This is to certify that the
dissertation entitled

USING MULTIDIMENSIONAL ITEM RESONSE THEORY TO EXAMINE MEASUREMENT EQUIVALENCE: A MONTE CARLO INVESTIGATION

presented by

Linda Baumunk Chard

has been accepted towards fulfillment
of the requirements for the

Ph.D degree in Measurement and Quantitative Methods

Mark D. Reckase
Major Professor's Signature

4/25/05
Date

USING MULTIDIMENSIONAL ITEM RESPONSE THEORY TO EXAMINE MEASUREMENT EQUIVALENCE: A MONTE CARLO INVESTIGATION

By

Linda Baumunk Chard

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Measurement and Quantitative Methods

2005

ABSTRACT

USING MULTIDIMENSIONAL ITEM RESPONSE THEORY
TO EVALUATE MEASUREMENT EQUIVALENCE:
A MONTE CARLO INVESTIGATION

By

Linda Baumunk Chard

This dissertation seeks to examine the accuracy of the *W*-index, a new multidimensional item response theory (MIRT) index of comparative fit to a measurement model involving multiple group respondents. To do this, the study utilizes simulated data with known properties. Specifically, it focuses on measurement equivalence as determined by similar factor structure, demonstrated by comparable model fit across groups. Additionally, the study examines the effects that variation in three experimental factors may have on the effectiveness of the *W*-index procedure as a scaling method. In particular, it examines how sample size, strength of intertrait correlation, and percentage of items lacking equivalence influence the detection of a lack of measurement equivalence within an MIRT structure. Finally, to illustrate a practical use of the *W*-index to examine measurement equivalence, it is applied to measures of "Teacher collective responsibility for student learning" collected from seven U.S. school districts. Here the purpose is to evaluate whether a battery of 26 items that were supposed to measure the latent trait of teacher collective responsibility for student learning actually did measure the same construct across groups.

The results show that the *W*-index procedure is a reliable MIRT method to identify a lack of measurement equivalence under certain conditions. Specifically, those conditions include a sample size of 2000 for any case or 1000, if the requirement for a weak intertrait correlation (.02) is met. Additionally, the small sample size of 150 may not result in an "Acceptable" identification of lack of equivalence, regardless of the other criteria. Contrary to expectation, the percentage of items lacking ME was not a critical factor for accurate identification with the *W*-index procedure.

DEDICATION

This work is dedicated, in loving memory, to my father,
my earliest and most demanding teacher.

ACKNOWLEDGEMENTS

No dissertation is ever completed in isolation, without the direction, assistance, and encouragement of countless professors, associates, friends, and family, and this is no exception. There is a lengthy list of people who have given me more than I can ever possibly repay, and I would like to acknowledge some of them.

The first is my spirited academic advisor, Dr. Ed Wolfe, who adamantly challenged all of his students to explore with fervor, to question with conviction, and to forge on regardless. Equally insistent was Dr. Mark Reckase, the chairperson for my dissertation committee. It is his ability to consistently ask the baffling questions that kept me searching for answers and delving deeper in the literature. Also, I own a great deal to Dr. Ken Frank, who first introduced me to the idea of teacher collective responsibility for student learning that later lead to my interest in measurement equivalence. Additionally, I wish to thank the other members of my committee, Dr. Joyce Grant, and Dr. Fred Oswald, for their insights, suggestions, and indispensable assistance.

I would like to also thank all the school administrators who allowed me to utilize their districts and to the teachers who so willing completed my survey without material compensation. Obviously, the data collection would have been much more cumbersome without this generous donation.

The final group I want to thank is the one to whom I owe the most: my family. Throughout this long and arduous journey, the value of their unquestioning love and support has been immeasurable.

TABLE OF CONTENTS

LIST OF TABLES x

LIST OF FIGURES xi

CHAPTER 1:
INTRODUCTION 1

Measurement Equivalence Defined 1
The Importance of Measurement Equivalence 2
Why Measurement Equivalence is not Routinely Investigated 4
Methods to Verify Equivalence 5
Structural Equation Modeling 5
Item Response Theory 5
Concerns of Measurement Equivalence Investigations 6
Research Questions 9

CHAPTER 2:
REVIEW OF THE LITERATURE 10

The Multidimensional Item Response Theory Approach 10
The Multidimensional Random Coefficients Multinomial Logit Model 12
Structure of the MRCMLM 13
The Context of Measurement Equivalence Investigations 15
Factors Studied in Measurement Equivalence Simulations 16
Sample size 16
Strength of intertrait correlation. 17
Number of items lacking equivalence 18
Common Methods to Assess Measurement Equivalence 19
Differential Item Functioning 19
Dimensionality 20
Model Fit 21

CHAPTER 3:
SIMULATION METHODOLOGY 24

Investigation Objective 24
The W-index: A Procedure to Access Across-groups Model Fit 24
Determination of W-critical Value 27
Assessment of Model Fit via ConQuest 28
Verification of Between-item Dimensionality 30
Simulation Study Overview 30
Multidimensional Item Response Data Sets 32

Constant Elements 32
Data Generation Procedure 32
Null Condition: P = 0 33
Experimental factors 34
Number of items lacking equivalence. 34
Sample size 35
Strength of intertrait correlation 35
Logistic Regression 35

CHAPTER 4:
SIMULATION RESULTS 37

The Null Condition 37
Descriptive Statistics 37
W-Critical Values for Simulated Data 37
Accurate Identification of Lack of Measurement Equivalence: Statistical Power 41
Results from Logistic Regression 42
Interactions 42
Main Effects 46
Variation in number of items lacking equivalence. 47
Variation in sample size. 48
Variation in intertrait correlation. 49

CHAPTER 5:
DISCUSSION OF SIMULATION RESULTS 50

Rates of Statistical Power 50
Variation in Number of Items Lacking Equivalence 50
Variation in Sample Size 52
Variation in Intertrait Correlation 53
The Effects of the Two-Way Interactions 54
Summary 55

CHAPTER 6:
REAL DATA METHODOLOGY 57

Survey Instrument 57
Instrumentation 57
Population 60
Data Collection 60
Data Analysis of the Survey Instrument 61
Verification of Between-Item Dimensionality 65
Determination of Model Fit 66
Exploratory Factor Analysis 68

CHAPTER 7:
REAL DATA RESULTS 69

Descriptive Statistics 69
W-critical Value from Simulated Real Data 70
Dissimilarity in Factor Loadings 70

CHAPTER 8:
REAL DATA DISCUSSION 73
Survey Items with Dissimilar Factor Loadings Across Groups 73
Implications of Efforts to Measure Teacher Collective Responsibility 77

CHAPTER 9:
CONCLUSIONS 79

Implications of the Findings 79
Consequences of Ignoring Measurement Equivalence 80
Issues for Future Research 84

APPENDICES

Appendix A. IRT Investigations of Effects of Variation in Experimental Factors 87

Appendix B. Teacher Collective Responsibility for Student Learning
Survey Instrument 93

Appendix C. SAS Code to Generate Data 97

Appendix D. WINSTEPS Code to Generate Data 100

Appendix E. SAS Code to Create W-statistic for Groups and Merge 101

Appendix F. SAS Code to Identify W-Critical Value for Null Condition 102

Appendix G. SAS Code to Identify Statistical Power Rate 103

Appendix H. Frequency Distribution of *W*-index – Simulated Null Condition 106

Appendix I. SAS code for Logistic Regression 156

Appendix J. Results of Logistic Regression 159

Appendix K. Statistical Power for W-index Procedure by Number of Items Lacking Equivalence.....162

Appendix L. Statistical Power of W-index Procedure by Sample Size.....163

Appendix M. Statistical Power of W-index Procedure by Intertrait Correlation.....164

Appendix N. Factor Loadings for Real Data Survey Instrument.....165

Appendix O. Eigenvalues and Scree Plot for Real Data.....166

Appendix P. Factor Correlations –Elementary and Secondary Real Data.....167

Appendix Q. Frequency Distribution of *W*-index - Real Data.....168

Appendix R. Exploratory Factor Analysis – Elementary Real Data.....171

Appendix S. Exploratory Factor Analysis – Secondary Real Data.....173

REFERENCES.....176

LIST OF TABLES

Table 1. Characteristics of Variation in Experimental Factors.....34

Table 2. Descriptive Statistics for Null Condition, Simulated Data.....37

Table 3. Descriptive Statistics for *W*-index, Null Condition.....38

Table 4. *W*-Critical Values for Null Condition.....39

Table 5. Type I Error Rates for Second Null Data Sets.....39

Table 6. Statistical Power of *W*-index 40

Table 7. Logistic Regression Results – Two-Way Interaction.....42

Table 8. Power of *W*-Index for the Sample Size-by-Intertrait Correlation Interaction.....42

Table 9. Statistical Power of *W*-Index for the Number of Items Lacking Equivalence-by-Sample Interaction.....44

Table 10. Logistic Regression – Main Effects Only..... 45

Table 11. Instrument Blueprint.....45

Table 12. Factor Correlations.....61

Table 13. Rating Scale Analysis.....62

Table 14. Descriptive Statistics for Real and Simulated Demographic Groups.....69

Table 15. *W*-statistic and Rejection Conclusion.....70

Table 16. Factor Loadings for Elementary and Secondary Real Data.....71

LIST OF FIGURES

Figure 1. Two-way Interaction of Sample Size and Intertrait Correlation on Statistical Power Developmental Model.....43

Figure 2. Two-way Interaction of Sample Size and Number of Items Lacking Equivalence on Statistical Power 44

Figure 3. Main Effect for Number of Items Lacking Equivalence 46

Figure 4. Main Effect for Sample Size.....47

Figure 5. Main Effect for Intertrait Correlation 48

Figure 6. Developmental Model 57

CHAPTER 1: INTRODUCTION

An essential attribute of any psychological or behavioral instrument is measurement equivalence. That is, the instrument must measure the intended construct equally well across measurement contexts such as instrument forms, measurement occasions, raters, or subpopulations. On the surface, this seems a simple concept. Unfortunately, this is not the case. In truth, the issue of measurement equivalence (ME) is multi-faceted and perplexingly complex, resulting in numerous definitions and varying procedures for investigation. The importance of ME is such that it is referred to by some as a "prerequisite" for group comparisons (Riordan, Richardson, Schaffer, & Vandenberg, 2001). Regardless, evaluations of measurement equivalence between groups are not routinely performed by data analysts. As a result, the validity of conclusions drawn from studies where measurement equivalence is not considered may be in question (Vanderberg & Self, 1993).

Measurement Equivalence Defined

The definition of measurement equivalence chosen for this study is that of Cheung and Rensvold (2002), who describe it as the condition whereby members of different groups associate survey items, or similar measures, with similar constructs. ME refers to "whether or not, under different conditions of observing and studying phenomena, measurement operations yield measures of the same attribute" (Horn & McArdle, 1992, p.117). The specific attribute examined, which will be addressed later in more detail, varies from study to study, depending on which psychometric properties are investigated. The primary question being asked in an examination of measurement

equivalence, as it is considered in the study presented here, is "do the measures being assessed represent the same construct between subgroups of the population being measured?" When applied to a psychological or behavioral instrument, a lack of ME indicates that measures from the instrument do not mean the same from one group to another (Cheung & Rensvold, 1999; Vandenberg & Lance, 2000). Thus, by definition, measures lack equivalence unless they measure the same construct with similar precision across groups or populations. Lack of equivalence can be inferred when the psychometric properties of an instrument are not comparable across groups (Hui & Triandis, 1985; Knight & Hill, 1998).

The Importance of Measurement Equivalence

ME is essential for all behavioral and psychological instruments because, according to Riodan and Vandenberg (1994), only when subjects from different groups ascribe essentially the same meaning to the scale or items can meaningful across-group comparisons be conducted. Routinely, researchers compare the mean response values for various demographic groups based on measures that are drawn from an instrument designed to measure a particular latent trait. From these observations, substantive inferences are made concerning between-group differences in the level of the construct purportedly represented by the measures. This creates a disconcerting situation: although the observed differences might well be due to the way the construct is conceptualized in each group rather than true group differences, a study of the measurement equivalence of the measures from the instrument for these groups is seldom conducted. Thus, the validity of these inferences is dependent on the often untested assumption that, across groups, the measures carry the same meaning for the construct. When this assumption of

measurement equivalence is in fact violated, absolute differences in scores between groups, and, therefore, inferences based on these differences, are likely to be misleading (Chan, 2000). This presents a serious problem for researchers. If the construct of interest is not measured equivalently across groups, then a comparison of means across groups may be inaccurate, unwarranted, or even meaningless (Golembiewski, Billingsley, & Yeager, 1976; Schmitt, 1982; Vandenberg & Self, 1993).

Some researchers, such as Horn and McArdle (1992), have recognized this fact and attempted to make others aware of it. They pointed out the problem of not conducting ME analyses by writing

> If there is no evidence indicating presence or absence of measurement equivalence-- the usual case -- or there is evidence that such equivalence is not obtained, then the basis for drawing scientific inference is severely lacking: findings of differences between individuals and groups cannot be unambiguously interpreted (p. 117).

In spite of this and similar attempts to alert researchers to the importance of establishing measurement equivalence, most seem to be unaware of or have elected to disregard the warnings. In a synthesis of the measurement equivalence literature completed in 2000 involving 65 studies, Vandenberg and Lance found a substantial number of cases where inaccurate inferences would have been made by the various researchers if they had not undertaken the ME tests. In this account, they insist that "tests of ME should be routinely conducted prior to conducting tests aimed at evaluating cross-group differences" (p. 47). Hence, to avoid costly errors and to produce compelling research results, prior to making direct between-group comparisons, it must be verified that the measures from the instrument being used do not lack measurement equivalence. According to Reise, Widaman, and Pugh (1993),

> Measurement equivalence is a basic requirement or prerequisite for studying group differences with statistical models. Once measurement equivalence is established, additional theoretically important questions may be addressed, including questions regarding group differences in means or variances on the latent variables identified (p. 562).

To do this, it is essential that reliable and valid methods for evaluating measurement equivalence are developed. These methods can then be routinely applied to psychological or behavioral instruments before comparisons of groups are made. Once it has been verified that the measures do not lack ME, the means of latent variables can be suitably compared (Bollen & Long, 1993; Byrne, Shavelson, & Muthén, 1989; Millsap & Everson, 1991; Riordan & Vandenberg, 1994).

Why Measurement Equivalence is not Routinely Investigated

The use of the term "equivalence" is relatively new, but the underlying concept goes as far back as the work of Karl Pearson in the early 1900s (Millsap & Meredith, 2004). Even though a considerable amount of time has passed since its conception, ME still does not enjoy the usage it warrants, given its importance. According to Steenkamp and Baumgartner (1998), the exclusion of a verification of measurement equivalence from routine data analysis exists for a variety of reasons. First, there is a bewildering array of types and classifications of equivalence found in the literature. Also, there is little consistency in the use of the term ME in the literature. Moreover, many researchers are relatively unfamiliar with models that incorporate the means of latent and observed variables. This is compounded by the fact that there are substantial methodological complexities involved in testing for measurement equivalence, particularly if the data is multidimensional. In real-world contexts, the latter is often the case. Added to this, many of the existing methods are inappropriate for certain types of investigations, particularly

those involving real data and assumptions of unidimensionality or normality. Finally, there is an absence of clear guidelines as to how to ascertain whether or not a measure exhibits "adequate" equivalence. In totality, these factors result in uncertainty, confusion, and the avoidance by many of crucial measurement equivalence substantiation.

Methods to Verify Equivalence

Structural Equation Modeling

In measurement equivalence examinations, the most commonly employed statistical procedure is structural equation modeling (SEM), which uses confirmatory factor analysis (CFA) procedures. In doing this, the most conventional procedure to verify that the items on a given instrument do not lack equivalence is the demonstration of equality of factor loadings (Byrne, Shavelson, & Muthén, 1989; Horn & McArdle, 1992; Rensvold & Cheung, 2001; Schmitt, 1982; Vandenberg & Lance, 2000; Vandenberg & Self, 1993). A second common criterion for equivalence investigation is equality of factor covariances (Schaubroeck & Green, 1989; Schmitt, 1982; Vandenberg & Self, 1993). A third is the equality of the error variance/covariance matrices (Byrne, 1994; Drasgow & Kanfer, 1985; Marsh & Hocevar, 1985; Mullen, 1995). Finally, the equality of variance/covariance matrices of latent variables is a fourth common SEM criterion for evaluation (Byrne, 1994; Jackson, Wall, Martin, & Davids, 1993; Marsh, 1993, Marsh & Hocevar, 1985).

Item Response Theory

Item response theory (IRT), a measurement model that has been widely adopted in the psychometric literature, has been less visibly investigated as a means for evaluating ME. As an alternative to SEM, IRT methods can, in some cases, "provide different and

potentially more useful information for the establishment of measurement invariance" (Meade, Lautenschlager, Michels, & Gentry, 2004, p. 362). In its favor is the fact that IRT methods are not forced to meet the normal distribution assumption that plagues existing methods based on CFA. Thus, they are more appropriate in situations in which the assumption of normality may not be met. It is also to their advantage that sample-free item parameter estimates and test-free ability estimates can be obtained (De Champlain & Gessaroli, 1996).

As a result of increased use, within the IRT framework, several approaches to investigating ME have been devised. Among these is that of model fit. This procedure is based on the views of researchers such as Hambleton, Swaminathan, and Rogers, who contend that "Equivalence only holds when the fit of the model to the data is exact in the population" (1991, p. 23). This notion is the focus of the research presented in this dissertation. Specifically, this dissertation seeks to evaluate the performance of a new index for evaluating ME using a measure of model fit between groups of respondents to a survey instrument using item response theory in a multidimensional setting.

Concerns of Measurement Equivalence Investigations

Because measurement equivalence investigations that examine factorial structure in multidimensional item response theory (MIRT) are relatively new, as with almost any fledgling area of research, there are still some unresolved concerns. The first concern is one that is basic to any study. That is, what method or procedure is most effective for the proposed investigation? In previous studies, some investigators have found a particular IRT or MIRT-based procedure to be effective while others find it is not. As a result, the researcher is left in a quandary as to what procedure may effectively be used in a given

situation. This may, in part, account for the less frequent use of MIRT procedures as compared to the more popular SEM methods.

Another concern arises from the relatively small number of measurement equivalence investigations currently being conducted, particularly using MIRT. Because the number is small, there are fewer well-established guidelines or quantitative criteria that may be used to make critical decisions in MIRT than in SEM. For instance, there is a conspicuous absence of clear guidelines as to how to ascertain whether or not a measure exhibits "adequate" equivalence. Additionally, dissimilar findings have been presented due to the fact that, although the intent of the studies is the same, the designs may not be. Prime examples of this are found in the research reports of the effects on the detection rate of lack of ME as a result of variation in the measurement context. With time and additional studies that are similar in design, this concern may be overcome. However, such is not now the case.

A review of the literature confirms that there are not as many investigations concentrating on ME as other research areas. This supports the concern by investigators that there simply are not enough corroborating studies of equivalence, particularly ones that attempt to determine the condition under which competing methods result in different conclusions. This view is expressed by Vandenberg (2002), who is one of the many researchers calling for additional studies involving measurement equivalence analyses. This view is also supported by another group of researchers, of which Vandenberg is a part (Riordan et al., 2001), who also actively seeks an increase in Monte Carlo studies to determine the accuracy of the existing methodologies intended to identify a lack of measurement equivalence. In his writings, Vandenberg strongly advocates

research that compares the efficiency of one procedure to that of another under a variation in measurement context. His concern is that there is developing an "unquestioning faith on the part of some that the technique [being used] is correct or valid under all circumstances" (p. 140). As a result of the insistence, a number of investigators conducted promising research to examine equivalence using both of the two most common methods: SEM and IRT (Facteau & Craig, 2001, Maurer, Raju, & Collins, 1998; Raju, Laffitte, & Byrne, 2002; Reise et al., 1993). However, at this point, this number is also small.

A sizeable number of researchers have employed structural equation modeling methods to address the equivalence issue essential for convincing and compelling comparisons of group means. However, generally speaking, those who apply IRT models have not followed their lead. Thus, these investigators inadvertently run the risk of drawing conclusions that may be misleading, inaccurate, or even erroneous. To address some of the concerns found in equivalence investigations and the lack of generally accepted methods for determining a lack of measurement equivalence in the commonly adopted framework of item response theory, this study focuses on the following issues. First, it examines the accuracy of a new multidimensional item response theory (MIRT) index of comparative fit to a measurement model with multiple groups of respondents, referred to as the *W*-index. To do this, this study utilizes simulated data with known properties. Specifically, it focuses on measurement equivalence as determined by similar factor structure, demonstrated by comparable model fit across groups. Second, this study examines the effects that variation in the measurement context may have on the effectiveness of the *W*-index MIRT procedure as a scaling method. In particular, it

examines how the percentage of items lacking equivalence, sample size, and strength of intertrait correlation influence the detection of a lack of measurement equivalence within an MIRT structure. Finally, to illustrate a practical use of the *W*-index to examine measurement equivalence, it is also applied to measures of "teacher collective responsibility for student learning" collected from seven US school districts. Here the purpose is to evaluate whether a battery of 26 items that were supposed to measure the latent trait of teacher collective responsibility for student learning actually did measure the same construct across groups.

