.8044	1.0417	-0.6371	-0.9205	0.7124	-1.4625	0.1923	1.4569	0.8681	0.2486
	I9	0.1988	-0.7967	-1.2251	2.0624	-0.5704	-0.6953	1.1563	-1.3218	-0.6321	1.63	0.4669	0.3685
	I10	-0.3581	1.4248	-0.3557	0.3738	1.1609	-0.1614	1.0125	-1.1115	-0.2506	1.5709	0.394	0.6049
	I11	-0.4946	-0.1162	-0.6062	0.8977	0.148	-0.1531	0.3846	-0.8011	0.4065	1.2594	1.1148	-0.0813
	I12	0.6927	0.0676	-0.8098	0.0996	-0.5265	-1.4271	0.4656	-1.1967	1.2907	1.4342	0.6992	0.4397
	I13	-1.3119	1.5337	-1.286	0.6622	-2.1514	0.4626	1.1478	-1.203	-0.8171	1.276	0.9047	-0.9886
	I14	-1.0381	0.3236	-1.3927	1.3456	-0.1526	-0.9461	0.6485	-0.8386	1.4766	1.5672	0.5902	-0.2999
	I15	0.9909	0.407	-1.1768	1.2063	0.645	-1.3853	0.4582	-1.9471	0.1163	1.4603	0.7381	0.1092
	I16	1.5929	0.7725	-0.7223	1.5213	0.3655	-0.5605	0.8707	-1.55	-0.7071	1.1661	0.5775	-1.206
S2	I17	0.2301	-0.6547	-1.7665	-0.8673	-1.771	0.5647	0.5558	1.0716	1.3316	0.3074	-1.2076	1.3648
	I18	0.5278	-0.6501	0.8807	0.8161	0.9172	0.4678	0.8124	0.9784	0.6403	-0.1702	-1.3277	-0.1395
	I19	0.6985	1.053	1.6187	0.1574	1.778	1.18	0.5287	1.2539	1.4322	-0.1172	-1.8126	0.5888
	I20	1.3471	0.2574	-0.395	0.2611	-0.8563	0.875	0.5344	1.3005	0.0317	0.102	-1.6731	0.454
	I21	-1.0738	0.4524	-0.5817	-0.261	-1.7043	0.6025	1.2247	1.1342	-0.5102	-0.0898	-1.6519	0.3448
	I22	1.1396	-1.1684	-0.6882	0.5219	0.342	1.7362	1.274	1.3641	-0.2673	0.3414	-1.6786	-0.2075
	I23	-1.7139	1.0348	0.2582	0.5956	0.6521	1.6677	0.6195	1.4108	0.813	-0.031	-1.5087	-0.9504
	I24	-2.6658	-0.7622	0.5851	0.7647	-2.2573	0.2182	0.7755	0.8547	-0.8867	0.1877	-1.0195	1.4897
	I25	-0.8492	-2.9385	0.3443	0.2234	3.2191	-0.7114	0.5005	0.7525	-2.2573	-0.1995	-1.072	1.7689
	I26	1.8683	1.0119	0.532	-0.0239	0.267	1.8223	0.7708	1.4017	-1.2351	-0.1392	-1.6157	-1.3334
	I27	0.1456	-0.1909	1.1116	0.5251	0.0941	1.3774	1.2883	0.9666	0.5092	-0.215	-1.1946	-1.1777
	I28	0.3836	-0.7041	0.2125	-0.0381	0.4086	1.2302	0.8227	1.1326	0.4855	0.1555	-1.5608	0.2185
	I29	-0.4877	1.5034	-0.7475	0.3294	0.8893	1.4088	0.8804	0.8464	1.1502	-0.1296	-1.9397	0.0521
	I30	0.8208	-0.6878	0.167	-0.0242	-0.6117	0.7379	0.4677	1.2626	0.5592	0.1632	-1.2677	-0.6775
	I31	-1.0512	-0.3818	0.2926	0.7733	0.77	1.672	0.8277	1.3571	0.048	-0.2204	-1.5289	-0.9517
	I32	-0.9961	0.3062	-0.8535	-0.1809	0.6026	1.7312	0.8131	1.5176	-0.7556	0.1867	-1.2768	-1.2228
S3	I33	-0.4524	1.2534	1.1665	-1.1696	-0.0195	0.4964	-1.6128	0.13	0.3348	-1.4763	0.7119	0.6581