Research Questions

Thus, to accomplish the intended purposes, the following questions are posed for this study:

1) Can the *W*-index method using factorial structure equality accurately identify a lack of measurement equivalence in a survey instrument?
2) Is the accuracy of the *W*-index of measurement equivalence using factorial structure equality affected by variations in the number of items lacking equivalence?
3) Is the accuracy of the *W*-index of measurement equivalence using factorial structure equality affected by variations in sample size?
4) Is the accuracy of the *W*-index of measurement equivalence using factorial structure equality affected by variations in the strength of the intertrait correlation?

CHAPTER 2: REVIEW OF THE LITERATURE

This chapter reviews the multidimensional item response theory approach to measurement equivalence investigation, some of the most common methods that employ this approach, and results of prior studies involving ME. Additionally, a detailed discussion is presented of the multidimensional random coefficients multinomial logit model (MRCMLM) used in the study.

The Multidimensional Item Response Theory Approach

Early investigations of measurement equivalence were performed as a result of attempts to identify violations of the unidimensionality assumption that is commonly evoked for the sake of simplifying the creation of measures from responses to an educational or psychological instrument. Researchers quickly discovered that in real-world contexts, the unidimensional assumption is often difficult to support (Nandakumar, 1994). As a result, multidimensional item response theory models gained some popularity. Although investigations of measurement equivalence using multidimensional item response theory (MIRT) are comparatively new, the basic procedures are not. According to Hambleton & Swaminathan (1985), basic IRT methods have been employed for almost 50 years. A review of the current ME literature involving MIRT methods verifies that, although still relatively small, there is a notable growth in the number of studies in recent years. One reason for this is that improved computer software production has facilitated the application of all IRT methods to investigate a lack of ME and has now placed the complexity of multidimensional investigations within the capabilities of nearly all researchers. This has significantly increased the ability of MIRT methods to compete with the more well-established SEM methods.

Multidimensional item response theory procedures are systems designed to determine consistent features of persons and items that influence responses, within a multidimensional framework. In many cases, MIRT models are expansions of unidimensional models that stipulate a nonlinear monotonic item response function to account for the relationship between examinee level on a latent variable and the probability of a particular item response (Linden & Hambleton, 1997; Lord, 1980). According to Reckase (1997), multidimensional item response theory (MIRT), consists of a general class of models that describe the interaction between persons and test items where

> the characteristics of the person are described using a vector of hypothetical constructs. Further, the characteristics of the test items are described using a set of item parameters and a functional form that relates location in the space defined by the vector of person parameters to the probability of correct response to each item (p. 25).

Here the focus is on modeling the relationship between person and test items. Thus, the individual characteristics of the items are the center of attention in the investigation. This is rooted in of the thinking of Lord (1980), who supported a need

> to describe the items by item parameters and the examinees by examinee parameters in such a way that we can predict probabilistically the response of any examinee to any item even if similar examinees have never taken similar items before (p. 11).

In MIRT, initially, a model is created representing the interaction between persons and test items. The intent is to accurately reproduce the probability of a correct response to an item for individuals at a particular point in the θ space. Each item is of concern as it is examined for appropriate fit. Concern is raised if there is a discrepancy in the predicted probabilities for a particular range of abilities (Drasgow, Levine, & McLaughlin, 1991). Here the focus is on conditional measures of fit.

The estimate for a given person is based on observed item responses given the item parameters (Meade et al., 2004). The exact nature of the model to be used in the investigation is determined by a set of item parameters that are potentially unique for each item. In a simulation study, there are numerous item response models to select from. Thus, it is of importance to select a model representative of the specific situation of interest and the nature of the data to be generated. One such model that is representative of the data in this study is the multidimensional random coefficients multinomial logit model (MRCMLM).

The Multidimensional Random Coefficients Multinomial Logit Model

In the social sciences, log-linear models have been employed for several decades (Keldermna & Rijkes, 1994; Knoke & Burke, 1980) with numerous multidimensional item response theory models being used (Ackerman, 1992; Camilli, 1992; Embretson, 1991; Glas, 1992; Luecht & Miller, 1992; Oshima & Miller, 1992; Reckase, 1985). Of the many current methods available for use with multidimensional data, the one chosen for this study is the Multidimensional Random Coefficient Multinomial Logit Model (MRCMLM; Adams, Wilson, & Wang, 1997), which is a multidimensional extension of the Rasch model (Xie, 2001).

The MRCMLM was selected for this study for multiple reasons. First, it is appropriate for the real data, which is known to be multidimensional. Second, it does not necessitate a large sample size--the sample size for the real data example used in this dissertation is 616. Third, Adams et al. (1997) demonstrated the MRCMLM was a mathematically tractable and flexible multidimensional model that produces parameter estimates that are readily interpretable. Fourth, it draws on the (often strong) relationship

between the latent dimensions to produce more accurate parameter estimates and individual measurements. Last, and most importantly, as an adaptation of an IRT method, the model does not necessitate meeting the normality assumption that other often-employed methods, particularly in structural equation modeling, do.

Although the name MRCMLM is rather long and, at first, daunting, it can be broken down into meaningful factors. Beginning with the left most word in the title, the M, "multidimensional", refers to the ability of the model to incorporation several latent traits. This is particularly helpful in working with real data that is seldom "truly unidimensional." RC or "random coefficients" indicates that the model incorporates random effects. This is slightly misleading, as it is actually a "mixed" model that is capable of incorporating both fixed and random effects. MLM, "multinomial logit model" (Amemiya, 1985) refers to a regression model that is applicable when the dependent variable takes on discrete values (Adams & Wilson, 1996). This regression model is used to decompose the location parameter into factors called base parameters. Although just the 1-parameter model using only the location parameter is presented here, there is also a 2-parameter model that uses both slope and location (Valbuena, 2002).

Structure of the MRCMLM.

The following explanation of the MRCMLM is adapted from that given by Briggs and Wilson (2003). The MRCMLM assumes a set of D traits underlie the respondents' responses. In the MRCMLM, the position of a person (n) on the D-dimensional latent space is represented by a vector of latent traits $\theta_n = [\theta_{n1}, \theta_{n2},\ldots,\theta_{nD}]$, where the D dimensions may be non-orthogonal. These vectors can be appended across persons to create an N x D matrix of positions in the latent space, Θ. An item difficulty index, δ_{ik},

depicts the relative difficulty of surpassing threshold k of item i (i.e., responding with category x rather than category x-1 on the rating scale, where there are k-1 categories). Item difficulties can be appended to create a vector of item difficulties, δ. A response in category k in dimension d of item i is scored b_{ikd}.

The probability of a response in category x for item i is modeled as

$$\pi_{nix} = \frac{\exp(b'_{ix}\theta + a'_{ix}\delta)}{\sum_{x=1}^{X_i} \exp(b'_{ix}\theta + a'_{ix}\delta)} \qquad (1)$$

The b_j parameters are called category difficulties or thresholds. Each is defined as the point on the theta scale (the trait level) at which the probability is 50% that the item response is greater than threshold j (Reise et al., 1993). The intended dimensional structure of the model is depicted using two matrices composed of vectors that relate each item to the underlying dimensions. These two are the design matrix (**A'**) and the scoring matrix (**B'**).

The design matrix, $\mathbf{A'} = (a_{11}, a_{11}, \ldots, a_{ik})$, consists of item scores mapped to their intended dimensions, for each item. The number of rows is equal to the total number of response categories for all generalized items.

To create the scoring matrix, **B'**, the scores across D dimensions can be collected into a column vector $\mathbf{b'}_{ik} = [b_{ik1}, b_{ik2}, \ldots, b_{ikD}]$, then collected into the scoring submatrix for item i, $\mathbf{B'}_i = (b_{i1}, b_{i2}, \ldots, b_{iK})$, and then collecting into a scoring matrix $\mathbf{B'} = (\mathbf{B'}_1, \mathbf{B'}_2, \ldots, \mathbf{B'}_I)$ for the whole test.

The Context of Measurement Equivalence Investigations

Previously, the most common venues for studies of ME were across cultures (Jansens, Brett, & Smith, 1995; Reise et al., 1993; Riordan & Vandenberg, 1994; Windle, Isawaki, & Lerner, 1988). However, additional interest in cross-group measurement equivalence has resulted in both increased use in this area and a salient expansion to others. Many of these additional investigations are across a variety of demographic groups other than those defined by ethnicity. Some of the other group classifications include gender (Byrne, 1994; Collins, Raju, & Edwards, 2000), differing levels of academic achievement (Byrne et al., 1989), rater groups (Facteau & Craig, 2001; Pentz & Chou, 1994), and aspects of industrial organization (Drasgow & Kanfer, 1985).

Another prominent focus of investigations involving measurement equivalence is the stability of measures across measurement conditions, such as different media of measurement administration like those found in a web-based survey versus a paper-and-pencil survey (Donovan, Drasgow, & Probst, 2000; Meade et al., 2004; Taris, Bok, & Meijer, 1998). Still others are concerned with stability of measurement over time (Golembiewski et al., 1976; Riordan et al., 2001; Taris et al., 1998). Even the already strong interest in cross-culture investigations of ME has increased recently (Ghorpade, Hattrup, & Lackritz, 1999; Ployhart, Wiechmann, Schmitt, Sacco, & Rogg, 2002; Steenkamp & Baumgartner, 1998). This upsurge may be attributed partially to the explosive growth of international markets and the ascendancy of multinational organizations (Triandis, 1994).

Factors Studied in Measurement Equivalence Simulations

The effect of a great many contextual factors on the accurate verification of ME has been investigated. Some of the most frequently included factors in simulation and Monte Carlo investigations are the effects of test length (De Champlain & Gessaroli, 1991; De Champlain, Gessaroli, Tang, & De Champlain, 1998; Flowers, Oshima, & Raju, 1999), the effects of intertrait correlation (Gosz & Walker, 2002; Hambleton & Rovinelli, 1986; Nandakumar, 1994; van Abswoude, van der Ark, & Sijtsma, 2004), and the effects of theta location (Seraphine, 2000). Other studies have examined the effects of number of traits (van Abswoude et al., 2004), the effects of the number of variant items (Gosz & Walker, 2002; Hambleton & Rovinelli, 1986; van Abswoude et al., 2004), the effects of sample size (De Champlain & Gessaroli, 1991; De Champlain et al., 1998), and the effects of number of scale (Seraphine, 2000). A listing of these studies, as well as their findings and other pertinent information, is presented in Appendix A.

Sample size.

One of the largest groups in these studies focuses on the influence of sample size on the rate of accurate detection of lack of ME (Boles, Dean, Ricks, Short & Want, 2000; Davidson & Chen, 1991; Facteau & Craig, 2001; Flowers, 1996; Idaszak, Bottom, & Drasgow, 1988; Knol & Berger, 1991; Luczak, Raine, & Venables, 2001; Martin & Firedman, 2000; Meade et al., 2004; Schaubroeck & Green, 1989; Schmitt, 1982; Vandenberg, 2002; Vandenberg & Self, 1993; Yoo, 2002). Several previous simulation studies have used as a "large" sample size 1000 or 2000 (Cohen & Kim, 1992, 1993; Lim & Drasgow, 1990), while 150 is common for a "small" sample size (Hidalgo-Montesinos & Lopez-Pina, 2002; Meade et al., 2004).

Typical of the findings that identification of lack of ME is more accurate with larger sample sizes are those from De Champlain and Gessaroli (1996). Their study was designed to identify lack of ME through dissimilar dimensionality across groups using the G^2 statistic with *TESTFACT.* The results showed a very slight increase in accuracy (as displayed by a decrease in the rate of false acceptance) when the sample sizes was increase from 250 to 500 (.07 to .06), but was significantly more accurate when the sample size was increased to 1000 (.02). In line with this, additional studies involving samples sizes of 150 (Hidalgo-Montesinos & Lopez-Pina, 2002; Meade et al., 2004) determined that identification of a lack of ME was not as accurate with this small sample size. Thus, based on findings such as these, it is hypothesized that, in this study, the rate of accurate identification of lack of equivalence will be smallest when the samples size is small (n = 150) and will increase with an increase in sample size, such that the best rate is obtained when the sample size is largest (n = 2000).

Strength of intertrait correlation.

There are also some notable findings concerning the effect of the strength of the intertrait correlation, as identified by a variety of procedures, utilizing commercially produced software. Generally, the accuracy of the procedures decreases with an increase in the intertrait correlation. However, there is no agreement as to the point at which accurate identification can no longer be made. As might be expected, the specific intertrait correlation values needed for accurate identification of lack of ME vary from procedure to procedure. For example, Nandakumar (1994) found Stout's t-statistic, as implemented in *DIMTEST,* to be effective when the intertrait correlations were as high as .70. In another study, Gosz and Walker (2002) found that although one test of ME

(implemented in *NOHARM;* Fraser, 1985) accurately identified lack of equivalence only up to intertrait correlations of .50, another (implemented in *TESTFACT;* Wilson, Wood, & Gibbons, 1991) continued to performed well, even with high intertrait correlations of .90. Using *TESTFACT* to identify false acceptance rather than accurate rejection, De Champlain and Gessaroli (1996) reported a perfect rate for false acceptances (0.00) when the intertrait correlation was zero. But that rate (indicating inaccuracy) rose to 0.10 when the intertrait correlation was increased to .70. These variations in findings come as no surprise, based on the diversity of methods. Nevertheless, it poses a problem for the researcher as to what criteria to use. From these studies, a definitive conclusion can not been drawn as to a value that signifies the point at which identification can no longer accurately be made for all procedures currently available. For this study, the hypothesis is made that, in line with some prior research, accurate identification of lack of equivalence will be made with intertrait correlations of .40 or less, and the accuracy rate will decrease with an increase in the strength of the intertrait correlation.

Number of items lacking equivalence.

There is a similar diversity in findings on the effect of number or percent of items lacking equivalence. One example comes from a study by Hambleton and Rovinelli (1986) involving six tests for lack of ME. They found that *TESTFACT* was effective when only 30% of the total instrument items lacked equivalence. However, for the other 5 tests in the same study, (three methods of linear factor analysis, a residual analysis, and Bejar's method), they reported that for accurate identification, these test required 50% of the total number of items lack ME. As with other experimental factors, the situation exists that, across procedures and indices, the percentage of items on the instrument

needed for accurate identification of lack of ME varies. Again, it is difficult to make a direct comparison between findings, with different IRT or MIRT methods, different variations in contextual settings, and different research designs. In the investigation presented here, the maximum percentage of items lacking equivalence being investigated is 23% (6 items). Thus, based on previous findings, it is hypothesized that in this study, the most accurate identification of lack of equivalence will be made with the largest number of items (6 items or 23%) but will decrease when a smaller percentage of items lack equivalence.

Common Methods to Assess Measurement Equivalence

Differential Item Functioning

Within the IRT framework, there are multiple methods to investigate a lack of measurement equivalence (McKinley & Mills, 1985). Regrettably, none of these has been universally accepted. Of these, the most common method to assess equivalence is an examination of differential item functioning (DIF) across groups of interest. An item is defined to have DIF if respondents with the same ability but from different groups do not have the same probability of endorsing the item (Hambleton et al., 1991). Numerous indices exist for this purpose, but all of those indices are designed to determine whether the responses of members of subgroups or subpopulations to a particular item are consistent with their joint responses to the remaining items on the instrument. Hence, DIF indices seek to determine whether ME exists between subgroups with respect to their responses to individual items on the instrument. This item-level concept has also been expanded to a more extensive examination that includes overall *test* differential

functioning, as well as *item* differential functioning in a recently-emerging concept known by the acronym DFIT (Raju, van der Linden, & Fleer, 1995).

Dimensionality

Other prior investigations of ME have been concerned with differential dimensionality between subgroups. Most of the indices designed for this purpose are commonly used to evaluate threats to the unidimensionality, although they could be adapted for the purpose of evaluating whether differential dimensionality between subgroups exists. Additionally, many of these procedures have software specifically designed to facilitate their application. One of the best known is Stout's t-statistic test of essential dimensionality, facilitated by the computer programs *DIMTEST* (Stout, 1987), *DETECT*, and *Poly-DIMTEST*. *DIMTEST* has been shown repeatedly to effectively identify dimensionality in single test situations (De Champlain & Gessaroli, 1991; Hattie, 1996; Nandakumar, 1994; Seraphine, 2000; van Abswoude et al., 2004). Other well-known tests include Bock's full information factor analysis G^2_{diff} statistic (1988), used in *TESTFACT*; McDonald's nonlinear factor analysis (*NOHARM,* 1981, 1993) and the Holland and Rosenbaum's method (1986).

In spite of their appropriateness for some investigations, for a simulation study involving Likert-scale survey items and multidimensionality, these methods are inappropriate for two reasons. First, they are designed for a single test administered to a single group of examinees within an exploratory factor framework. As noted by Byrne and Campbell (1999), even though a given measurement may report accurately *within* each of two or more groups, there is no guarantee that the measurement will operate equivalently *across* groups. Winter and Prohaska (1983) support this view in their

statement that "a measurement tool which works for one group may not work for another" (p. 422). Second, some of the indices employed are intended for dichotomous items and may not be effective when applied indiscriminately to polytomous or Likert-scale data (Adams et al., 1997). Rather, a multidimensional, or MIRT, procedure that can accommodate Likert-scale response items and multiple examinee groups is required for this study.

Model Fit

A third more serviceable procedure to identify a lack of measurement equivalence is to compare the model fit or value of the fit function across groups. Customarily, fit is assessed at the item level by a statistic that depicts the congruence between the proportion of item responses in a particular category predicted and the proportion of responses in a particular category observed in the data (Hui & Triandis, 1985; Knight & Hill, 1998).

One common index used for this is the likelihood ratio (LR) test (Thissen, Steinberg, & Wainer, 1988, 1993). In a unidimensional setting where the LR is to be used, a baseline model is generated in which all item parameters for all test items are constraint so that item parameters for like items are equal across measurement contexts. This model provides a baseline likelihood value, L_c, for item fit to the model (the *c* standing for compact). Additionally, a second nested model is generated with some parameter(s) changed. The specific change is defined by the design of the investigation. From this model, a likelihood value, LR_{a_i}, is also obtained (the *a* standing for augmented). The two values are then compared, creating a likelihood ratio, LR_i, such that

$$LR_i = \frac{L_C}{L_{a_i}} \tag{2}$$

where L_c is the likelihood function of the baseline model and LR_{A_i} is the likelihood function in which item parameter(s) of item *i* are allowed to vary (Meade et al., 2004). From this, a natural log transformation is taken, which results in a test statistic, $\chi^2_{(M)}$, distributed as a chi-square, where

$$\chi^2(M) = -2\ln(LR_i) = -2\ln L_c + 2\ln L_{a_i} \tag{3}$$

with M equal to the difference in the degrees of freedom between models.

In reality, this is a "badness-of-fit" test, where a statistically significant result implies the baseline model fits significantly more poorly than the manipulated model. Thus, a rejection of the null hypothesis indicates that there is a difference between the two models or that there is a lack of equivalence with regard to item *i*. To complete the investigation, the LR test is applied individually to each item in the instrument in order to verify equivalence for all items. As would be expected, it is highly unlikely that a ratio exactly equals one, indicating parameter equality across groups, for all items. Rather, a ratio is sought that is not significantly different from one. Thus, the assessment is more an evaluation of partial equivalence accompanied by an evaluation of the degree to which variance will be tolerated.

This concept of model fit has also been expanded for application to the multidimensional situation. Here a fit statistic commonly reported is identified by the term "deviance," which is defined as

$$\text{Deviance} = -2 * (Lm - Ls) \quad (4)$$

where *Lm* denotes the maximized log-likelihood value for the model of interest, and *Ls* is the log-likelihood for the saturated model (http://www.statsoft.com/textbook/glosd.html). This statistic is distributed as a chi-square with degrees of freedom equal to the number of parameters that are unconstrained in *Lm* as compared to *Ls*. The deviance statistic is not typically interpreted on its own. Rather, it provides a numerical value for the degree to which the fit of the model estimated from the given parameters deviates from the model generated by the data.