I34	-0.0748	-0.076	0.5506	-0.6035	0.3219	-0.0485	-1.2278	-0.0674	0.1745	-1.3035	0.8177	0.0811
I35	-0.255	-0.0933	1.602	-0.8914	0.4987	-1.0333	-1.6114	0.1916	0.8284	-1.2462	0.768	0.8406
I36	-1.1317	0.7542	1.3941	-1.4824	0.3306	-0.8824	-1.9069	0.2789	-0.4602	-1.5724	0.7402	-0.3955
I37	-0.036	-0.264	1.1562	-0.5901	0.4468	-0.2594	-1.2987	0.1038	-0.0041	-1.4439	0.4365	0.1277
I38	0.2226	-1.0454	0.7133	-1.2313	-0.585	0.1548	-1.551	0.7006	-0.0848	-1.5888	0.7293	-0.1298
I39	1.19	-1.6104	1.5345	-2.2742	-0.9184	-0.6172	-1.7298	0.027	-0.3602	-1.3808	0.7967	-1.0065
I40	0.1773	-0.2134	0.9647	-0.3773	0.0992	-0.1573	-1.3314	0.0071	-0.0182	-1.3266	0.6563	-0.3306
I41	-0.9712	0.4062	-0.0878	-1.4638	0.3482	-0.5459	-1.8213	-0.3293	0.5462	-1.5595	0.7617	-0.439
I42	-0.0323	0.544	1.7815	-1.4135	0.9923	-0.0459	-1.8129	0.0228	0.0105	-1.3999	1.055	0.3914
I43	-0.119	0.3051	0.6062	-0.8692	-0.1823	0.1489	-1.0322	-0.3996	-0.1882	-1.4119	0.5179	-0.4629
I44	-0.3644	-0.1451	0.4996	-0.8503	0.12	0.041	-1.1687	0.3884	0.0505	-1.3416	0.5033	0.0655
I45	1.5757	-0.2542	1.2996	-1.8607	0.4631	-0.5017	-1.1028	0.9643	-1.2157	-1.7087	0.2832	0.4549
I46	0.7287	0.8225	0.834	-1.3931	0.1857	0.1657	-1.5578	-0.2592	-0.3056	-1.4605	0.569	0.1276
I47	0.0925	-1.4863	-0.2953	-1.4338	-0.3896	-0.7505	-1.9394	-0.2499	-0.0996	-1.3282	0.8987	0.0102
I48	-1.3896	-0.6132	0.44	-0.6054	0.7984	-0.7008	-1.4776	0.0469	-0.5342	-1.3307	0.5248	1.0499

REFERENCES

Albano, A. D. (2011). *Equate: Statistical methods for test equating*. Retrieved from http://CRAN.R-project.org/package=equate.

Aldenderfer, M. S., & Blashfield, R. K. (1984). *Cluster analysis*. Beverley Hills, CA: Sage.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing.* Washington DC: AERA.

Angoff, W. H. (1984). *Scales, norms, and equivalent scores*. Princeton, NJ: Educational Testing Service.

Azzalinin, A., & Genton, M. G. (2008). Robust likelihood methods based on the skew-*t* and related distributions. *International Statistical review, 76*(1), 106-129.

Baker, F. B., & Al-Karni, A. (1991). A comparison of two procedures for computing IRT equating coefficients. *Journal of Educational Measurement, 28*(2), 147-162.

Béguin, A. A., & Hanson, B. A. (2001, April). *Effect of noncompensatory multidimensionality on separate and concurrent estimation in IRT observed score equating.* Paper presented at the annual meeting of the National Council on Measurement in Education, Seattle, WA.

Béguin, A. A., Hanson, B. A., & Glas, C. A. W. (2000, April). *Effect of multidimensionality on separate and concurrent estimation in IRT equating.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588-606.

Bock, R. D., Thissen, D., & Zimowski, M. F. (1997). IRT estimation of domain scores. *Journal of Educational Measurement*, *34*(3), 197-211.