CHAPTER 3: SIMULATION METHODOLOGY

In the next three chapters, a study is described in which simulated data were used to determine the degree of accuracy in identifying a lack of ME using an MIRT index of model fit under variations in measurement context. This chapter explains the methodology and gives a detailed description of the index as well as the software used.

Investigation Objective

The intent of this study is to examine the use of a new index, the *W*-index, which can be utilized in the context of multidimensional item response theory (MIRT) for the purpose of identifying a lack of measurement equivalence (ME) between subpopulations of survey respondents. The position is taken that a lack of equivalence is established by demonstrating different factor structures for the same latent construct across groups of interest (Buss & Royce, 1975; Mullen, 1995) as exemplified by lack of model fit. This is based on the definition of equivalence employed by Hambleton et al, (1991), who stated that "equivalence only holds when the fit of the model to the data is exact in the population" (p. 23). Thus, if a difference across groups is found in the degree to which the given model fits the data, the instrument lacks measurement equivalence.

The *W*-index: A Procedure to Access Across-groups Model Fit

The following section describes the index developed for this study, which is based on a comparison of model fit between two groups and can be used to assess measurement equivalence within an MIRT context. The procedure relies on a comparison of the deviances of item responses from each group to a common MIRT configuration. The group for whom an expected MIRT structure is specified is the reference group; the other group is the focal group.

Although the deviance statistic provides a measure of model fit for a given situation, there is no existing index to compare fit across models, thereby determining if one model fits significantly best or worse than another under varying conditions. For that reason, the *W*-index, was developed for this study. To compute this, first, a proportionality constant (PC) was created, defined by

$$PC = \frac{deviance}{(n-p)} \tag{5}$$

where n = sample size; p = number of parameters estimated.

Then the PC value for focal group was compared to that for the reference group as a ratio:

$$W = \frac{PC_{focal}}{PC_{reference}} \tag{6}$$

Thus, this ratio may be distributed in a form similar to an *F*-statistic, as it meets the definition imposed by Hays (1988) for the *F* variable as "a random variable formed from the ratio of two independent chi-square variables, each divided by its degrees of freedom (1988, p. 332). The required assumption of normality for the F-ratio is met by sufficiently large sample size under the Central Limit Theorem.

The null hypothesis to be tested is

H_0: *W*–index = 1,

indicating the fit of the data to the model is statistically equivalent across groups.

A rejection of the null hypothesis, at the customary rate of $\alpha = .05$, indicates a lack of equivalence because the fit to the model of the data response sets for the reference and focal groups differ by more than can be expected due to random sampling.

It is important to point out that a conclusive determination of the lack of measurement equivalence should not be made solely on the rejection of or failure to reject the null hypothesis. Two situations exist that warrant additional substantive investigation. First, there is the possibility that a large number of items lack equivalence for both groups of interest. Such a situation would result in similar exceptionally large deviance values. Thus, the resultant *W*-index would be statistically close to 1, leading to a failure to reject the null hypothesis. Therefore, an inspection of the relative size of the deviance as well as the total number of percentage of items lacking equivalence should also be completed to verify items are not "equally bad" across groups.

Additionally, it is important to note that in some cases including items that lack measurement equivalence across groups may not necessarily be undesirable. For example, in prior cross-national investigations, it has been clearly established that some constructs are consistently interpreted differently due to cultural differences (Cunningham, Cunningham, & Green, 1973; Cole & Maxwell, 1985; England & Harpaz, 1983; Hui & Triandis,1985; Mullen, 1995; Singh, 1995; Steenkamp & Baumgartner, 1998). The recognition and acknowledgement of this fact is important in a thorough measurement equivalence examination. As a result, the identification of items displaying dissimilar factor loadings should be followed by an assessment of the content of these items and an attempt to quantify why such dissimilarity exists.

Determination of W-critical Value

Unfortunately, the exact shape of the null distribution of the *W*-index is unknown. Hence, we relied on a Monte Carlo approximation of that sampling distribution for the sake of identifying appropriate critical values in the study reported here. Specifically, pairs of item response datasets were generated that were in accord with the MIRT model adopted for the reference group, and deviance statistics were computed based on the fit of each dataset to the MIRT model posited to be optimal for the reference group. The *W*-index for each pair of datasets was computed from each corresponding pair of deviance statistics, and a frequency distribution of the *W*-index was obtained for a large number of iterations of this process. The resulting frequency distribution allowed us to determine the *W*-critical value for a particular configuration. By placing the focal group (i.e., the group for whom the MIRT model is expected to be sub-optimal), in the numerator of the fraction, it is expected to observe the *W*-index with values greater than 1.00 because the fit of the data to the specified model is expected to be worse than it is for the reference group. Thus, this allows for the adoption of one-tailed hypothesis tests. The *W*-critical value obtained from the frequency distribution of the simulated data could then be used to examine the lack of ME for the demographic groups under variations in experimental factors. Because the deviance statistic has been shown to be a viable procedure for determining model fit (Adams et al., 1997), it is hypothesized that in this study, the *W*-index, based on the deviance statistic, will accurately identify a lack of measurement equivalence as demonstrated by unsatisfactory model fit and dissimilar factorial structure across groups.

Assessment of Model Fit via *ConQuest*

This dissertation employs a piece of software entitled *ConQuest* (Wu, Adams, & Wilson, 1998) to facilitate identification of across-group model fit using the MRCMLM. The program utilizes marginal maximum likelihood to estimate γ, the matrix of regression coefficients, Σ, the variance-covariance matrix, and ξ, the item parameter vector of the MRCMLM. The following is a summary of the complete explanation of this procedure presented by the authors in the manual, *ACER ConQuest: Generalized item response modelling software* (1998):

First, the unconditional, or marginal, item response model is obtained, which is

$$f_x(x;\xi,\gamma,\Sigma) = \int_{\theta} f_x(x;\xi \mid \theta) f_\theta(\theta;\gamma,\Sigma) d\theta \qquad (8)$$

From this, the likelihood function is given by

$$\Lambda = \prod_{n=1}^{N} f_x(x_n;\xi,\gamma,\Sigma) \qquad (9)$$

where N is the total number of sampled persons.

Differentiating with respect to each of the parameters and defining the marginal posterior as

$$h_{Y}\left(\theta_{n};W_{n},\xi,\gamma,\Sigma \mid x_{n}\right)=\frac{f_{n}\left(x_{n};\xi \mid \theta_{n}\right)f_{\theta}\left(\theta_{n};W_{n},\xi,\gamma,\Sigma\right)}{f_{x}\left(x_{n};W_{n},\xi,\gamma,\Sigma\right)} \tag{10}$$

provides the following system of 3 likelihood equations:

$$A'=\sum_{n=1}^{N}\left[x_{n}-\int_{\theta_{n}} E_{x}\left(z \mid \theta_{n}\right)h\theta\left(\theta_{n};Y_{n},\xi,\gamma,\Sigma \mid x_{n}\right)d\theta_{n}\right]=0, \tag{11}$$

$$\hat{\gamma}=\left(\sum_{n=1}^{N}\bar{\theta}_{n}W'_{n}\right)\left(\sum_{n=1}^{N}W_{n}W'_{n}\right)^{-1}, \text{ and} \tag{12}$$

$$\hat{\Sigma}=\frac{1}{N}\sum_{n=1}^{N}\int_{\theta_{n}}\left(\theta_{n}-\gamma W_{n}\right)\left(\theta_{n}-\gamma W_{n}\right)'h_{\theta}\left(\theta_{n};Y_{n},\xi,\gamma,\Sigma \mid x_{n}\right)d\theta \tag{13}$$

where $E_{x}\left(z \mid \theta_{n}\right)=\Psi\left(\theta_{n},\xi\right)\Sigma z\exp\left[z'\left(b\theta_{n}+A\xi\right)\right];$ (14)

and $\bar{\theta}_{n}=\int_{\theta_{n}}\theta_{n}h_{\theta}\left(\varsigma_{n};Y_{n},\xi,\gamma,\Sigma \mid x_{n}\right)d\theta.$ (15)

This system of three equations may then be solved using an EM algorithm following the approach of Bock and Aitken (1981).

In *ConQuest,* the estimation algorithms can be either adaptations of the quadrature method described by Bock and Aitken (1981) or the Monte Carlo method of Volodin and Adams (1995). The choice of which to use is based on the number of dimensions involved. Quadrature is the default method for fewer than three dimensions; the Monte Carlo method is used otherwise. The fit of the model is ascertained by generalizations of the Wright and Masters (1982) residual-based methods that were developed by Wu (1997), using the deviance statistic. This program formally checks model fit by alternatively positing dimensionality structures and comparing the fit between the latent construct and the observed score of these nonlinear models.

Verification of Between-item Dimensionality

There is an important distinction between "within-item" and "between-item" dimensionality in MRCMLM. In order to have "between-item dimensionality" the items must have a significant loading (> 0.4) on only one factor (Wu et al., 1998). For the real data, it was necessary to verify such a condition existed. However, for this portion of the investigation, the data were simulated to meet this requirement, thus justifying the use of the between-item feature in *ConQuest.*

Simulation Study Overview

For the simulation, the computer program *SAS 8e* (2004) and *WINSTEPS* (1999) were utilized to generate multidimensional data similar to those collected for the National Board for Professional Teaching Standards, using the *Teacher Collective Responsibility Survey Instrument*—the instrument for which responses were analyzed in the real data example section of this dissertation. The instrument and cover letter are included in Appendix B. The first step in the investigation was to generate a number of item response

data sets. This was accomplished with the assistance of *SAS8e* (2004) and *WINSTEPS* (1999). (See Appendices C and D) The first group generated was that for the baseline condition. The baseline (null case) was defined to have no items lacking measurement equivalence (referred to in the following discussions as the $p = 0$ condition). That is, the factorial structure was the same for both groups of interest. Next, each data set was submitted to *ConQuest* using a correctly specified model. Here a deviance statistic was obtained. The deviance statistics from the null data sets were used to create the *W*-index value (Appendix E). *SAS 8e* was used to determine the sampling distribution and the accompanying critical value for a hypothesis test using $\alpha = .05$ for the *W*-index (Appendix F). The *W*-critical values were verified by additional null data sets generated using the same procedure. Following this, data sets were created in which there was a lack of measurement equivalence (referred to in the following conditions as the $p \neq 0$ conditions). Here the intent was to identify how often a true lack of measurement equivalence could be detected by calculating the statistical power rate for the null hypothesis of equal model fit across groups. These were fully crossed with 4 variations in sample size and 3 variations in strength of intertrait correlation. From this, an evaluation of the accuracy of the *W*-index procedure for identifying a lack of measurement equivalence in measures from a controlled situation with known parameters was made (Appendix G). For further information to aid the investigation, a logistic regression that included all interactions and main effects was also completed.

Constant Elements

In alignment with the real data, the simulated data response sets consisted of 26, four-option, Likert scale items. Additionally, the discrimination parameters (α) were constant both within and between items (i.e., we assumed that the data conformed to a Rasch model). Also, the number of rating scale categories was set to equal 4 ($k = 4$) for all items and across all remaining conditions. As another constant element, the distances between the item category thresholds (taus) were set to be equal (-1, 0, and 1). The data were generated to be multidimensional, with two dimensions. In the null condition only, where no items lack equivalence (p = 0), 13 items loaded identically on each dimension for both the focal and reference groups. In the other conditions, where some items lack equivalence ($p \neq 0$), the factor loadings for the 26 items are different for the reference and focal groups.

Data Generation Procedure

The data generation followed procedures suggested by Wherry, Naylor, Wherry, and Fallis (1965). First, a set of two randomly generated simulee traits (thetas) was created, each from a N(0,1) distribution, for each simulated response. This produced a multidimensional setting, with D = 2. The correlation between the trait distributions was varied as an experimental factor. In addition, a delta, or item difficulty parameter, was randomly generated from a N(0,1) distribution for each item. For each matched pair of simulee traits (thetas) and item difficulty (delta), an item response was calculated based on a multidimensional Rasch Rating Scale Model, which is

$$\pi_{nix} = \frac{\exp \sum_{j=0}^{x} \left(\theta_n - \delta_i - \tau_j \right)}{\sum_{k=0}^{m_i} \exp \sum_{j=0}^{k} \left(\theta_n - \delta_i - \tau_j \right)} \qquad (7)$$

where, τ_j represents the relative difficulties of the various item category thresholds that were common across all items.

The response category for each item was determined by comparing the calculated category probabilities of a given response to an item by a simulee with a number sampled at random from a U[0,1] distribution. If the sampled number was less than the calculated probability for the threshold between the first and second rating scale categories, then the item response was scored as the first category. If the sampled number was larger than this calculated probability but less than a second threshold's probability, the item response was scored as the second category, and so on. The process was completed for each simulee on each of the items.

Null Condition: P = 0

The first data configuration constitutes the null situation, in which equivalence holds across groups. These data sets define the sampling distribution for the *W-index* against which the remaining simulated data sets were compared. In these data sets, no items lacked measurement equivalence. This was established by generating data for two groups of simulees using the same factor structure for both the focal and reference groups. Here the value of p, or number of items lacking equivalence, was set equal to zero (p = 0). A separate version of the null condition was created for each cell of the experimental

design described in the following sections (i.e., for each combination of sample size and intertrait correlation). 200 null data sets were generated for each group for each cell of the experimental design, thus producing 4,800 data sets. In addition to these data sets, a separate grouping of data sets was also generated via the same procedure to verify findings from the original data sets. This consisted of 100 sets for both the reference and focal groups for each of the null conditions, resulting in an additional 2,400 data sets.

Experimental factors

Using the same procedure, additional data sets were generated in which experimental factors were varied. 50 data sets per group per cell of the experimental design were created. The factors included in the study were sample size, strength of correlation between latent traits, and number of items displaying a lack of equivalence. The values for each of these used in the study are displayed in Table 1.

Table 1. Characteristics of Variation in Experimental Factors

CHARACTERISTICS	VALUES			
Number of items lacking equivalence	$p_1 = 0$*	$p_2 = 2$	$p_3 = 4$	$p_4 = 6$
Sample size	$n_1 = 150$	$n_2 = 500$	$n_3 = 1000$	$n_4 = 2000$
Intertrait correlation	$r_1 = .20$	$r_2 = .40$	$r_3 = .60$	

**Note*: This particular condition serves as a reference condition for the sake of evaluating the Type II error rate.

These factors were fully crossed, thus producing 3,600 data sets. Subsequently, the effects of these three factors on the detection rate of the *W*-index method were examined via the simulations.

Number of items lacking equivalence.

Unfortunately, there were no specific guidelines that have been clearly identified

as to the ideal number of items displaying a lack of equivalence on a given instrument to ensure correct verification. However, based on previous research (Raju et al., 1995) and the real data, values were selected that could be expected in a survey instrument of 26 items: 2, 4, and 6 items. Taking into consideration rounding, two items is approximately 8% of the items on the full instrument and 15% of one factor. Four items is approximately 15% of the total instrument and 31% of one factor. Six items is 23% of the instrument and 46% of one factor. Again, the reference group was defined as having no items lacking equivalence or $p = 0$.

Sample size.

In the experimental design there were four levels of sample size investigated ($n_1 = 150$, $n_2 = 500$, $n_3 = 1000$, $n_4 = 2000$), with sample size held constant for both the focal and reference groups. These sample sizes were chosen to be representative of those considered in similar prior research.

Strength of intertrait correlation.

The second factor under investigation was the magnitude of the intertrait correlation. The values selected were .20, .40, and .60. As there were no specific guidelines that have been established from previous research for these, .20 and .60 were selected because they represent the range from a weak to a strong correlation; .40 was selected because it is the average intertrait correlation for the real data in this study.

Logistic Regression

Additionally, the results of the experiment were analyzed using logistic regression. In this situation, correct identification of lack of ME was the dependent variable and the

previous three experimental factors were the independent variables. Significance was determined through an examination of the Wald Chi-Square statistic, at $\alpha = .05$.

CHAPTER 4: SIMULATION RESULTS

In this chapter, the results obtained from the simulation portion of the investigation are presented.

The Null Condition

To create the null condition (p_1) in which no items lacked equivalence, the factorial structure for the focal group (group 2) was defined to be identical to that for the reference (group 1): items 1 through13 loaded on theta 1 and items 14 through 26 loaded on theta 2 for both groups. This condition was fully crossed with the four sample sizes and the three intertrait correlation values.

Descriptive Statistics

The descriptive statistics for the simulated null data sets are given in Table 2. Overall, the means for each group under all conditions were close to the value of 2.50 and were closer to that value as the sample size increased. A similar trend exists for the standard deviation, which centered around the value of 1.13. Generally, the data were slightly platykurtic (with an average around -0.80) and symmetrical (with an average value around 0.00).

The descriptive statistics for the *W*-index for the simulated null condition are given in Table 3.

W-Critical Values for Simulated Data

The critical values obtained from the frequency distribution of the *W*-index for all cells of the null condition at $\alpha = .05$ are shown in Table 4. (The complete frequency distribution output is included in Appendix H)

Table 2. Descriptive Statistics for Null Condition, Simulated Data

Intertrait Correlation	Sample Size	Group	Mean	Standard Deviation	Kurtosis	Skewness
$r = 0.2$	150	1	2.51	1.13	-0.74	-0.02
		2	2.51	1.14	-0.80	-0.02
	500	1	2.50	1.13	-0.82	-0.02
		2	2.50	1.14	-0.86	-0.00
	1000	1	2.51	1.13	-0.87	-0.01
		2	2.51	1.13	-0.86	-0.01
	2000	1	2.50	1.13	-0.87	0.01
		2	2.50	1.13	-0.87	-0.00
$r = 0.4$	150	1	2.56	1.13	-0.84	-0.06
		2	2.58	1.14	-0.87	-0.09
	500	1	2.49	1.13	-0.85	0.02
		2	2.51	1.13	-0.87	-0.01
	1000	1	2.50	1.13	-0.89	-0.01
		2	2.50	1.13	-0.89	-0.01
	2000	1	2.49	1.14	-0.89	0.03
		2	2.49	1.14	-0.88	0.02
$r = 0.6$	150	1	2.48	1.13	-0.80	0.04
		2	2.48	1.14	-0.86	0.04
	500	1	2.50	1.14	-0.89	-0.02
		2	2.51	1.14	-0.89	-0.02
	1000	1	2.50	1.13	-0.81	0.01

Table 2 (cont.)

	2	2.50	1.13	-0.79	0.01
2000	1	2.51	1.13	-0.81	-0.02
	2	2.50	1.13	-0.79	0.01

Table 3. Descriptive Statistics for *W*-index, Null Condition

Sample Size	Intertrait Correlations	Mean	Standard Deviation
150	.02	1.0022	0.013
	.04	0.9987	0.013
	.06	1.0003	0.015
500	.02	1.0001	0.007
	.04	1.0000	0.007
	.06	1.0000	0.008
1000	.02	0.9999	0.005
	.04	0.9996	0.005
	.06	0.9994	0.005
2000	.02	1.0002	0.004
	.04	0.9999	0.003
	.06	0.9999	0.004

Table 4. *W*-critical Values for Null Condition

	Intertrait Correlation		
	r = .20	r = .40	r = .60
Sample Size			
150	1.02	1.02	1.02
500	1.01	1.01	1.01
1000	1.01	1.01	1.01
2000	1.01	1.01	1.01

To insure the accuracy of these values, a verification was completed by first generating a second group of 100 data sets for both the focal and reference groups, (2,400 data sets) and then making use of the critical values acquired from the first set. The Type I Error Rates from the second simulated data sets are shown in Table 5.

Table 5. Type I Error Rates for Second Simulated Null Data Sets

Intertrait Correlation		.20	.40	.60
	Reject	Frequency	Frequency	Frequency
Sample Size				
150	0	0.94	0.95	0.95
	1	0.06	0.05	0.05
500	0	0.96	0.95	0.94
	1	0.04	0.05	0.06
1000	0	0.95	0.94	0.95
	1	0.05	0.06	0.05
2000	0	0.96	0.95	0.96
	1	0.04	0.05	0.04

Accurate Identification of Lack of Measurement Equivalence: Statistical Power

The critical values shown in Appendix H were used to evaluate the rate at which the *W*-index correctly rejected a false null hypothesis (statistical power) for each cell of the experimental design utilized in the simulation. This power rate for each condition is given in Table 6.

Table 6. Statistical Power* of *W*-index

p	2			4			6		
r	.20	.40	.60	.20	.40	.60	.20	.40	.60
n									
150	.12	.10	.16	.24	.16	.06	.28	.18	.10
500	.16	.16	.12	.38	.28	.08	.22	.20	.12
1000	.52	.26	.12	.62	.38	.26	.68	.30	.22
2000	1.00	.92	.60	.90	.92	.60	1.00	.90	.64

* Power is the proportion of cases for which an accurate identification of lack of equivalence is made.

p = number of items lacking equivalence

r = intertrait correlation

n = sample size

The power rates, or proportion of cases for which an accurate identification of lack of equivalence was made, range from a low of .06 to a high of 1.00. Generally the rates are smallest with small sample size and large intertrait correlation. The trend is for power to be larger with larger sample size and with smaller intertrait correlation.

Results from Logistic Regression

Interactions

First, using *SAS 8e,* a logistic regression was completed that included the three-way interactions (Appendix I). Initial analysis of the univariate relationships between the experimental factors and statistical power indicated that sample size exhibits a quadratic influence on statistical power, so two three-way interactions were examined—one between number of items lacking equivalence, intertrait correlation, and sample size and the other between number of items lacking equivalence, intertrait correlation, and the square of the sample size. The results showed that neither of these three-way interactions was statistically significant (Appendix J). Next, a simpler model that excluded the three-way interactions but included all two-way interactions (with both linear and quadratic trends for the sample size factor) was fit to the data. This model revealed that neither the intertrait correlation-by-sample size squared term nor the number of items-by-intertrait correlation term contributed to the model, so those terms were removed (Appendix J). The reduced model contained two statistically significant two-way interactions. The results are given in Table 7.

The first statistically significant two-way interaction was between sample size (n) and intertrait correlation (r) ($\chi^2_{Wald} = 22.21$, $p < .0001$). Table 8 displays a two-way table summarizing the power rates for the sample size-by-intertrait correlation interaction. These power rates are also depicted in Figure 1. The results indicate a similar overall trend for the lower two intertrait correlations of .20 and .40 across sample sizes. This differs slightly from the higher intertrait correlation rate of .60.

Table 7. Logistic Regression Results - Two-Way Interactions

Parameter	DF	Estimate	Standard Error	Wald Chi-Squared	Pr > ChiSq
Intercept	1	0.77	0.46	2.87	.09
p	1	0.14	0.09	2.42	.12
r	1	1.06	0.67	2.49	.11
n	1	0.00	0.00	0.65	.42
n^2	1	0.00	0.00	24.09	<.0001
n*p	1	0.00	0.00	6.54	.01
n^2*p	1	0.00	0.00	8.01	.005
n*r	1	0.00	0.00	22.21	<.0001

p = number of items lacking equivalence

r = intertrait correlation

n = sample size

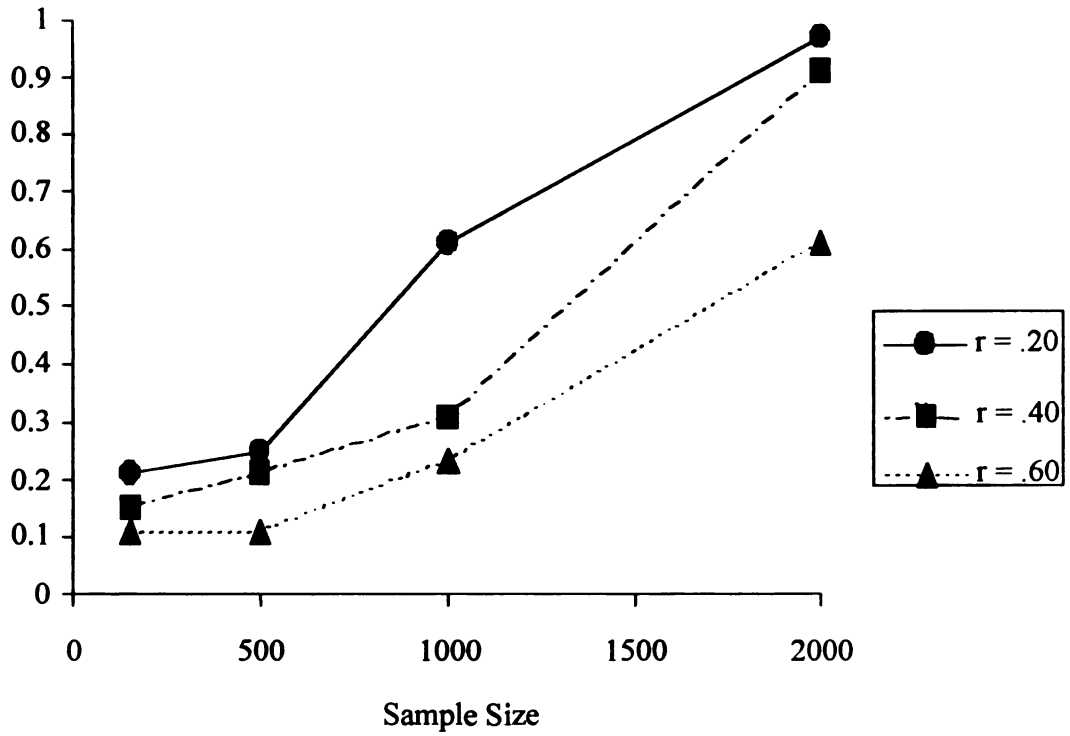
Table 8. Power of *W*-Index for the Sample Size-by-Intertrait Correlation Interaction

n/r	.20	.40	.60
150	0.21	0.15	0.11
500	0.25	0.21	0.11
1000	0.61	0.31	0.23
2000	0.97	0.91	0.61

r = intertrait correlation

n = sample size

Figure 1. Two-way Interaction of Sample Size and Intertrait Correlation on Power



The graph also suggests a possible sigmoid relationship between the sample size and the intertrait correlation with respect to statistical power. However, the trend seems slight within the range of sample sizes considered in this study, so this term was subsequently dropped from the model.

The second statistically significant two-way interaction included the quadratic trend between sample size (n^2) and the number of items exhibiting lack of ME (p) ($\chi^2_{Wald} = 8.01$, $p = .005$). Table 9 displays a two-way table summarizing the power rates for the sample size-by-number of items lacking equivalence interaction, also depicted in Figure 2.

Table 9. Statistical Power of *W*-Index for the Number of Items Lacking Equivalence-by-Sample Size Interaction

n/p	2	4	6
150	0.13	0.15	0.19
500	0.15	0.25	0.18
1000	0.30	0.42	0.40
2000	0.84	0.81	0.85

p = number of items lacking equivalence

n = sample size

Figure 2. Two-way Interaction of Number of Items Lacking Equivalence and Sample Size on Statistical Power

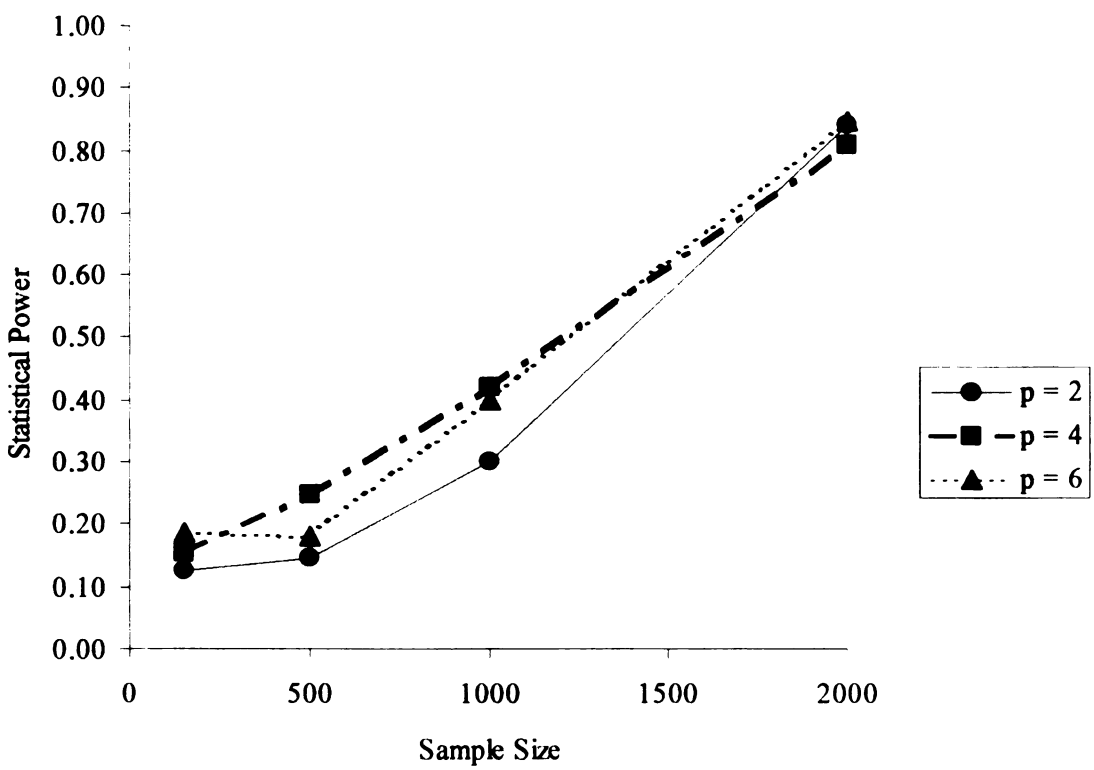


Figure 2 shows for smaller sample size, the increase is steepest for 4 items lacking equivalence. For larger sample size, the rate of increase is slightly more steep when 2 items lack equivalence. However, except for the decrease in rate for $p = 6$, n = 500, the rate of acceleration in power across sample size is very similar for all values of number of items lacking equivalence. In fact, over the range of sample sizes that are typically recommended for use with complex IRT models (> 1000), the variation is slight, and the trend seems to be nearly linear. Hence, this interaction term was dropped from the model.

Main Effects

The final model was fit to these data for the sake of directly evaluating three of the research hypotheses. The results of fitting the data to a main effects model (which included a quadratic term for sample size) are shown in Table 10. These results are discussed in the following three subsections.

Table 10. Logistic Regression Results – Main Effects

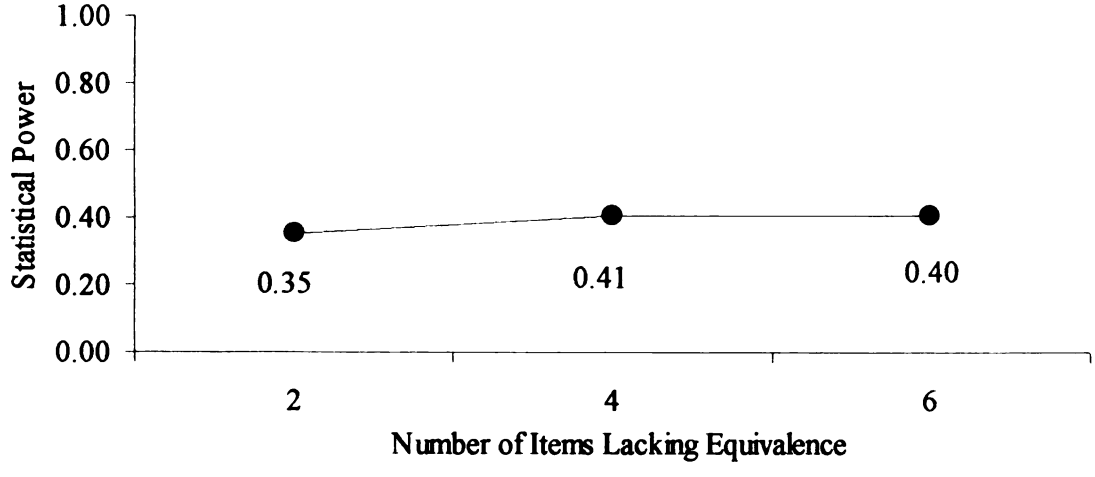
Parameter	DF	Estimate	Standard Error	Wald Chi-Squared	Pr > ChiSq
Intercept	1	0.47	0.24	3.77	.05
p	1	-0.02	0.03	0.30	.58
r	1	3.75	0.39	93.25	<.0001
n	1	0.00	0.00	1.95	.16
n^2	1	0.00	0.00	22.53	<.0001

p = number of items lacking equivalence
r = intertrait correlation
n = sample size

Variation in number of items lacking equivalence.

For the number of items lacking equivalence, the results show that as this number increased, statistical power did not tend to increase by much. In fact, the effect is not statistically significant ($\chi^2_{Wald} = 0.30$, $p = .58$). The power increased only slightly between the first two levels of this factor and not at all between the second two levels—specifically, the average statistical power equals .35, .41, and .40 for 2, 4, and 6 items lacking measurement equivalence, respectively, as shown in Figure 3. Additionally, the results, as displayed in Appendix K, show that when the sample size and intertrait correlation were held constant, statistical power increased for 14 of the 24 cases (58%). There were 3 cases (~13%) in which the power stayed the same as the number of items lacking equivalence increased. In 7 cases (~29%), there was a decrease in power associated with an increase in number of items lacking equivalence.

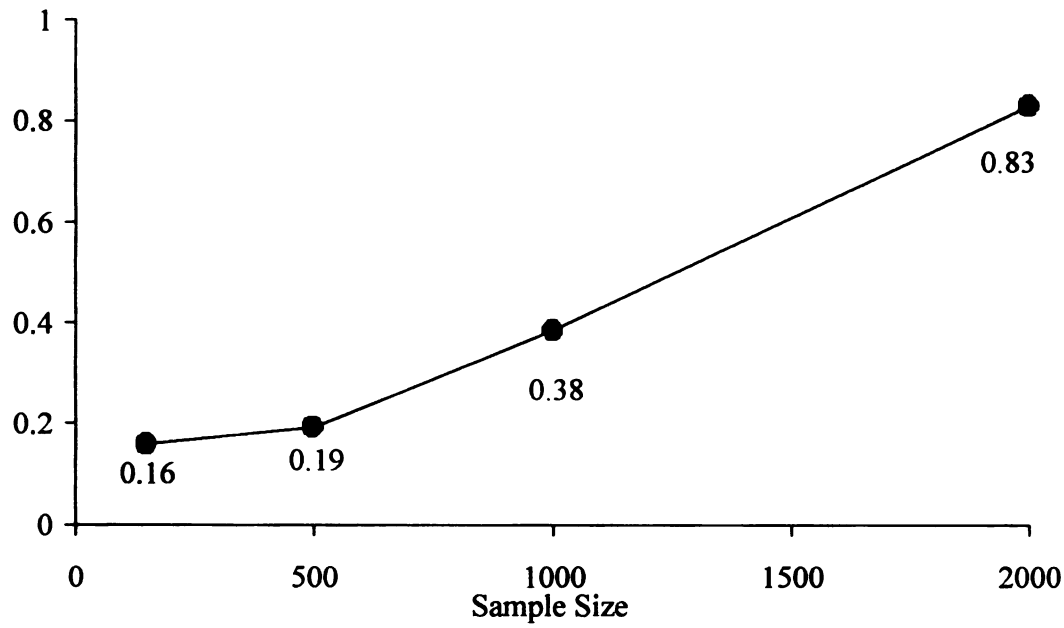
Figure 3. Main Effect for Number of Items Lacking Equivalence



Variation in sample size.

The results show that the increase in sample size over intertrait correlation and number of items lacking equivalence resulted in a quadratic increase in power. This outcome is statistically significant, ($\chi^2_{Wald} = 22.53$, $p < .0001$). Specifically, the average statistical power for sample sizes of 150, 500, 1000, and 2000 equal .16, .19, .38, and .83, respectively (as shown in Figure 4).

Figure 4. Main Effect for Sample Size

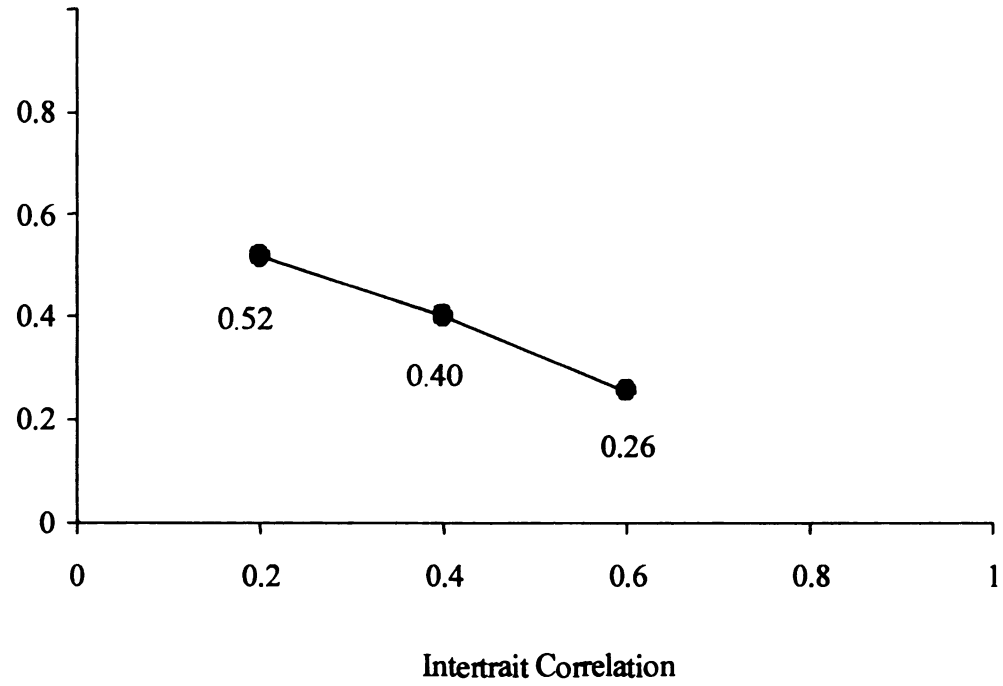


In this study, the power increased 93% of the time (25 out of 27 cases) (Appendix L) with an increase in sample size. Specifically, the largest values for power where obtained when the sample size was 2000, where the maximum value was 1.00. Power decreased markedly as the sample size decreased, to a minimum of .06, when the sample size was 150.

Variation in intertrait correlation.

With regard to changes in intertrait correlation, the results show a strong inverse relationship with statistical power. This outcome was also statistically significant, ($\chi^2_{Wald} = 93.25$, $p < .0001$). The average power rate across all levels of the remaining factors for intertrait correlations equal to .20, .40, and .60 were .52, .40, .26, respectively (as shown in Figure 5).

Figure 5. Main Effect for Intertrait Correlation



Approximately 92% of the time, as the intertrait correlation increased, the value for power decreased (22 out of 24 cases) over all variations in sample size and number of items lacking equivalence. In all cases but one, the largest accurate identification rates for a given number of items lacking equivalence, across sample size, were those for r = .20 and decreased substantially as the intertrait correlation increased.

CHAPTER 5: DISCUSSION OF SIMULATION RESULTS

In this chapter, results from the simulation phase of the investigation are discussed.

Rates of Statistical Power

As there were no well-established guidelines for accurate identification rates for lack of measurement equivalence, those used were based on the prior research of Flowers, Raju, and Oshima (2002), which also involved statistical power. They were:

Unacceptable: Power < 0.2

Marginally acceptable: $0.2 \leq$ Power < 0.4

Acceptable: $0.4 \leq$ Power < 0.6

Good: $0.6 \leq$ Power

Effects of Variation in Experimental Factors on Accuracy Rate

Variation in Number of Items Lacking Equivalence

The relevant research question being addressed is:

> ***Is the accuracy of the W-index of measurement equivalence using factorial structure equality affected by variations in the number of items lacking equivalence?***

The findings from this investigation show that as the number of items lacking equivalence increased over all values of sample size and intertrait correlation, statistical power also increased in 14 out of 24 cases (Appendix K). This is to say that 58% of the time, increasing the number of items lacking equivalence resulted in a higher power rate; 42% of the time it did not. In other words, in this study, increasing the number of items lacking equivalence did not consistently increase power significantly (Figure 3). Thus, in this study, a variation in the number of items lacking equivalence did not consistently

result in a corresponding change in statistical power. Additionally, a smaller number of items lacking equivalence did not automatically result in small statistical power. For example, "Good" identification was made when only 2 items (8% of the total) lacked equivalence across intertrait correlation when the sample size was 2000.