Braun, H. I., & Holland, P. W. (1982). Observed-score test equating: A mathematical analysis of some ETS equating procedures. In P. W. Holland and D.B Rubin (Eds.), *Test equating* (pp. 9-49). New York: Academic.

Brossman, B. G. (2010). *Observed score and true score equating procedures for multidimensional item response theory.* Unpublished doctoral dissertation, University of Iowa, Iowa City, IA.

Budescu, D. (1985). Efficiency of linear equating as a function of the length of the anchor test. *Journal of Educational Measurement, 22*(1), 13-20.

Cai, L. (2012). *flexMIRT* [Computer program]. Seattle, WA: Vector Psychometric Group.

Carroll, J. D., Williams, B. A., & Levine, M. V. (2007). Multidimensional modeling with unidimensional approximations. *Journal of Mathematical Psychology, 51*(4), 207-228.

Chen, H., Cui, Z., Zhu, R., & Gao, X. (2010). *Evaluating the effects of differences in group abilities on the Tucker and the Levine observed-score methods for common-item nonequivalent groups equating.* (Research Report Series 2010-1). Iowa City, IA: American College Testing.

Davey, T., Oshima, T., & Lee, K. (1996). Linking multidimensional item calibrations. *Applied Psychological Measurement, 20*(4), 405-416.

Davison, M. L. (1985). Multidimensional scaling versus components analysis of test intercorrelations. *Psychological Bulletin*, *97*(1), 94-105.

Davison, M. L., & Sireci, S. G. (2000). Multidimensional scaling. In H.E.A Tinsley & S. Brown (Eds.), *Handbook of multivariate statistics and mathematical modeling* (pp. 325-349). Washington, DC: American Psychological Association.

Davison, M. L., & Skay, C. L. (1991). Multidimensional scaling and factor models of test and item responses. *Psychological Bulletin, 110*(3), 551-556.

Divgi, D. R. (1985). A minimum chi-square method for developing a common metric in item response theory. *Applied Psychological Measurement, 9*(4), 413-415.

Dorans, N. J., & Feigenbaum, M. D. (1994). Equating issues engendered by changes to the SAT and PSAT/NMSQT. In I. M. Lawrence, N. J. Dorans, M. D. Feigenbaum, N. J. Feryok, A. P. Schmitt, & N. K. Wright (Eds.), *Technical issues related to the introduction of the new SAT and PSAT/NMSQT* (ETS Research Memorandum No. RM-94-10). Princeton, NJ: Educational Testing Service.

Dwyer, A., Boughton, K. A., Yao, L., Steffen, M., & Lewis, D. (2006). *A comparison of subscale score augmentation methods using empirical data.* Paper presented at the meeting of the National Council on Measurement in Education, San Francisco, CA.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahweh, NJ: Lawrence Erlbaum Associates.

Fleishman, A. I. (1987). A method for simulating non-normal distributions. *Psychometrika, 43*, 521-532.

Gessaroli, M. E. (2004). *Using hierarchical multidimensional item response theory to estimate augmented subscores*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.

Haberman, S. J. (2008). When can subscores have value? *Journal of Educational and Behavioral Statistics, 33*(2), 204-229.

Haberman, S. J., Sinharay, S., & Puhan, G. (2009). Reporting subscores for institutions. *British Journal of Mathematical and Statistical Psychology*, *62*(1), 79-95.

Haebara, T. (1980). Equating logistic ability scales by a weighted least squares method. *Japanese Psychological Research, 22*(3), 144-149.

Haladyna, S. J., & Kramer, G. A. (2004). The validity of subscores for a credentialing test. *Evaluation and the Health Professions, 24*(7), 349-368.

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

Hanson, B. A., & Béguin, A. A. (2002). Obtaining a common scale for item response theory item parameters using separate versus concurrent estimation in the common-item equating design. *Applied Psychological Measurement, 26*(1), 3-24.

Hanson, B. A., & Zeng, L. (2004). PIE computer software. Iowa City, Iowa: CASMA.

Harris, D. J., & Kolen, M. J. (1990). A comparison of two equipercentile equating methods for common item equating. *Educational and Psychological Measurement, 50*, 61-71.