Although somewhat unexpected, these findings are not out of line with those from other current ME investigations (Furlow & Fouladi, 2005; Meade, Ellington, & Graig, 2004) where it was also found that the number of deviant items did not have the expected effect. There may be a plausible reason for this finding. Consider for a moment the variance/covariance matrix that contains $\frac{i^2 - i}{2}$ items in the off-diagonals. In this study with 26 items, this amounts to 325 elements in the off-diagonals. For each single item that lacks equivalence (3.8% of the total items), the lack of fit for the one item affects 25 entries in the covariance matrix, computed as $\left(\sum_{1}^{n}(i-n)\right)$, where n = number of items lacking equivalence. Thus, there is lack of fit for 7.7% of the elements in the covariance matrix (25/325). For 2 items lacking equivalence (7.7% of the total items), 15.1% of the interitem covariances (49/325) are effected. For 4 items, (15.4% of the total items), 94 items in the covariance matrix or 28.9% are effected. Having 6 items (23.1% of the total items) that lack equivalence would affect 41.5% of the matrix elements (135/325). This constitutes a considerable amount of misfit. In fact, although the largest number of items lacking equivalence considered in the study made up only 23% of the total on the instrument, their lack of fit to the model affected the fit of almost half of the items in the covariance matrix. However, if the test contained more items, the effect would be greatly reduced. Say, for example, the test contained 100 items. With i = 100, there are 4,950

elements in the off-diagonals. For the same number of items lacking equivalence (2 or 2% of the total items), only 194 or 4% of the matrix elements would be affected, which would, undoubtedly, yield quite different results, as the same number of items resulted in a much smaller percentage of misfit. Consequently, a great deal less misfit would result in a smaller deviance statistic, which would result in a *W*-index closer 1, which would result in a failure to reject the null hypothesis. Thus, the failure to see a consistent effect on the statistical power of the *W*-index connected to the number of items lacking equivalence in this particular study may well be a result of over sensitivity of the index as a result of small number of items on the test. Most fortunately for the procedure, acceptable rates were still achieved across the number of items lacking equivalence when other criteria, such as a large sample size and a small intertrait correlation, were met.

Variation in Sample Size

The results of variation in sample size (Appendix L) support the conclusion that, generally, a large sample size will result in a high rate of correct identification of lack of measurement equivalence, with other factors being the same. Specifically, as hypothesized, the largest sample size (n = 2000) yielded results in the highest category of "Good" across the board. Rates were also "Good" for samples sizes of 1000, if the intertrait correlation was .20. For the smallest samples size of 150 all the other rates were "Unacceptable" except in two situations where the intertrait correlations was .20. Here the rates were "Marginally acceptable."

With these results, we were now able to address the second research question:

Is the accuracy of the W-index of measurement equivalence using factorial structure equality affected by variations in sample size?

In this study, variations in sample size were shown to affect the accuracy of the *W*-index in identifying a lack of measurement equivalence, with larger sample sizes being associated with higher accuracy, as reflected by a measure of power or percentage of times a correct identification of lack of equivalence was made. Specifically, a sample size of 2000 yielded "Good" results in all situations, while all of the identification rates from sample sizes of 150 were, at best, "Marginally acceptable" varying from a low of 6% to a high of only 28%.

These results were consistent with other IRT studies that revealed identification of a lack of ME was not as accurate with a small sample size of 150 (Hidalgo-Montesinos and Lopez-Pina , 2002) and more accurate with large sample sizes (De Champlain et al.,1998; De Champlain & Gessaroli,1998; Meade et al., 2004). Specifically, the sample size supported most strongly by this study for "Good" results was n = 2000. "Acceptable" rates were obtained for n = 1000 if the intertrait correlation was maximally .20.

Variation in Intertrait Correlation

The third research question is:

> *Is the accuracy of the W-index of measurement equivalence using factorial structure equality affected by variations in the intertrait correlation?*

The findings are that variations in the strength of the intertrait correlation do affect the accuracy of the *W*-index method. In this study, a smaller intertrait correlation resulted in more accurate identification of lack of equivalence in 92% of the cases, across samples size and number of items lacking equivalence. Additionally, the strength of the intertrait correlation has a strong inverse relationship with accurate identification of ME:

as the intertrait correlation increases, statistical power decreases (Appendix M). These results, also, are in line with the research hypothesis that the accuracy of the method would be lower when the intertrait correlation was higher. Specifically, the rates were acceptable for all cases where $r = .20$ and the sample size was 1000 or greater. For intertrait correlations of both .40 and .60, a minimum sample size of 2000 is needed to achieve a "Good" rate.

Although a great deal of prior research involves unidimensional data, the findings from this specific multidimensional investigation were in line with others, such as that completed by van Abswoude et al. (2004), who also concluded that larger intertrait correlation was associated with less accurate identification of lack of measurement equivalence.

The Effects of the Two-Way Interactions

The statistically significant two-way interactions in this study were sample (1) size-by-intertrait correlation and (2) number of items lacking equivalence-by-squared sample size. Even though the effects of both were slight, they do have implications that should be recognized. First, based on the results from this study, an increase in sample size alone, without considering the intertrait correlation, may not guarantee the results desired. For example, when the sample size is smallest, increasing only the sample size from that of n = 150 to the next larger size of 500 increases the rate but does not move the statistical power into the "Acceptable" category for all cases, nor does increasing just the sample size to an even larger value of 1000. In order to reach the "Acceptable" category, an intertrait correlation of .20 is also required. This illustrates the

effect of the two-way interaction identified between sample size and intertrait correlation. Hence, it may be deduced that although a large sample size is desirable, it alone does not guarantee maximum results. It is recommended for best results that the strength of the intertrait correlation also be considered.

Similarly, the second two-way interaction between the squared sample size and the number of items lacking equivalence also supports the findings that a large number of items lacking equivalence by itself is insufficient to achieve "Good" identification of lack of ME. For example, when there are 6 items (23%) lacking equivalence, if the sample size is 150 or 500, power is only .19 and .18, respectively. However, for the same percentage of items lacking equivalence, if the samples size is increased to 2000, the value for power is increased to .85. Thus, for maximum results, a large number of items lacking equivalence needs to be coupled with a large sample size.

Summary

Taken in totality, the results from this investigation provide an answer to this investigation's overarching research question, which is

> *Can the W-index method using factorial structure equality accurately identify a lack of measurement equivalence in a survey instrument?*

Supporting the hypothesis that the *W*-index would accurately identify a lack of ME in measures from a survey instrument, the answer to this most important question is a qualified "yes, it can," in certain situations. In this study, results in the "Good" category were obtained with the largest sample size of 2000 for all values of intertrait correlation and number of items lacking equivalence. Additionally, "Acceptable" results were obtained for n = 1000, if the intertrait correlation was kept at .20. Conversely, no results

in the "Acceptable" category were found when the sample size was 150, regardless of the other factors. This is in line with prior research that also found a small sample size to yield unacceptable results and a large same size to be advantageous.

As an additional qualifier to the use of the *W*-index, if attempts are made to increase statistical power by increasing sample size, it is recommended that the requirement of weak intertrait correlation (.20 or less) not be overlooked. Also, this study found that, contrary to what was expected, a large number of items lacking equivalence is not an assumption that must be met for accurate identification of lack of ME when using the *W*-index procedure.

CHAPTER 6: REAL DATA METHODOLOGY

This chapter presents the second phase in the investigation, which is a demonstration of the use of the *W*-index method to identify a lack of measurement equivalence by applying it to real data measures. The source for the real data is a study conducted through the National Board for Professional Teaching Standards using the *Teacher Collective Responsibility Survey Instrument* (Appendix B). The statistical tests, and measurement models, as well as some of the computer software, used for the real data portion of the study are analogous to those used for the simulation.

Survey Instrument

Instrumentation

The instrument is composed of 26, four-option, Likert-scale items. Approximately 180 items covering the aspects of the Developmental Model (Figure 6) were originally generated for the instrument developed by the author. A review of these items was completed by four, full-time college professors at a Land Grant, research-extensive university in the United States. Although from various departments, all the reviewers were within the College of Education and all were involved in research concerning "Teacher collective responsibility for student learning." As a result of suggestions made by the review team, appropriate modification and deletions were made to the instrument. The resulting final item distribution by item number blueprint for the instrument is given in Table 11. There were some additional demographic questions on the original instrument not included in this study.

Figure 6. Developmental Model

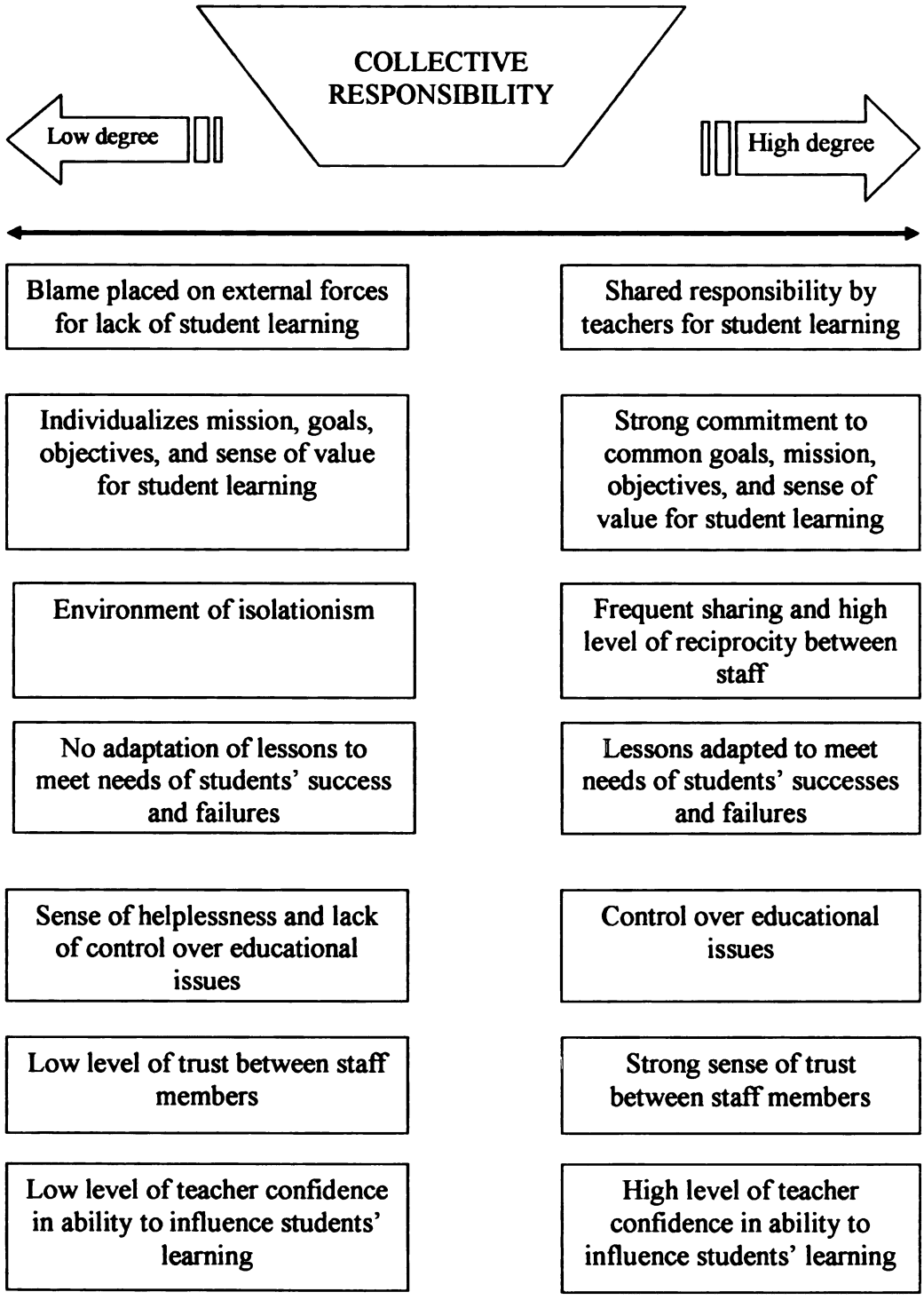


Table11. Instrument Blueprint

	COMPONENTS				
	Quad I Reporter: School	Quad II Reporter: Classroom	Quad III Identifier: Classroom	Quad IV Identifier: School	Total in Category
	Item Number				
1. Shared responsibility by teachers for student learning	7	1	21	14	4
2. Lesson adaptation	8	X	3	16	3
3. Teacher confidence in ability to influence students' learning	9	20	22	15	4
4. Commitment to common mission, goals, objectives and sense of value for student learning	10	24	23	18	4
5. Sharing, and reciprocity between staff	11	4	6	X	3
6. Sense of trust between staff members	12	5	25	17	4
7. Control over learning environment	13	2	26	19	4
Total	7	6	7	6	26

The cover letter for the instrument (Appendix B) assured the participants that their participation was entirely voluntary, their responses kept confidential, and that they could withdraw at any time.

Population

The target population in this study was U.S. public school teachers in grades PreK - 12. For clarification, "teachers" included all full and part-time classroom instructors, as well as other non-administrative personnel who routinely interact with students, such as counselors, media specialists, speech therapists, classroom consultants, and others. The sample population for this study consisted of 616 teachers in seven mid-western U.S. school districts. There were 407 females (66%) and 209 males (34%). Individual respondents were not identified. The school districts varied in size, with the largest PreK-12 student population being 38,139 and the smallest 1,387. The percentage of disadvantaged students in the districts ranged from a high of 50.3% to a low of 9%.

The demographic groups selected for this study were classified by grade level taught: secondary or elementary. Secondary was defined as grades 9 through 12 and elementary as pre-kindergarten through 8. The study included 370 secondary (60%) and 246 elementary teachers (40%).

Data Collection

Obtaining the data for the NBPTS was a two-step process. First, permission to administer the survey was granted by the Superintendent and/or the Board of Education in seven districts. Additionally, building administrators were contacted at individual schools within those districts. Secondly, at a routinely scheduled faculty meeting, the survey was introduced and distributed by the author, with typical completion taking 10 to 15 minutes.

As was expected, the response rate from this type of administration was high. This resulted in 616 usable surveys.

Data Analysis of the Survey Instrument

Prior to its use in this study, a data analysis was complete on this instrument to verify the rating scale. Values for the item parameter were obtained using *WINSTEPS* (1999) and *SAS 8e* (2004). For this analysis, the following aspects of the survey instrument were investigated: dimensionality, reliability, fit indices, and rating scales.

First, using *SAS 8e*, an exploratory factor analysis (EFA) was performed, resulting in the identification of four underlying factors for the instrument. Table 12 displays the correlation between the factors, which range from a low of .25 to a high of .46.

Table 12. Factor Correlations

	Inter-Factor Correlation			
	Factor1	Factor2	Factor3	Factor4
Factor1	1.00	.46	.28	.30
Factor2	.46	1.00	.31	.25
Factor3	.28	.31	1.00	.36
Factor4	.30	.25	.36	1.00

Essential unidimensionality for each of the four factors was determined by an additional investigation using the eigenvalue criteria and the scree plot. Based on this, further analysis was completed separately for each of the four subscales. A summary of the results from the total analysis of the separate scales is displayed in Table 13.

Table 13. Rating Scale Analysis Summary Statistics

	FACTOR				
	1	2	3	4	26-item instrument
Reliability - Standardized	0.87	0.89	0.78	0.81	0.93
Reliability - Raw Score	0.86	0.89	0.77	0.81	0.92
$Z(MS_{unweighted})$					
Person	0.99	0.99	1.02	1.00	1.01
Item	0.99	0.96	0.99	0.95	0.99
Linacre's Criteria					
Sample Size	Y	Y	Y	Y	
Unimodal	Y	Y	Y	Y	
Measure(θ)	Y	Y	Y	Y	
$MS_{unweighted}$	Y	Y	Y	Y	
τ's increase	Y	Y	Y	Y	
τ's distance	Y	Y	N	Y	
$Coherence_{Measures}$	Y	Y	N	Y	
$Coherence_{Categories}$	N	N	N	N	

Reliability was evaluated using *Cronbach's coefficient alpha* (internal consistency):

$$\alpha = \frac{k}{k-1}\left[1 - \frac{\Sigma S^2{}_i}{S^2{}_{TOTAL}}\right] \qquad (17)$$

where k = number of items in a scale, S^2_i = squared standard deviation for all items, and S^2_{TOTAL} = square of the standard deviation of the total scores for all examinees tested. This resulted in standardized reliability coefficients ranging from 0.78 to 0.89 across the subscales, and 0.93 for the total instrument.

To evaluate fit, the standardized unweighted mean-square statistic was calculated, for items in each subscale:

$$Z_{MS-unweighted} = \frac{\sum_{i=1}^{I} z^2_{ni}}{I} \qquad (18)$$

where z^2_{ni} is the square of the standardized residual for the response of person *n* to item *i*. The standardized unweighted mean-square statistic was also obtained for persons as well as items. The mean-square statistic depicts the degree to which observed ratings are in accord with those predicted by the measurement model. Numerous large residuals typically indicate that the measurement model does not sufficiently explain the observations. An examination using this fit index indicated no misfitting items. However, for the person fit statistics, 81 out of 616 (13.1%) had standardized unweighted mean squares greater than 2.0. Most displayed an "extreme checker" pattern of answering 1,4,1,4, etc. This lead to the conclusion that the questions were answered with disregard to the wording of the item, which did not warrant changes to the instrument.

The rating scale analysis developed by Linacre (2002) provided additional information about the degree to which respondents utilized the response scale in the manner that was intended. Each of the eight Linarce requirements was applied to each of

the four instrument subscales. First, the frequency distribution for each subscale showed that each category had a minimum of 10 observations. It also supported a unimodal progressive increase and/or decrease in the frequency with which each ordered rating category was chosen. The average respondent measure ($M_{(\theta n)}$) associated with each category measure was also examined. To meet Linacre's requirement, these averages should increase with the values of the rating scale categories. Next, the value of the unweighted mean squared fit statistic, evaluating the similarity of the observed to expected ratings, was examined to verify values less than 2.0. The category thresholds (τ's) were examined because the values of these indices should increase with the values of the rating scale categories. Additionally, adjacent category thresholds were examined to verify they were at least 1.4 logits apart and no more than 5 logits apart. The final item examined was the coherence statistics, both for the ratings and for the measure. In both cases, the values should be greater than 39%.

The results of the analysis were that, except for the coherence, each of the subscales met all eight of the requirements sufficiently well. Thus, based on the results, it was concluded that the items satisfied the Linacre rating scale requirements enough to deduce teachers employed the rating structure in the manner the author intended. In other words, the data analysis verified the rating scale of the instrument.

Model Selection

The model selected for this investigation was the MRCMLM (the Multidimensional Random Coefficients Multinomial Logit Model). As stated previously, this was selected because of its appropriatness to this real survey data, which is known to be multidimensional. Additionally, the MRCMLM does not necessitate a large sample

size. The sample size for the real data example is 616. As a third reason, Adams et al., (1997) demonstrated the MRCMLM was a mathematically tractable and flexible multidimensional model that produces parameter estimates that are readily interpretable. Fourth, it draws on the relationship between the latent dimensions to produce more accurate parameter estimates and individual measurements. Last, and most importantly, as an adaptation of an IRT method, the model does not necessitate meeting the normality assumption.

Verification of Between-Item Dimensionality

As was noted in the stimulation portion of the investigation, when using the MRCMLM, there is an important distinction between "within-item" and "between-item" dimensionality. In situations where between-item dimensionality exists, the items have a significant loading (> 0.4) on one factor, but may have non-significant loadings on one or more additional factors (Wu et al., 1998). An analogous term that might be more common from exploratory factor analysis (EFA) is "simple structure." Where "within-item dimensionality" exists, the items have significant loadings on more than one factor. To determine which of these situations existed, an exploratory factor analyses was performed on the survey instrument, using *SAS 8e* (2004). This identified four underlying factors for the instrument (Appendix N). That is, at a value of .4, each item loaded on only one factor. However, each also had non-zero but non-significant loadings on other factors. Because of this, to establish the dimensionality, additional investigations of these four factors were completed using the eigenvalue criteria and the scree plot (Appendix O). These tests supported the initial findings of essential unidimensionality for each factor. Therefore, the use of the between-item MRCMLM was justified for use with the real data.

Determination of Model Fit

The procedure followed for the real data study was in line with that developed for the simulation study. For each group, the index of model fit, or deviance statistic, was determined, using *Conquest.* Next, the proportionality constant was computed for each group, as defined by

$$PC = \frac{deviance}{(n-p)} \tag{5}$$

where n = sample size; p = number of parameters estimated

Then, the *W*-index was computed, again using *SAS 8e.* It is defined as the PC value for the focal group compared to that for the reference group as a ratio, or:

$$W = \frac{PC_{focal}}{PC_{reference}} \tag{6}$$

Again, this ratio of the PC for the focal group to the PC for the reference group creates the *W*-index used to test the null hypothesis, which is

Ho: *W*-index = 1

That is, there is no statistically significant difference in the fit of the model across groups of interest. If the null hypothesis is rejected, there is lack of measurement equivalence across the demographic groups. If we fail to reject, the conclusion is that no lack of measurement equivalence is detected and, consequently, the two groups are interpreting the construct of interest in the same manner. As in the simulation, a one-tailed hypothesis

test was used in this situation because the focal group (i.e., the group for whom the MIRT model is expected to be sub-optimal) was placed in the numerator of the fraction. Thus, it is expected that the *W*-index will have values greater than 1.00 because the fit of the data to the specified sub-optimal model is expected to be worse than it is for the reference group.

Identification of Critical Values

The procedure to determine the critical values for the real data was also closely aligned with that for the simulated data. First, multiple "simulated real data" data sets were generated. 100 data sets for the focal group and 100 for the reference were created, using *Matlab*, with the data having the same factorial structure, intertrait correlation and sample size as the real data. That is, there were 26, four-response, Likert-scale items in the data set. Like the real data, these data sets had four factors. Items 1, 2, 4, 5, 20, and 24 loaded on Factor 1; items 7 through 13 loaded on Factor 2; items 3, 6, 21, 22, 23, 25, and 26 on Factor 3; and 14 through 19 on Factor 4. Items included in Factor 1 are those in which the teacher acts as a reporter within the school as a whole. For those in Factor 2, the teacher again is asked to act as a reporter but within the individual classroom. For Factor 3, the questions ask the teacher to be an identifier of what is seen or perceived in the classroom of others. Finally, Factor 4 items ask the teacher to identify the collective responsibility through the entire school.

The variance/covariance matrix and item means used by *Matlab* to generate the data sets for the reference group were identical to those obtained from the real data for secondary teachers. To represent the same factorial structure for the focal group, thus

creating a null condition, an identical variance/covariance matrix was used. However, to create similarity to the real data, the means from the elementary teachers was used for the *Matlab* data set generation of the focal group data. In accordance with the real data, the sample size for the simulated real data reference group null data set was 370; the focal had a sample size of 246.

Following the format of the simulation phase of the investigation, the computer program *ConQuest* was used to obtain the deviance statistic for each pair of data sets from the demographic groups. *SAS 8e* was then used to obtain the PC and *W*-index for each. Again, as was done in the simulation, using *SAS 8e,* a frequently distribution of these *W*-index values for the simulated real data was obtained and the *W*-critical value identified at $\alpha = 0.05$ (one-tailed test). The *W*-critical value obtained from the frequency distribution of the simulated real data was then used to examine the lack of ME for the demographic groups in the real data.

Exploratory Factor Analysis

As an additional examination of the similarity or differences in the factorial structure of the data for each of the demographic groups, a separate Exploratory Factor Analysis (EFA) was conducted for each. For this, the Promax Rotated Factor Pattern was used because the factors are correlated (Appendix P).

CHAPTER 7: REAL DATA RESULTS

This chapter presents the results of the application of the *W*-index to identify a lack of measurement equivalence in real survey data.

Descriptive Statistics

The descriptive statistics, as well as the deviance statistic, for each of the demographic groups in the real and simulated real data, identified by grade level taught, are displayed in Table 14.

Table 14. Descriptive Statistics for Real and Simulated Real Demographic Groups

	Elementary		Secondary	
	Real	Simulated Real	Real	Simulated Real
Sample Size	246	246	370	370
Mean	3.20	3.15	2.90	2.91
Standard Deviation	0.36	0.34	0.34	0.33
Kurtosis	-0.66	-0.59	0.54	0.38
Skewness	-0.04	-0.05	0.48	0.45
Deviance	11865.79	12468.96	17867.58	18674.40
Deviance/df	57.05	59.95	53.82	56.25

One thing that should be noted from Table 14 is that the difference between the means of the elementary and secondary groups in the real data are farther apart than the means of the same groups in the simulated real data. This is quite probably due to the adoption of the real item difficulty parameter estimates for both groups. Also, the

difference in the deviance statistic used to compute the *W*-index between the elementary and secondary groups is quite large for both the real and simulated real data.

W-critical Value from Simulated Real Data

From the frequency distribution of the *W*-index values from the simulated real data, the *W*-critical value identified at $\alpha = 0.05$ (one-tailed test) was determined to be 1.04. The complete frequency distribution is included in Appendix Q. This critical value was then used with the real data to examine the lack of equivalence across the demographic groups of interest. The *W*-index and conclusion to reject are given in Table 15.

Table 15. *W*-index and Rejection Conclusion

Group	*W*-index	*W*-critical value	Conclusion
Elementary/Secondary	1.06	1.04	Reject Null Hypothesis

Dissimilarity in Factor Loadings

The *SAS* output obtained from the exploratory factor analysis (EFA) that displays the factor loadings for the elementary group is included in its entirely in Appendix R and that for the secondary group in Appendix S. A summary of the findings is shown in Table 16.

The results of the EFA show that Factor 2 (Reporter in School) has the most loadings in common for the two demographic groups: all of items 7 through 13 load on the same factor for both groups. Factor 3 (Identifier in Classroom) also has similar loadings for both groups for all but one item. Items 21, 22, 23, 25, and 26 load on the same factor.

Table 16. Factor Loadings for Elementary and Secondary Real Data

	Elementary	Secondary
FACTOR	Item Number	
1 - Reporter in Classroom	4, 5, 24	1, 2, 4, 5, 20
2 - Reporter in School	7, 8, 9, 10, 11, 12, 13	7, 8, 9, 10, 11, 12, 13
3 - Identifier in Classroom	20, 21, 22, 23, 25, 26	21, 22, 23, 25, 26
4 - Identifier in School	15, 16, 17	15, 16, 17, 19
5	1, 2, 3,	5, 18
6	6, 14, 19	14, 24 (neg), 25(neg)

Only item 20 does not match; it loads on Factor 3 (Identifier in Classroom) for the elementary but not the secondary group. Factor 4 (Identifier in School) is almost identical to Factor 3: items 15 through 17 load on it for both groups, and item 19 loads for secondary only. The loadings for Factor 1 (Reporter in Classroom) are less consistent. Items 4 and 5 load for both groups. However, items 1, 2 and 20 load for the secondary, while item 24 loads for elementary.

It should be noted in the output from the EFAs that there are two additional factors, Factors 5 and 6, and there are some items for each group that loaded on these. These were not included in the original factor configuration because they did not meet Stevens' (1966) criteria for "reliably defined." However, here their presence points out an obvious difference in the factorial structure between the elementary and secondary

groups. Elementary has 3 strong loadings on both Factor 5 and Factor 6. The loadings for secondary on Factor 5 are weaker and there are only 2 of them. On Factor 6, secondary has only 1 positive and 2 negative loadings. For Factor 5, there are no common loadings. Items 5 and 18 load for the secondary and items 1through 3 load for the elementary. Factor 6 does have one common item: 14. Additionally, items 6 and 19 load for the elementary while items 24 and 25 load negatively for the secondary. Thus, the results from the EFA show clearly that the factorial structures are not the same for the elementary and secondary groups. In order words, the results of the *W*-index procedure that identified a lack of ME for the measures obtain with this instrument are supported by the observable difference in the factorial structure identified through EFA.

There is an additional difference in the factorial structure between the two demographic groups to be noted from the EFA output. For the elementary group, there are no items that have a significant loading (> .4) on more than one factor. Thus, the elementary group exhibits simple structure. On the other hand, secondary does not. It has more non-zero loadings to accompany a few cross loadings.

CHAPTER 8: REAL DATA DISCUSSION

In this chapter, results from the real data phase of the investigation are discussed.

The application of the critical value obtained through the simulation to the real data measures resulted in a rejection of the null hypothesis that the fit of the data to the model for the elementary group (focal) and the secondary (reference) group were statistically the same (Table 14). Thus, it is concluded that the instrument measures lack equivalence, with regard to the demographic groups in this study: elementary teachers and secondary teachers. That is, the results show the battery of 26 items that were supposed to measure the latent trait of teacher collective responsibility for student learning did not in fact measure the same construct across groups identified by grade level taught. This is taken as indicating the two groups are not interpreting the construct in the same way, which is to say that collective responsibility has a different meaning for elementary teachers than it does for secondary teachers.

Survey Items with Dissimilar Factor Loadings Across Groups

In addition to the initial investigation, the results of the separate EFAs conducted for each demographic group helped to identify specific items with dissimilar factor loadings across groups. The results show the greatest differences in factor loadings between elementary and secondary groups were for Factor 1 (teacher as reporter in own classroom). The specific items that should have but did not load on Factor 1 (Reporter in Classroom) for the elementary (and did for the secondary) are 1, 2, and 20.

Item 20 says "Other teachers come to me for help with instructional issues." Since the question gives no explanation of the situation, teachers must interpret it based on their personal experiences. This lends itself to an understandable difference that exists between

elementary and secondary teachers, based on dissimilar perspectives and unlike definitions of what constitutes "coming for help." Due to both the physical structure and the collaborative environment of most elementary buildings, it is much easier for elementary teachers to contact peers and engage them in professional conversation involving instructional issues (DuFour, 1997). Thus, it is quite probable that one teacher could seek assistance from another in a casual, non-intrusive manner. In contrast, the secondary teachers are typically much more secluded from each other (Bryk & Driscoll, 1988). Therefore, the act of seeking help is a more overt and structured behavior, which may lessen its frequency of occurrence. As a consequence, it is likely that the concept of "seeking help with instructional issues" is interpreted differently for elementary and secondary teachers. Therefore, because the situation in the question was not clearly defined, based on their prior experiences, it is likely that the two groups interpret it differently. Thus, a difference in factor loading could be expected, which is what, in fact, the findings show.

Item 1 is "In this school, teachers feel responsible that all students learn." Here also, it is quite conceivable there is a discernable difference between elementary and secondary teachers based on a lack of clear definition for "responsible." Due to the obvious fact that elementary students are younger than secondary, elementary teachers feel a greater urgency to assume a care-taking or "responsibility" role than secondary teachers do (Meier, 1995). Therefore, it is likely that the two groups will not answer the question in the same manner because they do not have a common meaning for "responsible." As an additional contributing factor, a vast majority of elementary teachers are female, whereas a greater number of secondary teachers are male. Prior research has

shown that these two groups view differently their roles as teachers, including the degree to which they are responsible for their students (Bress, 2000; Yuen & Ma, 2002). As a consequence, the difference is reflected in dissimilar factorial structure for the elementary and secondary groups on this item.

Item 2 states, "In this school, teachers hold prominent leadership roles." Once again, there is a reasonable explanation as to why this item was interpreted differently by elementary and secondary teachers. In educational literature, it is well documented that elementary and secondary teachers view their role in the governance of the school in a different light (Deal & Peterson, 1994; Lee, Dedrick, & Smith, 1991). Studies have found that the position of being a "leader" as well as the expectations for such are viewed more positively by secondary than elementary teachers. Secondary teachers have more confidence in their ability to fill the leadership role and more readily accept them (Peterson & Deal, 1998). Thus, the dissimilarity between elementary and secondary in the loading of this item due to a difference in interpretation of the construct is in line with findings from prior research.

From this brief discussion of the lack of ME manifested in dissimilar factorial structure of the responses from elementary and secondary teachers in this investigation, it becomes obvious there are inherent differences between the two. Even though both groups deal with the education of children, the circumstances under which they work are quite different, a difference that can not be ignored. Rather, to achieve maximum results in attempting to use survey instruments in situations involving teachers throughout the PreK-12 school setting, care must be taken in providing a common conceptual framework and associated vocabulary. This may be established through prior in-service programs or

additional explanation provided within the text of the measurement instrument. If this is not done, the validity of conclusions drawn from studies where measurement equivalence is not considered may be in question (Vanderberg & Self, 1993). Thus, the results of the efforts may be discounted by the skeptics, regardless of the amount of work or expense that has been invested.

It is important to note that from this singular investigation, it can not be concluded that in all situations elementary and secondary teachers vary in their definition of collective responsibility. It is possible that in some situations the necessary establishment of commonality has been achieved. It does, however, point out the fact the ME substantiation is needed before the inevitable comparisons of mean values can be accurately made. This is extremely important because if the construct of interest, whether it be collective responsibility or something else, is not measured equivalently across groups, then a comparison of means across groups may be inaccurate, unwarranted, or even meaningless (Golembiewski et al., 1976; Schmitt, 1982; Vandenberg & Self, 1993). This has an important implication for the field of education, as the substantiation is not routinely done. Thus, those who are in a position to do so, such as administrators and research specialists, but elect not to substantiate measurement equivalence may be unknowingly contributing to the lack of credibility of American schools perceived by the general public. It would be a simple task to strength educational research findings by verifying that the measures from the instrument used in the investigation do not lack measurement equivalence. Thus, comparison of mean values on whatever is being measured could be made with the confidence that differences in mean values are

reflections of true differences in the construct, not artifacts of differences in construct meaning.

Implications of Efforts to Measure Teacher Collective Responsibility

Through prior research, higher collective responsibility has been linked to greater student academic achievement (DuFour, 1997; DuFour & Eaker, 1998; Lee & Smith, 1996). As a result, a growing number of schools are attempting to accelerate academic achievement by also increasing teacher collective responsibility for student learning. Knowing that collective responsibility may not be viewed by secondary and elementary teachers in the same way has strong implications for these efforts.

First, when programs, such as professional development designed to increase collective responsibility, are being prepared for presentation to an entire PreK-12 audience, to be effective, it must be recognized that before any progress can be made in improving collective responsibility, first, a consensus must be reached as to its meaning. It would be futile to proceed without doing so. From the beginning, input from all sectors of the school community is vital in order to establish agreement. Thus, it is critical to the success of such a professional development program that administrators demand total faculty involvement at the onset to establish the essential common vocabulary needed for consistent interpretation of collective responsibility.

Second, attempts to measure initial levels of collective responsibility across grade levels would, most probably, be inaccurate and misleading unless the instrument being used has been examined, and it has been verified that the measures from it do not lack measurement equivalence. Without such verification, there is no way to establish with complete certainty that differences in mean values reflect true differences in the level of

collective responsibility or other construct. This makes it virtually impossible to determine if increases are needed when it is not possible to determine with a high degree of accuracy the current level of collective responsibility of the teachers.

Finally, following the professional development programs or interventions, attempts to measure changes or new levels of collective responsibility where measurement equivalence has not been substantiated run the risk of being invalid, thereby resulting in unwelcomed and costly errors. Although administrators or researchers may be able to show significant differences in mean values over time, those changes are highly suspect if verification of measurement equivalence of the instrument being used has not been done. Rather than reflecting true increases (or decreases) in the level of collective responsibility of the faculty, they may only be the result of converging definitions brought about by in-service programs. Thus, those who are in a position of authority have an obligation to ensure every effort has been made to avoid faulty inferences and incorrect conclusions by every means possible, including substantiation of measurement equivalence.

The points outlined in the preceding paragraphs are applicable not only to teacher collective responsibility for student learning but also for efforts to measure any latent trait. The measurement of any latent trait is difficult due to the fact that, by definition, a latent trait is unseen. However, this does not mean that it is also necessarily undefined. Rather, in working with any latent trait, a common vocabulary, meaning, and understanding can be achieved if sufficient effort is applied. The verification that the measures from the instrument being used for such do not lack equivalence is one effort that can, and should be applied in all situations to achieve reliable and compelling research findings.

CHAPTER 9: CONCLUSIONS

Implications of the Findings

The results from this investigation show that the *W*-index procedure is a reliable MIRT method to identify a lack of measurement equivalence under certain conditions. Specifically, those conditions include a sample size of 2000 for any case or 1000, if the requirement for a small intertrait correlation (.20) is met. Additionally, it is important to note that the small sample size of 150 may not result in an "Acceptable" identification of lack of equivalence, regardless of the other criteria. This is an important finding for educational research because here the issue of sufficient sample size is often ignored or overlooked in the zeal for a convenient or available sample. This study shows clearly that with this procedure, as with many others, small sample size produces marginally acceptable results, at best. Thus, researchers who opted to use this method with a sample of less than 500 are running the risk of inaccurate results and faulty conclusions, even though other criteria are met.

With regard to the intertrait correlation, the findings were also in line with what was expected from prior research. In most cases (92%), as the intertrait correlation increased, the accuracy of the procedure decreased. Thus, the *W*-index procedure would be most appropriate for use with multidimensional instruments where the factors have a weak correlation (at .20 or less). This requirement is a reasonable restriction for instrument developers who can control the strength of the intertrait correlation on their instrument. It may not be as reasonable for those who are attempting to verify ME on measures obtained from an existing instrument.

A somewhat surprising third finding from this study is that a larger number of items lacking equivalence did not necessarily result in an acceptable power rate. In only 58% of the cases did an increase in number of items lacking equivalence results in increased statistical power. Thus, for this method, a minimum number of items lacking equivalence is not an assumption that must be met. In fact, acceptable identification rates were obtained for as few as 2 items (or 8%) lacking equivalence, when other criteria of large sample size and small intertrait correlation were met. The number of items lacking equivalence was a contributor, but not the sole determining factor, for accurate results with the *W*-index procedure. Although contrary to what was hypothesized, this may actually be considered a positive finding for instrument developers who are aware that a large percentage of items lacking ME is not an assumption that must be met in order to utilize the *W*-index procedure.

Some mention should be made of the fact that there were two two-way interactions found: between sample size and intertrait correlation and number of items lacking equivalence and sample size squared. However, an extensive discussion is unwarranted, as both were removed from the final model due to the fact that even though they were statistically significant, they were not substantively meaningful.

Consequences of Ignoring Measurement Equivalence

As stated at the onset, an essential attribute of any psychological or behavioral instrument is that it measures the intended construct equally well across groups. That is, the measures possess measurement equivalence. Thus, if the substantiation of ME is not undertaken, the researcher runs the risk that the instrument does not possess the most fundamental of attributes. Without first establishing ME, it is possible, and even probable,

that the instrument may not meet the required "prerequisites" for group comparisons (Riordan et al., 2001). If it is not verified that the construct of interest is the same for all groups, comparisons of it, as measured by a mean value or some other quantitative method, can not be made. Attempts to do so revert to the cliché of comparing "apples to oranges." This concern is supported by researchers, such as Riodan and Vandenberg (1994) who state that only when subjects from different groups ascribe essentially the same meaning to the scale items can meaningful across-groups comparison be conducted. If this is not done, mean differences may only be an artifact of lack of equivalence, not true differences in the construct being measured. Many individual researchers, as well as research groups, have warned that the result of ignoring the ME investigation is that the customary comparison of means across groups may be inaccurate, unwarranted, or even meaningless (AERA, APA, & MNME, 1999; Bejar, 1980; Golembiewski et al., 1976; Schmitt, 1982; Vandenberg & Self, 1993). Conversely, when the investigation of lack of ME is completed, the researcher can assert findings based on mean differences with the assurance that the same construct has been measured across groups.

When the lack of ME has not been tested, there is also a problem with the inferences and recommendations based on mean score differences. According to Chan (2000), these, too, may be inaccurate and, therefore, also have a high probability of being misleading. This results in a major problem, as the validity of the conclusions drawn from these studies may be questionable (Vandenberg & Self, 1993). Without validity, results are meaningless. Hence, to avoid costly errors and to produce compelling findings, the substantiation of ME must be added as an essential factor for convincing research.

Limitations of This Study

There are some important limitations of this examination to note. First, the simulation study and the *W*-critical value used as an index derived from that simulation are based on data that is generated to perfectly fit the MIRT model. However, the reality of real data is that it does not perfectly fit the model. Thus, although the *W*-index may be shown to produce accurate results in the situation modeled, there is no guarantee without further substantiation that it may be generalized to all situations encountered.

Second, there are other factors in the simulation phase that limit the generalizability of the findings in this study. For example, the assumption was made that the data conformed to a Rasch model. Also, the number of dimensions in the simulation was limited to two. Additionally, several elements were held constant. Those were 1) the discrimination parameters (α, both within and between items), 2) the number of rating scale categories, 3) equal taus or distances between item category thresholds, and 4) the number of items on the instrument for all conditions. These conditions are certainly not applicable to all situations, and, therefore, restrict the generalizability of the findings.

As a third limitation of the study, only 200 data sets for each null condition for each group and 50 data sets for each of the groups per cell for the other cells and groups were generated. An increase in number of data sets generated that may be needed to verify that similar results are obtained in future studies is actually more than just being "of value." It may actually be required because well-established critical values to be use with this procedure have not yet been determined.

Fourth, the effects of only three experimental factors on the accuracy rate of the method were investigated. There are numerous other factors that have been shown in

previous research to affect the accuracy rate of the method being using. Among these are 1) the effects of theta location (Seraphine, 2000); 2) the effects of test length (De Champlain & Gessaroli, 1996; De Champlain et al., 1998; Flowers et al., 1999); 3) number of traits (van Abswoude et al., 2004); and 4) the effects of number of scale (Seraphine, 2000). It would be important in future investigation of the *W*-index method to include as many of these factors as is feasible.