Hirsch, T. (1989). Multidimensional equating. *Journal of Educational Measurement, 26*(4), 337-349.

Holland, P. W., & Thayer, D. T. (1987). *Notes on the use of log-linear models for fitting discrete probability distributions* (Technical Report 87-79). Princeton, NJ: Educational Testing Service.

Holland, P. W., & Thayer, D. T. (2000). Univariate and bivariate loglinear models for discrete test score distributions. *Journal of Educational Behavioral Statistics, 25*(2), 133-183.

Holland, P. W., von Davier, A. A., Sinharay, S., & Han, N. (2006). *Testing the untestable assumptions of the chain and postratification equating methods for the NEAT design* (Research Report RR-6-17). Princeton, NJ: Educational Testing Service.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55.

Huang, P., Wu, Y., & Chen, Y. (1991). *IRT item parameter linking: Relevant issues for the purpose of item banking.* Paper presented at the International Academic Symposium on Psychological Measurement, Tainan, Taiwan.

Kang, T. (2006). *Model selection methods for unidimensional and multidimensional IRT models.* Unpublished doctoral dissertation, University of Wisconsin-Madison, Madison, WI.

Kaskowitz, G. S., & De Ayala, R. J. (2001). The effect of error in item parameter estimates on the test response function method of linking. *Applied Psychological Measurement, 25*(1), 39-52.

Kelley, T. L. (1927). *The interpretation of educational measurement*. New York: World Book.

Kim, S.-H., & Cohen, A. S. (1992). Effects of linking methods on detection of DIF. *Journal of Educational Measurement*, *29*(1), 51-66.

Kim, S.-H., & Cohen, A. S. (1998). A comparison of linking and concurrent calibration under item response theory. *Applied Psychological Measurement, 22*(2), 131-143.

Kim, S., & Lee, W. (2006). An extension of four IRT linking methods for mixed-format tests. *Journal of Educational Measurement*, *43*(1), 53-76.

Kolen, M. J. (1990). Does matching in equating work? A discussion. *Applied Measurement in Education, 3*(1), 97-104.

Kolen, M. J. (1985). Standard errors of Tucker equating. *Applied Psychological Measurement*, *9*(2), 209-223.

Kolen, M. J., & Brennan, R. L. (1995). *Test equating: Methods and practices*. New York, NY: Springer-Verlag.

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

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Beverly Hills, CA: Sage.

Kuder, G. F., & Richardson, M. W. (1937).The theory of the estimation of test reliability. *Psychometrika*, *2*(3), 151-160.

Lawrence, I., & Dorans, N. (1990). Effect on equating results of matching samples on an anchor test. *Applied Measurement in Education*, *3*(1), 19-36.

Li, Y. H., & Lissitz, R. W. (2000). An evaluation of the accuracy of multidimensional IRT linking. *Applied Psychological Measurement*, *24*(2), 115-38.

Livingston, S. A. (2004). *Equating test scores (without IRT).* Princeton, NJ: Educational Testing Service.

Livingston, S. A., Dorans, N. J., & Wright, N. K. (1990). What combination of sampling and equating methods works best? *Applied Measurement in Education, 3*(1), 73-95.

Lord, F. M., & Wingersky, M. S. (1984). Comparison of IRT true-score and equipercentile observed-score "equatings." *Applied Psychological Measurement, 8*(4), 452-461.

Loyd, B. H., & Hoover, H. D. (1980). Vertical equating using the Rasch model. *Journal of Educational Measurement, 17*(3), 179-193.

Lyren, P. (2009). Reporting subscores from college admission tests. *Practical Assessment, Research, and Evaluation, 14*(4), 1-10.

Macro, G. L. (1977). Item characteristic curve solutions to three intractable testing problems. *Journal of Educational Measurement, 14*(2), 139-160.

Macro, G. L., Petersen, N. S., & Stewart, E. E. (1983). A test of the adequacy of curvilinear score equating models. In D. Weiss (Ed.), *New horizons in testing* (pp. 147-176). New York: Academic.

Meng, Y. (2012). *Comparison of kernel equating and item response theory equating methods.* Unpublished doctoral dissertation, University of Massachusetts Amherst: Amherst, MA.