Fifth, in addition to investigating only 3 factors, within each of those factors there are additional limitations. With regard to the number of items lacking equivalence, only 2, 4, and 6 items lacking ME were included. These constitute 8%, 15%, and 23%, respectively, of the total items. It would be helpful in the future to consider other numbers. The situation where only one items lacks ME should have been included, as that is a situation frequently found with survey instruments. Also, only 3 values for intertrait correlation were considered. Many previous studies using other techniques have included both larger and smaller values. Thus, it is not possible to make a direct comparison with these findings, which is an additional limitation of the study.

Finally, the most significant limitation of this investigation is that, the accuracy of the *W*-index to identify lack of measurement equivalence was not compared to any existing method. Thus, it is difficult to draw a conclusion as to whether or not this is a better method than what now exists because prior research using methods other than this have different designs. As a consequence, it is not possible to accurately gauge how this procedure would compare to others under like conditions. Hence, in future studies, it would be of value to compare its accuracy to another in the same study with an identical study design, hereby, providing a direct comparison.

Issues for Future Research

Among the many issues connected to the use of an MIRT procedure to investigate ME still waiting to be addressed, there are two that I feel are of most importance for future research. The first pressing issue is the development of a practical fit index for MIRT models involving small sample size. Of course, this would also necessitate accompanying guidelines and critical values. The establishment of a widely-accepted and easy-to-use MIRT fit index would, without a doubt, be a valuable contribution, as it has the potential to rival SEM indices and significantly increase the use of IRT and MIRT procedures in ME investigations.

A second significant contribution to the item response theory repertoire as a result of future research should be the development of modification indices for MIRT that are similar to those currently used in SEM and CFA for situations where a lack of measurement equivalence is established. Presently, this is completed in IRT by the "brute force" method of testing all models that differ from a given model by adding a single parameter estimate or by relaxing a single constraint. Obviously, with large models, this is time-consuming and incredibly inefficient. Thus, the development of such indices would be another valuable contribution that could also lead to increased use of multidimensional item response theory procedures, as being called for by the IRT community. Unfortunately, the use of MIRT procedures for measurement equivalence verification lags far behind that of SEM. By making available to researchers viable IRT and MIRT procedures, there is a strong possibility that this situation will change in the future.

APPENDICES

Appendix A. IRT Investigations of Effects of Variation in Experimental Factors

Effects of Sample Size

Sample size	Author(s)	Title of study	Purpose	Method/Indices	Results
n = 250 500 1,000	DeChamplain, Gessaroli, Tang, and DeChamplain (1998)	Assessing the Dimensionality of Polytomous Item Responses with Small Sample Sizes and Short Test Lengths, A Comparison of Procedures	To examine the effectiveness of *Poly-DIMTEST* and *LISREL8* when compared with polytomous unidimensional data sets simulated to vary as a function of test length and sample size.	*Poly-DIMTEST* t-empirical statistic *LISREL8* Chi square fit statistic Examined Type I error rate	*Poly-DIMTEST* - Type I error probabilities near nominal values for all ***n*** ; unaffected by manipulation of ***n***. *LISREL8* - Severely inflated Type I error rates for all conditions, except 10-item data sets; n=500 and 1,000 Neither procedure worked well for n<500 With n>500 *LISREL8* chi square statistic can be useful for assessment of dimensionality.
n = 500 1,000	DeChamplain and Gessaroli (1991)	Assessing Test Dimensionality Using an Index Based on Nonlinear Factor Analysis	To examine effectiveness of IFI in identifying dimensionality To compare performance of IFI with the t-statistic of W. Stout (1987).	Sum of squares of residual covariances chi-square statistic (IFI index) Stout's t-statistic *NOHARM* Examined rejection rate	The IFI squares of residual covariances showed fairly high rejection rates of unidimensionality with two-dimensional data. The t-statistic seemed best suited for long tests having large ***n***, while the IFI might be preferable for smaller test lengths or smaller samples. For multivariate normal data, the chi-square statistic is as sensitive to departures from dimensionality as the Stout t statistic, especially for small sample sizes and short tests.

Effects of Number of Items per Trait

Items per trait	Author(s)	Title of study	Purpose	Method/Indices	Results
k = 7 (16%) 21 (47%) (45 items)	vanAbswoude, van der Ark, Sijtsma (2004)	A Comparative Study of Test Data Dimensionality Assessment Procedures Under Nonparametric IRT Models	To compare effectiveness of MSP (Mokken Scale *DETECT*, *DIMTEST* and HCA/CCPROX (Hierarchical Agglomerative Clustering/ customized proximity measure)	Cluster Analysis For MSP, the scalability coefficient H For *DETECT* and HCA/CCPROX, the expected conditional covariances Dmax For *DIMTEST*, Stout's t-index	In finding simulated structure, both latent traits (*DETECT* and HCA/CCPROX) were superior to normed unconditional covariances (MSP). *DETECT* and *DIMTEST* did not always reflect the true dimensionality of the item pool. Effectiveness of all methods decreased with increase in intertrait correlation.
k = 20 (50%) 30 (75%) (40 items)	Hambleton & Rovinelli (1986)	Assessing the Dimensionality of a Set of Test Items	To compare 1) linear factor analysis; 2) residual analysis; 3) nonlinear factor analysis; and 4) Bejar's method.	Factor Analysis using Three-parameter logistic model	Linear factor analysis in all instances overestimated the number of underlying dimensions. Nonlinear factor analysis, with linear and quadratic terms, led to correct determination of the item dimensionality. Both residual analysis and Bejar's method disappointing. Results suggest extreme caution in using linear factor analysis, residual analysis, and the Bejar method, Nonlinear factor analysis most promising.

Effects of Test Length

Test length	Author(s)	Title of study	Purpose	Methods/Indices	Results
x = 10 items 20 items	De Champlain, Gessaroli, Tang, and De Champlain (1998)	Assessing the Dimensionality of Polytomous Item Responses with Small Sample Sizes and Short Test Lengths, A Comparison of Procedures	To examine effectiveness of *Poly-DIMTEST* and *LISREL8*, compared with polytomous unidimensional data sets simulated to vary as a function of test length and sample size.	*Poly-DIMTEST* t-empirical statistic *LISREL8* Chi square fit statistic Examined Type I error rate	*LISREL8* - Severely inflated Type I error rates obtained with for all conditions, except the 10-item data sets; 500 and 1,000 simulees. *Poly-DIMTEST* - Type I error probabilities at or near nominal values for all sample sizes; unaffected by the manipulation of sample size. *Poly-DIMTEST* T-statistic lacks the power needed to use with samples of fewer than 20 items.
x = 15 items 45 items	De Champlain and Gessaroli (1991)	Assessing Test Dimensionality Using an Index Based on Nonlinear Factor Analysis	1) To examine effectiveness of IFI in identifying dimensionality 2) To compare performance of IFI with the T-statistic of W. Stout (1987).	Sum of squares of residual covariances (IFI index) Stout's t-statistic *NOHARM* chi-square statistic Examination of rejection rate	The IFI squares of the residual covariances showed fairly high rejection rates of unidimensionality when two-dimensional data were generated. The t -statistic seemed best suited for long tests with large sample size, while the IFI preferable for smaller test lengths or smaller samples. For multivariate normal data, the chi square statistic is as sensitive to departures from dimensionality as the Stout T statistic, especially for small sample sizes and short tests.

Effects of Number of Traits

Number of traits	Author(s)	Title of study	Purpose	Methods/Indices	Results
x = 2 4	vanAbswoude, van der Ark, & Sijtsma (2004)	A Comparative Study of Test Data Dimensionality Assessment Procedures Under Nonparametric IRT Models	To compare effectiveness of MSP (Mokken Scale for polytomous items), *DETECT*, *DIMTEST* and HCA/CCPROX (Hierarchical Agglomerative Clustering/ customized proximity measure)	Cluster Analysis For MSP, the scalability coefficient H For *DETECT* and HCA/CCPROX, the expected conditional covariances Dmax For *DIMTEST*, Stout's t-index	Methods that use covariances conditional on the latent trait (*DETECT* and HCA/CCPROX) were superior in finding the simulated structure to the method that used normed unconditional covariances (MSP). *DETECT* and *DIMTEST* did not always reflect the true dimensionality of the item pool. Effectiveness of all methods decreased with increase in intertrait correlation

Effects of Theta Location

Θ location	Author(s)	Title of study	Purpose	Methods/Indices	Results
Θ = 0.00 .25 .50 .75 1.00 1.25 1.50	Seraphine (2000)	The Performance of *DIMTEST* when Latent Trait and Item Difficulty Distributions Differ	To examine effect of shifts in difficulty (b) and examinee trait (Θ) shifts in location on *DIMTEST*, with ceiling effects of minimum competency testing.	*DIMTEST*, Stout's t-statistic Examined Type I error rate sample size = 1,500 Test length =50 Intertrait correlation r= .3.	The power of *DIMTEST* was reduced as the location shifted upward and the scale shifted downward. This held only when the AT1 items were selected using principal axis factor analysis. When AT1 items were selected using expert opinion, the procedure was only slightly affected by changes in location and scale.

Effects of Shift in Difficulty and Examinee Trait

Scale	Authors	Title of study	Purpose	Methods/Indices	Results
s = 1.0 .9 .8 .7	Seraphine (2000)	The Performance of *DIMTEST* when Latent T rait and Item Difficulty Distributions Differ	To examine effect of shifts in difficulty (b) and examinee trait (Θ) shifts in location on *DIMTEST*, with ceiling effects	*DIMTEST*, Stout's t-statistic Type I error rate sample size n = 1,500 test length x=50 Normal distribution Θ correlation r= .3.	Power of *DIMTEST* was reduced as the location shifted upward and the scale shifted downward. This held only when the AT1 items were selected using principal axis factor analysis. When AT1 items were selected using expert opinion, the procedure was only slightly affected by changes in location and scale.

Effects of Strength of Intertrait Correlation

Intertrait correlation	Author(s)	Title of study	Purpose	Methods/Indices	Results
r = 0 .2 .4 .6 .8 1.0	vanAbswoude, van der Ark, Sijtsma (2004)	A Comparative Study of Test Data Dimensionality Assessment Procedures Under Nonparametric IRT Models	To compare the effectiveness of 4 methods, *MSP* (Mokken Scale for polytomous items), *DETECT*, *DIMTEST* and *HCA/CCPROX* (Hierarchical Agglomerative Clustering/ customized proximity measure)	Cluster Analysis For *MSP*, the scalability coefficient H For *DETECT* and *HCA/ CCPROX*, the expected conditional covariances Dmax For *DIMTEST*, Stout's T-index	Methods that use covariances conditional on the latent trait *DETECT* and *HCA/CCPROX* were superior in finding the simulated structure to the method that used normed unconditional covariances *MSP*). *DETECT* and *DIMTEST* did not always reflect the true dimensionality of the item pool. Effectiveness of all methods decreased with increase in intertrait correlation.

Intertrait correlation	Author(s)	Title of study	Purpose	Methods/Indices	Results
r = .50 .75 .90	Gosz, & Walker (2003)	Empirical Comparison of Multidimensional Item Response Data Using *TESTFACT* and *NOHARM*	To compare *TESTFACT* and *NOHARM*	EFA ANOVA (Tukey) Root mean square deviations (RMSD) Type I error	*TESTFACT* outperformed *NOHARM* with 24 bi-dimensional items and high intertrait correlation (.9) For 12 bi-dimensional items, *NOHARM* did better, especially when the correlation between dimensions was lower (0.5 or 0.75).
r = .3 .7	Nandakumar (1994)	Assessing Dimensionality of a Set of Item Responses, Comparison of Different Approaches	To compare the performance of *DIMTEST*, the approach of Holland and Rosenbaum [covariance based, non-factor analytic procedure] and nonlinear factor analysis for unidimensionality uses simulation and real data	*DIMTEST*, t-index; Approach of Holland and Rosembaum multi-dimensional compensatory model, conditional covariance Non-linear analysis, goodness-of-fit statistic; Mean and s.d. of squared residuals and absolute residuals	All 3 models correctly confirm unidimensionality equally. *DIMTEST* had greater power in detecting multi-dimensionality for correlations between abilities as high as .7. H&R's and the nonlinear factor analysis methods demonstrated good power provided the correlation between abilities was low (r = .3). Same for simulated and real data

Appendix B. Teacher Collective Responsibility for Student Learning Survey Instrument

As a part of a research project through the College of Education at Michigan State University, teachers in your school are being asked to respond to the following survey. The project is called "National Board Certified Teachers as an Organizational Resource." The research focus is on understanding the relationship between National Board Certified Teachers and school-level collective responsibility. The data collected from this survey will be used in this project.

Please indicate your voluntary agreement to participate by providing your signature below, then completing and returning this survey. All data collected will be kept confidential. Participating in this study is voluntary, and this survey is expected to take approximately 15 minutes to complete. You may choose not to answer any question or stop at any time.

Although your confidentiality will be protected in all publications by using a pseudonym for each school as well as identification numbers for individual teachers, you or others may be able to discern some of the identities based on reported attributes of the school and person. Some questions may request sensitive information about your commitment to your students and relationships with colleagues and parents. To minimize risks, only the investigators will know respondents' identities and this information will not be shared with anyone beyond the research team, including other teachers and school officials. Further, data will not be reported in a manner that allows individuals to be identified. Your privacy will be protected to the maximum extent allowable by law. Note that nothing will be published from these data until 2004.

If you have questions or concerns regarding your rights as a study participant, or are dissatisfied at any time with any aspect of this study, you may contact – anonymously, if you wish – Ashir Kumar, M.D., Chair of the University Committee on Research Involving Human Subjects (UCRIHS) by phone: (517) 355-2180, fax: (517) 432-4503, e-mail: ucrihs@msu.edu, or regular mail: 202 Olds Hall, East Lansing, MI 48824.

If you have any questions about this study, please feel free to contact the individuals below:

Gary Sykes
410A Erickson Hall
East Lansing, MI
(517) 353-9337
E-mail: garys@msu.edu

Linda Chard
118 Erickson Hall
East Lansing, MI
(810) 603-1940
E-mail: chardlin@msu.edu

You indicate your voluntary agreement to participate by signing below, and completing and returning this questionnaire.

Signature _____

Date _____

Name (*please print*) _____

Background Characteristics

Please circle the appropriate response.

Gender: Female Male

Teaching area this year (circle all that apply)

Art
Career and Technical Education
English
Health Education
Math
Music
Science
School Counseling
Social Studies
Special Education, K - 12
World Languages other than English
Other -- specify_____

Grade level taught this year (circle all that apply)

Pre-K K 1 2 3 4 5 6 7 8 9 10 11 12 Not in a classroom

Race (circle all that apply)

Asian
African American/Black, non-Hispanic
Hispanic/Latino
Native American/American Indian
Caucasian/White, non-Hispanic
Other -- specify _____

Collective Teacher Beliefs

This survey is designed to help us gain a better understanding of faculty perceptions of their school and the learning environment. Please respond to each of the questions by considering the ***current*** conditions in your school.
Your answers are confidential.

Directions: Please indicate level of agreement with each statement by circling the descriptor that best depicts your opinion. The scale of responses ranges from "Strongly Disagree" (1) to "Strongly Agree" (4).

	Strongly Disagree	Disagree	Agree	Strongly Agree
1. In this school, teachers feel responsible that all students learn.	(1)	(2)	(3)	(4)
2. In this school, teachers hold prominent leadership roles.	(1)	(2)	(3)	(4)
3. Teachers in this school are prepared to teach the subjects they are assigned.	(1)	(2)	(3)	(4)
4. Teachers in this school adapt their lessons to enable students to learn.	(1)	(2)	(3)	(4)
5. Teachers in this school help each other do their best.	(1)	(2)	(3)	(4)
6. In this school, teachers frequently discuss instructional improvement.	(1)	(2)	(3)	(4)
7 In this school, teachers are supportive of each other.	(1)	(2)	(3)	(4)
8. I know what happens in other teachers' classrooms.	(1)	(2)	(3)	(4)
9. I observe positive ways teachers relate to their students.	(1)	(2)	(3)	(4)
10. I know how other teachers deal with difficult students in their classrooms.	(1)	(2)	(3)	(4)
11. I have observed other teachers who try to help students who are failing.	(1)	(2)	(3)	(4)
12. I know in which classrooms students are showing academic growth.	(1)	(2)	(3)	(4)
13. I know the extent to which teachers exchange educational materials and techniques.	(1)	(2)	(3)	(4)
14. I know the extent to which other teachers in this school are applying new teaching techniques.	(1)	(2)	(3)	(4)
15. I am responsible for the performance of all of my students.	(1)	(2)	(3)	(4)
16. I know how to teach students with diverse abilities.	(1)	(2)	(3)	(4)

Directions: Please indicate level of agreement with each statement by circling the descriptor that best depicts your opinion. The scale of responses ranges from "Strongly Disagree" (1) to "Strongly Agree" (4). Your answers are confidential. Please respond to each of the questions by considering the ***current*** conditions in your school.	Strongly Disagree	Disagree	Agree	Strongly Agree
17. It is my responsibility to make sure my class runs smoothly every day.	(1)	(2)	(3)	(4)
18. I know how to teach students from diverse backgrounds.	(1)	(2)	(3)	(4)
19. I feel it is necessary to adapt my teaching methods to meet my students' needs.	(1)	(2)	(3)	(4)
20. Other teachers come to me for help with instructional issues.	(1)	(2)	(3)	(4)
21 I work with staff and administration to solve school-related problems.	(1)	(2)	(3)	(4)
22. I help resolve conflicts between the school and parents/community.	(1)	(2)	(3)	(4)
23. I share a common mission with others in this school.	(1)	(2)	(3)	(4)
24. I work with others to control disruptive behavior.	(1)	(2)	(3)	(4)
25. I work with other teachers and administrators to keep students interested in school.	(1)	(2)	(3)	(4)
26. I work with other teachers and /or administrators on instructional improvement.	(1)	(2)	(3)	(4)

Appendix C. SAS Code to Generate Data

```
%macro iter(iter,cell,n,r,i,p,sd1,sd2,sd3,sd4,sd5,sd6,tau1,
tau2,tau3,rs);

/****  seed values***********************************/

  %let seed1=%eval(&iteration*&cell*&sd1);
  %let seed2=%eval(&iteration*&cell*&sd2);
  %let seed3=%eval(&iteration*&cell*&sd3);
  %let seed4=%eval(&iteration*&cell*&sd4);
  %let seed5=%eval(&iteration*&cell*&sd5);
  %let seed6=%eval(&iteration*&cell*&sd6);
  %let ns=%eval(2*&n);
  %let ntest=%eval(&n+1000);

/****  generate theta1 theta2 ************************/

data person;
  do person=1 to &ns.;
    base=rannor(&seed1.);
    r1=rannor(&seed2.);

    theta1=base;
    theta2=(&rs.*base)+((1-(&rs.**2))**.5)*r1;

  output;
 end;
run;


/****  generate delta *********************************/

data item;
  array delta d1-d&i.;
   do over delta;
     delta=rannor(&seed4.);
  end;
run;

/****  fill arrays ***********************************/

data both;
 if _n_=1 then set item;
   set person;
   person=person+1000;

     array delta d1-d&i.;

     array prob1s pa1-pa&i.;
     array prob2s pb1-pb&i.;
     array prob3s pc1-pc&i.;

     array prob11s paa1-paa&i.;
     array prob12s pbb1-pbb&i.;
```

```
     array prob13s pcc1-pcc&i.;

     array scores sa1-sa&i.;
     array scoress sb1-sb&i.;

     array randvar ra1-ra&i.;
     array randvars rb1-rb&i.;

 do over prob1s;

     prob1s = exp(theta1-delta-&tau1.)/(1+(exp(theta1-delta-&tau1.)));
     prob2s = exp(theta1-delta-&tau2.)/(1+(exp(theta1-delta-&tau2.)));
     prob3s = exp(theta1-delta-&tau3.)/(1+(exp(theta1-delta-&tau3.)));

     prob11s = exp(theta2-delta-&tau1.)/(1+(exp(theta2-delta-&tau1.)));
     prob12s = exp(theta2-delta-&tau2.)/(1+(exp(theta2-delta-&tau2.)));
     prob13s = exp(theta2-delta-&tau3.)/(1+(exp(theta2-delta-&tau3.)));

/****  category classification *************************/

  randvar=ranuni(&seed5.);
  randvars=ranuni(&seed6.);

  scores=1;
  scoress=1;

    if randvar < prob1s then scores=2;
    if randvar < prob2s then scores=3;
    if randvar < prob3s then scores=4;

    if randvars < prob11s then scoress=2;
    if randvars < prob12s then scoress=3;
    if randvars < prob13s then scoress=4;

 end;
run;

/****  create data sets - person id, scores on thetas ***/

data winfile;
 file "C:\A_data\datag1_&cell._&iteration..dat";
  set both;
  where person le &ntest;
        put person @10 (sa1-sa13 sb14-sb&i.)(+(-1));
run;

data winfile;
 file "C:\A_data\datag2_&cell._&iteration..dat";

  set both;
  where person gt &ntest;

        if &p in(1) then do;
        put person @10 (sa1-sa13 sb14-sb&i.) (+(-1));
  end;

        if &p in(2) then do;
```

```
        put person @10 (sa1-sa15 sb16-sb&i.) (+(-1));
  end;

        if &p in(3) then do;
        put person @10 (sa1-sa17 sb18-sb&i.) (+(-1));
  end;

run;

/********  correlation between thetas  *****************/

proc corr nosimple;
   var theta1 theta2;
run;

%MEND iter;

/*
     cell=(n)(r)(p)
     n1=150, n2=500, n3=1000 n4=2000
     r1=.2  r2=.4  r3=.6 intertrait correlation
     p1=0   p1=2   p2=4 p3=6 items with different factor loading
%macro iter(iter,cell,n,r,i,p,sd1,sd2,sd3,sd4,sd5,sd6,tau1,tau2,tau3,rs)

*/

%iter(  1,111, 150,1,26,1,1,2,3,4,5,6,-1,0,1,.2);
.
.
.
%iter( 50,434,2000,3,26,4,1,2,3,4,5,6,-1,0,1,.6);
```