Milligan, G. W. (1996). Clustering validation: Results and implications for applied analyses. In P. Arabie, L.J. Hubert, & G. De Soete (eds.) *Clustering and classification,* River Edge, NJ: World Scientific.

Min, K.-S. (2003). *The Impact of Scale Dilation on the Quality of the Linking of Multidimensional Item Response Theory Calibrations*. Unpublished doctoral dissertation, Michigan State University, East Lansing, MI.

Mroch, A. A., & Bolt, D. M. (2006). A simulation comparison of parametric and nonparametric dimensionality detection procedures. *Applied Measurement in Education,19*(1), 67-91.

Muthén, L. K, & Muthén, B. O. (2007). *MPLUS* [Computer program]. Los Angeles, CA: Muthén & Muthén.

Ogasawara, H. (2002). Stable response functions with unstable item parameter estimates. *Applied Psychological Measurement, 25*(1), 53-67.

Oshima, T. C., Davey, T. C., & Lee, K. (2000). Multidimensional Linking: Four Practical Approaches. *Journal of Educational Measurement*, *37*(4), 357-73.

Per-Erick, L. & Vedman, J. (2013). *A comparison of methods for examining the added value of subscores.* Unpublished manuscript.

Petersen, N. S., Cook, L. L., & Stocking, M. L. (1983). IRT versus conventional equating methods: A comparative study of scale stability. *Journal of Educational Statistics, 8,* 137-156.

Pommerich, M., Nicewander, W. A., & Hanson, B. (1999). Estimating average domain scores. *Journal of Educational Measurement, 36,* 199-216.

Puhan, G., & Liang, L. (2011a). Equating subscores under the nonequivalent anchor test (NEAT) design. *Educational Measurement: Issues and Practice, 30*(1), 23-35.

Puhan, G., & Liang, L. (2011b). *Equating subscores using total scaled scores as an anchor.*(Research Report RR-11-07). Princeton, NJ: Educational Testing Service.

Puhan, G., Sinharay, S., Haberman, S. J., Larkin, K. (2010). The utility of augmented subscores in a licensure exam: an evaluation of methods using empirical data. *Applied Measurement in Education, 23*(3), 266-285.

R Development Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

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

Reckase, M. D., & McKinley, R. L. (1991). The discriminating power of items that measure more than one dimension. *Applied Psychological Measurement, 15*(4), 361-373

Revelle, W. (2013). *Psych*. R package version 1.3.10.12. Retrieved from http://personality-project.org/r.

Ricker, K. L., & von Davier, A. A. (2007). *The impact of anchor test length on equating results in a nonequivalent groups design* (Research Report RR-07-44). Princeton, NJ: Educational Testing Service.

Schlessinger, I. M., & Guttman, L. (1969). Smallest space analysis of intelligence and achievement tests. *Psychological Bulletin, 71*(2), 95-100.

Shepard, R. N. (1972). Introduction to volume I. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences* (Vol. 1, pp. 1-20). NewYork: Seminar Press.

Shin, D. (2007). A comparison of method of estimating subscale scores for mixed-format tests. *Pearson Educational Measurement.*

Skorupski, W. P., & Carvajal, J. (2010). A comparison of approaches for improving the reliability of objective level scores. *Educational and Psychological Measurement, 70*(3), 357-375.

Silver, N. C., & Dunlap, W. P. (1987). Averaging correlation coefficients: Should Fisher's z transformation be used? *Journal of Applied Psychology*, *72*(1), 146-148.

Simon, M. K. (2008). *Comparison of concurrent and separate multidimensional IRT linking of item parameters*. Unpublished doctoral dissertation, University of Minnesota, Minneapolis, MN.

Sinharay, S. (2010). How often do subscores have added value? Results from operational and simulated data. *Journal of Educational Measurement*, *47*(2), 150-174.

Sinharay, S., & Haberman, S. J. (2008). *Reporting subscores: A survey* (Research Report RM-08-18). Princeton, NJ: Educational Testing Service.

Sinharay, S., & Haberman, S. J. (2011a). *Equating of subscores and weighted averages under the NEAT design* (Research Report RR-11-01). Princeton, NJ: Educational Testing Service.

Sinharay, S., & Haberman, S. J. (2011b). Equating of augmented subscores. *Journal of Educational Measurement*, *48*(2), 122-145.

Sinharay, S., Haberman, S. J., & Puhan, G. (2007). Subscores based on classical test theory: To report or not to report. *Educational Measurement: Issues and Practice*, *26*(4), 21-28.

Sinharay, S., Puhan, G., & Haberman, S. J. (2011). An NCME instructional module on subscores. *Educational Measurement: Issues and Practice*, *30*(3), 29-40.

Skaggs, G., & Lissitz, R. W. (1986). An exploration of the robustness of four test equating models. *Applied Psychological Measurement, 10*(3), 303-317.

Stephens, C. N. (2012). *An investigation into the psychometric properties of the proportional reduction of mean squared error and augmented scores.* Unpublished doctoral dissertation, University of Iowa, Iowa City, IA.

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement,7*(2), 201-210.

Stone, C. A., Ye, F., Zhu, X., & Lane, S. (2010). Providing subscale scores for diagnostic information: A case study when the test is essentially unidimensional. *Applied Measurement in Education, 23*(1), 63-86.

Swediati, N. (1997). *Equating tests under the generalized partial credit model.* Unpublished doctoral dissertation, University of Massachusetts Amherst: Amherst, MA.

Tate, R. L. (2004). Implications of multidimensionality for total score and subscore performance. *Applied Measurement in Education, 17*(2). 89-112.

Wainer, H., Vevea, J. L., Camacho, F., Reeve, B. B., Rosa, K., & Nelson, L. (2001). Augmented scores—"borrowing strength" to compute scores based on small numbers of items. In D. Thissen & H. Wainer (Eds.), *Test scoring* (pp. 343-387). Hillsdale, NJ: Lawrenece Erlbaum Associates.

Wang, T., Lee, W., Brennan, R. L., & Kolen, M. J. (2008). A comparison of the frequency estimation and chained equipercentile methods under the common-item nonequivalent groups design. *Applied Psychological Measurement*, *32*(8), 632-651.

Way, W. D., & Tang, K. L. (1991). *A comparison of four logistic model equating methods.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Weeks, J. P. (2011). *Plink*. R package version 1.3-1. Retrieved from http://cran.r-project.org/web/packages/plink/.

Wei, Y. (2008). *A simulation study on the performance of four multidimensional IRT scale linking methods*. Unpublished doctoral dissertation, University of Florida, Gainesville, FL.

Wingersky, M. S., Cook, L. L., & Eignor, D. R. (1987). *Specifying the characteristics of linking items used for item response theory item calibration* (Research Report 87-24). Princeton, NJ: Educational Testing Service.

Wingersky, M. S., & Lord, F. M. (1984). An investigation of methods for reducing sampling error in certain IRT procedures. *Applied Psychological Measurement, 8*(3), 347-364.

Wuertz, D. (2012). fMultivar. Retrived from http://www.rmetrics.org.

Yao, L. (2011). Multidimensional linking for domain scores and overall scores for nonequivalent groups. *Applied Psychological Measurement, 35*(1), 48-66.

Yao, L., & Boughton, K. A. (2007). A multidimensional item response modeling approach for improving subscale proficiency estimation and classification. *Applied Psychological Measurement, 31*(2), 83-105.

Yao, L., & Boughton, K. A. (2009). Multidimensional linking for tests containing polytomous items. *Journal of Educational Measurement, 46*(2), 177-197.

Yen, W. M. (1987, June). *A Bayesian / IRT index of objective performance*. Paper presented at the annual meeting of the Psychometric Society. Montreal, Quebec, Canada.

Young, F.W., & Harris, D. F. (1993). Multidimensional scaling. In M.J. Noursis (Ed.). *SPSS for windows: Professional statistics* (computer manual, version 6.0) (pp.155-222). Chicago, IL: SPSS, Inc.

Zhang, J. (1996). *Some fundamental issues in item response theory with applications.* Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign, Urbana-Champaign, IL.

Zhang, J., & Stout, W. F. (1999). The theoretical DETECT index of dimensionality and its application to approximate simple structure. *Psychometrika*, *64*(2), 213-249.