Appendix D. WINSTEPS Code to Generate Data

```
START /WAIT WINSTEPS BATCH=YES Control-file Output-file
Extra=specifications
START /WAIT WINSTEPS BATCH=YES command.cmd 1111.out data=1111.dat
pfile=1111.prs ifile=1111.itm rfile=1111.res
START /WAIT WINSTEPS BATCH=YES command.cmd 1112.out data=1112.dat
pfile=1112.prs ifile=1112.itm rfile=1112.res
START /WAIT WINSTEPS BATCH=YES command.cmd 1113.out data=1113.dat
pfile=1113.prs ifile=1113.itm rfile=1113.res
START /WAIT WINSTEPS BATCH=YES command.cmd 1114.out data=1114.dat
pfile=1114.prs ifile=1114.itm rfile=1114.res
START /WAIT WINSTEPS BATCH=YES command.cmd 1115.out data=1115.dat
pfile=1115.prs ifile=1115.itm rfile=1115.res
START /WAIT WINSTEPS BATCH=YES command.cmd 1116.out data=1116.dat
pfile=1116.prs ifile=1116.itm rfile=1116.res

.
.
.
```

Appendix E. SAS Code to Create W-statistic for Groups and Merge

```
/****  1nulls_null *********************************/

%macro null(iter,cell,n,r,i,p,sd1,sd2,sd3,sd4,sd5,sd6,tau1,tau2,tau3,rs);

data d1;
 infile "C:\A_data\outputg1_&cell._&iter..txt" firstobs=11 obs=13;
    input @18 n1 / @19 deviance1 / @43 parameters1;
      df1 = n1-parameters1;
      pc1 = deviance1 / df1;
run;

data d2;
 infile "C:\A_data \outputg2_&cell._&iter..txt" firstobs=11 obs=13;
    input @18 n2 / @19 deviance2 / @43 parameters2;
       df2 = n2-parameters2;
       pc2 = deviance2 / df2;
run;

data both;
 file 'C:\A_data\output_1nulls.dat' mod;
   merge d1 d2;
   f=pc2/pc1;
   n=&n;
   r=&r;
   p=&p;
   iter=&iter;
   cell=&cell;
 put n1 deviance1 parameters1 df1 pc1 n2 deviance2 parameters2 df2 pc2
    f n r p cell iter;
run;

%mend null;
%iter(  1,111, 150,1,26,1,1,2,3,4,5,6,7,8,9,10,-1,0,1,.2);
.
.
.
%iter( 50,434,2000,3,26,4,1,2,3,4,5,6,7,8,9,10,-1,0,1,.6);
```

Appendix F. SAS Code to Identify W-Critical Value for Null Condition

```
/****  2nulls ****************************************/

data d1;
 infile 'c:\A_data\output_1nulls_null.dat';
 input n1 deviance1 parameters1 df1 pc1
  n2 deviance2 parameters2 df2 pc2
  f n r p cell iter;
run;

proc sort;
      by n r p;
run;

proc freq;
 title 'NULL DISTRIBUTIONS FOR EACH CELL OF THE EXPERIMENTAL DESIGN';
 where p = 1;
      by n r p;
 table f;
run;
```

Appendix G. SAS Code to Identify Statistical Power Rate

```
/****  allrates ***************************************/

data d1;
  infile 'C:\A_data\output_1nulls_null.dat';
    input n1 deviance1 parameters1 df1 pc1
      n2 deviance2 parameters2 df2 pc2
      f n r p cell iter;
run;

proc sort;
   by n r;
run;

data d2;
  infile 'C:\A_data\p2\output_1nulls_p2.dat';
     input n1 deviance1 parameters1 df1 pc1
     n2 deviance2 parameters2 df2 pc2
     f n r p cell iter;
run;

proc sort;
   by n r;
run;

data d3;
  infile 'C:\A_data\p3\output_1nulls_p3.dat';
     input n1 deviance1 parameters1 df1 pc1
     n2 deviance2 parameters2 df2 pc2
     f n r p cell iter;
run;

proc sort;
   by n r;
run;

data d4;
   infile 'C:\A_data\p4\output_1nulls_p4.dat';
     input n1 deviance1 parameters1 df1 pc1
     n2 deviance2 parameters2 df2 pc2
     f n r p cell iter;
run;

proc sort;
   by n r;
run;

data nulls;
 input n r p wcrit;
 cards;
 150 1 1 1.0213788742
 150 2 1 1.0196160983
 150 3 1 1.0238464383
 500 1 1 1.013231831
 500 2 1 1.0122569345
```

```
 500 3 1 1.0137880578
1000 1 1 1.007899174
1000 2 1 1.0080043459
1000 3 1 1.0074458755
2000 1 1 1.0064497658
2000 2 1 1.005433082
2000 3 1 1.0061964026
 150 1 2 1.0213788742
 150 2 2 1.0196160983
 150 3 2 1.0238464383
 500 1 2 1.013231831
 500 2 2 1.0122569345
 500 3 2 1.0137880578
1000 1 2 1.007899174
1000 2 2 1.0080043459
1000 3 2 1.0074458755
2000 1 2 1.0064497658
2000 2 2 1.005433082
2000 3 2 1.0061964026
 150 1 3 1.0213788742
 150 2 3 1.0196160983
 150 3 3 1.0238464383
 500 1 3 1.013231831
 500 2 3 1.0122569345
 500 3 3 1.0137880578
1000 1 3 1.007899174
1000 2 3 1.0080043459
1000 3 3 1.0074458755
2000 1 3 1.0064497658
2000 2 3 1.005433082
2000 3 3 1.0061964026
 150 1 4 1.0213788742
 150 2 4 1.0196160983
 150 3 4 1.0238464383
 500 1 4 1.013231831
 500 2 4 1.0122569345
 500 3 4 1.0137880578
1000 1 4 1.007899174
1000 2 4 1.0080043459
1000 3 4 1.0074458755
2000 1 4 1.0064497658
2000 2 4 1.005433082
2000 3 4 1.0061964026
;
run;
proc sort;
   by n r p;
run;

data all;
 merge d1 d2 d3 d4 nulls;
   by n r p;
   reject=0;
 if f gt wcrit then reject=1;
run;

proc freq;
```

```
 title 'CRITICAL VALUES';
   by n r;
   where p=1;
 table f;
run;

proc freq;
 title 'REJECTION RATES - Redo';
   by n r p;
 table reject ;
run;
```

Appendix H. Frequency Distribution of *W*-index – Simulated Null Condition

NULL DISTRIBUTIONS FOR EACH CELL OF THE EXPERIMENTAL DESIGN

---------- n=150 r=1 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9678607037	1	0.50	1	0.50
0.9702585889	1	0.50	2	1.00
0.9705550121	1	0.50	3	1.50
0.973800614	1	0.50	4	2.00
0.9766772449	1	0.50	5	2.50
0.9770294141	1	0.50	6	3.00
0.9771128119	1	0.50	7	3.50
0.9774853727	1	0.50	8	4.00
0.9776673858	1	0.50	9	4.50
0.9782251003	1	0.50	10	5.00
0.9810079153	1	0.50	11	5.50
0.9820104344	1	0.50	12	6.00
0.9824256144	1	0.50	13	6.50
0.9825898595	1	0.50	14	7.00
0.9836040811	1	0.50	15	7.50
0.9836133712	1	0.50	16	8.00
0.9837271063	1	0.50	17	8.50
0.9842680988	1	0.50	18	9.00
0.9845684825	1	0.50	19	9.50
0.9847164528	1	0.50	20	10.00
0.9847940263	1	0.50	21	10.50
0.9852903527	1	0.50	22	11.00
0.9854977723	1	0.50	23	11.50
0.9866566648	1	0.50	24	12.00
0.9871772976	1	0.50	25	12.50
0.9875992704	1	0.50	26	13.00
0.9876604206	1	0.50	27	13.50
0.9877027759	1	0.50	28	14.00
0.9880469567	1	0.50	29	14.50
0.9881935254	1	0.50	30	15.00
0.9882321333	1	0.50	31	15.50
0.9884408959	1	0.50	32	16.00
0.988931782	1	0.50	33	16.50
0.989281257	1	0.50	34	17.00
0.9894694683	1	0.50	35	17.50
0.9900411312	1	0.50	36	18.00
0.9906552232	1	0.50	37	18.50
0.990769053	1	0.50	38	19.00
0.990790643	1	0.50	39	19.50
0.9908362851	1	0.50	40	20.00
0.9909269528	1	0.50	41	20.50
0.9916605115	1	0.50	42	21.00
0.9919227464	1	0.50	43	21.50

---------- n=150 r=1 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9919547621	1	0.50	44	22.00
0.9920375977	1	0.50	45	22.50
0.9922828879	1	0.50	46	23.00
0.9922874678	1	0.50	47	23.50
0.9924995186	1	0.50	48	24.00
0.9928000185	1	0.50	49	24.50
0.9929522448	1	0.50	50	25.00
0.9933357729	1	0.50	51	25.50
0.9933806862	1	0.50	52	26.00
0.9937168233	1	0.50	53	26.50
0.9938376412	1	0.50	54	27.00
0.9938450071	1	0.50	55	27.50
0.9940788076	1	0.50	56	28.00
0.994162886	1	0.50	57	28.50
0.9941960868	1	0.50	58	29.00
0.9942179137	1	0.50	59	29.50
0.9947507496	1	0.50	60	30.00
0.9949426283	1	0.50	61	30.50
0.9950174963	1	0.50	62	31.00
0.9954709634	1	0.50	63	31.50
0.9955294357	1	0.50	64	32.00
0.9957050726	1	0.50	65	32.50
0.9958087363	1	0.50	66	33.00
0.9958602566	1	0.50	67	33.50
0.9960392284	1	0.50	68	34.00
0.9961076939	1	0.50	69	34.50
0.9961096202	1	0.50	70	35.00
0.9965145406	1	0.50	71	35.50
0.9970472826	1	0.50	72	36.00
0.9971870587	1	0.50	73	36.50
0.9973515406	1	0.50	74	37.00
0.9975096741	1	0.50	75	37.50
0.9976555538	1	0.50	76	38.00
0.9980975513	1	0.50	77	38.50
0.9986078311	1	0.50	78	39.00
0.9986389691	1	0.50	79	39.50
0.9992635901	1	0.50	80	40.00
0.9994025975	1	0.50	81	40.50
0.9994472143	1	0.50	82	41.00
1.0000435788	1	0.50	83	41.50
1.0003566167	1	0.50	84	42.00
1.0004716761	1	0.50	85	42.50
1.0008422301	1	0.50	86	43.00
1.0012231607	1	0.50	87	43.50
1.001686561	1	0.50	88	44.00
1.001741049	1	0.50	89	44.50
1.0021155379	1	0.50	90	45.00
1.0022419608	1	0.50	91	45.50
1.0026407888	1	0.50	92	46.00
1.0027242507	1	0.50	93	46.50

---------- n=150 r=1 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.0028206883	1	0.50	94	47.00
1.002854556	1	0.50	95	47.50
1.0028890205	1	0.50	96	48.00
1.0030225611	1	0.50	97	48.50
1.0030452381	1	0.50	98	49.00
1.0032989443	1	0.50	99	49.50
1.0033733322	1	0.50	100	50.00
1.0034820051	1	0.50	101	50.50
1.0034927798	1	0.50	102	51.00
1.0037919997	1	0.50	103	51.50
1.003851742	1	0.50	104	52.00
1.0041626949	1	0.50	105	52.50
1.0042304637	1	0.50	106	53.00
1.0042314605	1	0.50	107	53.50
1.0042966614	1	0.50	108	54.00
1.0043442583	1	0.50	109	54.50
1.0046438506	1	0.50	110	55.00
1.0049980129	1	0.50	111	55.50
1.0053287531	1	0.50	112	56.00
1.0053467097	1	0.50	113	56.50
1.0055498173	1	0.50	114	57.00
1.0055688584	1	0.50	115	57.50
1.0057819557	1	0.50	116	58.00
1.0059604338	1	0.50	117	58.50
1.0062174697	1	0.50	118	59.00
1.006269007	1	0.50	119	59.50
1.0062954422	1	0.50	120	60.00
1.0064733957	1	0.50	121	60.50
1.0065349654	1	0.50	122	61.00
1.0066884174	1	0.50	123	61.50
1.0070230548	1	0.50	124	62.00
1.007156872	1	0.50	125	62.50
1.0072952017	1	0.50	126	63.00
1.0074739802	1	0.50	127	63.50
1.0075075052	1	0.50	128	64.00
1.0078501634	1	0.50	129	64.50
1.0078951931	1	0.50	130	65.00
1.0083137103	1	0.50	131	65.50
1.0083362532	1	0.50	132	66.00
1.0085245567	1	0.50	133	66.50
1.0085692562	1	0.50	134	67.00
1.0088821636	1	0.50	135	67.50
1.008981829	1	0.50	136	68.00
1.009000934	1	0.50	137	68.50
1.0092961287	1	0.50	138	69.00
1.0095349417	1	0.50	139	69.50
1.0095624951	1	0.50	140	70.00
1.0097119135	1	0.50	141	70.50
1.0097210636	1	0.50	142	71.00
1.0097972451	1	0.50	143	71.50
1.0098245488	1	0.50	144	72.00
1.0101265866	1	0.50	145	72.50

---------- n=150 r=1 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.0101444842	1	0.50	146	73.00
1.0102182032	1	0.50	147	73.50
1.010347912	1	0.50	148	74.00
1.0104915211	1	0.50	149	74.50
1.0106447708	1	0.50	150	75.00
1.011386704	1	0.50	151	75.50
1.0116180662	1	0.50	152	76.00
1.0120211914	1	0.50	153	76.50
1.0121014629	1	0.50	154	77.00
1.0121679186	1	0.50	155	77.50
1.0124671276	1	0.50	156	78.00
1.0125496819	1	0.50	157	78.50
1.0126351568	1	0.50	158	79.00
1.0130826276	1	0.50	159	79.50
1.013185054	1	0.50	160	80.00
1.0133585386	1	0.50	161	80.50
1.0134752105	1	0.50	162	81.00
1.0135021346	1	0.50	163	81.50
1.0141087277	1	0.50	164	82.00
1.0142392118	1	0.50	165	82.50
1.0144090601	1	0.50	166	83.00
1.0146212322	1	0.50	167	83.50
1.0146251598	1	0.50	168	84.00
1.0148060006	1	0.50	169	84.50
1.0152004058	1	0.50	170	85.00
1.0152675071	1	0.50	171	85.50
1.0154105838	1	0.50	172	86.00
1.0156270113	1	0.50	173	86.50
1.0156898355	1	0.50	174	87.00
1.0160682393	1	0.50	175	87.50
1.0163283232	1	0.50	176	88.00
1.0165451583	1	0.50	177	88.50
1.0166976258	1	0.50	178	89.00
1.0168075685	1	0.50	179	89.50
1.0172213517	1	0.50	180	90.00
1.0173662823	1	0.50	181	90.50
1.017930144	1	0.50	182	91.00
1.0183206181	1	0.50	183	91.50
1.0185785725	1	0.50	184	92.00
1.0194197608	1	0.50	185	92.50
1.0196541515	1	0.50	186	93.00
1.0200678846	1	0.50	187	93.50
1.0206625694	1	0.50	188	94.00
1.0210652273	1	0.50	189	94.50
1.0213788742	1	0.50	190	95.00
1.0223335512	1	0.50	191	95.50
1.0241287541	1	0.50	192	96.00
1.0249349997	1	0.50	193	96.50
1.0252137342	1	0.50	194	97.00
1.026718264	1	0.50	195	97.50
1.0280414027	1	0.50	196	98.00

---------- n=150 r=1 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.0291432198	1	0.50	197	98.50
1.0308726592	1	0.50	198	99.00
1.0316199364	1	0.50	199	99.50
1.0322239861	1	0.50	200	100.00

---------- n=150 r=2 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9589713934	1	0.50	1	0.50
0.9704743139	1	0.50	2	1.00
0.9739070517	1	0.50	3	1.50
0.9739429152	1	0.50	4	2.00
0.9743310346	1	0.50	5	2.50
0.9744871411	1	0.50	6	3.00
0.9746300518	1	0.50	7	3.50
0.9750441489	1	0.50	8	4.00
0.9767200223	1	0.50	9	4.50
0.9768610809	1	0.50	10	5.00
0.97735553	1	0.50	11	5.50
0.9777745667	1	0.50	12	6.00
0.9779757473	1	0.50	13	6.50
0.9783883863	1	0.50	14	7.00
0.9796063964	1	0.50	15	7.50
0.9796692069	1	0.50	16	8.00
0.9800772304	1	0.50	17	8.50
0.9809723216	1	0.50	18	9.00
0.9817116928	1	0.50	19	9.50
0.9818698386	1	0.50	20	10.00
0.9819495619	1	0.50	21	10.50
0.9821284914	1	0.50	22	11.00
0.9822024045	1	0.50	23	11.50
0.9825211813	1	0.50	24	12.00
0.9825376795	1	0.50	25	12.50
0.982659004	1	0.50	26	13.00
0.9826789786	1	0.50	27	13.50
0.9838422786	1	0.50	28	14.00
0.9839171798	1	0.50	29	14.50
0.9854663176	1	0.50	30	15.00
0.9856447263	1	0.50	31	15.50
0.9856972569	1	0.50	32	16.00
0.9859328797	1	0.50	33	16.50
0.985981187	1	0.50	34	17.00
0.9861036701	1	0.50	35	17.50
0.98615911	1	0.50	36	18.00
0.9861755625	1	0.50	37	18.50
0.9864494145	1	0.50	38	19.00
0.9867870938	1	0.50	39	19.50
0.9869168836	1	0.50	40	20.00
0.987080964	1	0.50	41	20.50
0.9871175092	1	0.50	42	21.00

---------- n=150 r=2 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9874952525	1	0.50	43	21.50
0.9875571041	1	0.50	44	22.00
0.9878559941	1	0.50	45	22.50
0.9882878754	1	0.50	46	23.00
0.9885375843	1	0.50	47	23.50
0.9887783609	1	0.50	48	24.00
0.9887995294	1	0.50	49	24.50
0.988801335	1	0.50	50	25.00
0.9888978484	1	0.50	51	25.50
0.9890343432	1	0.50	52	26.00
0.9890992367	1	0.50	53	26.50
0.9895159111	1	0.50	54	27.00
0.9901422812	1	0.50	55	27.50
0.9905205293	1	0.50	56	28.00
0.9905950712	1	0.50	57	28.50
0.9907337933	1	0.50	58	29.00
0.9908865199	1	0.50	59	29.50
0.9915167692	1	0.50	60	30.00
0.9916630568	1	0.50	61	30.50
0.9918953357	1	0.50	62	31.00
0.9920732314	1	0.50	63	31.50
0.992291267	1	0.50	64	32.00
0.992365882	1	0.50	65	32.50
0.9925786385	1	0.50	66	33.00
0.9925838468	1	0.50	67	33.50
0.9926036874	1	0.50	68	34.00
0.9929117084	1	0.50	69	34.50
0.9930072625	1	0.50	70	35.00
0.9930328442	1	0.50	71	35.50
0.9931172331	1	0.50	72	36.00
0.9938891858	1	0.50	73	36.50
0.9940239035	1	0.50	74	37.00
0.9940639935	1	0.50	75	37.50
0.9942645712	1	0.50	76	38.00
0.994279609	1	0.50	77	38.50
0.9945221303	1	0.50	78	39.00
0.9945907067	1	0.50	79	39.50
0.9949274788	1	0.50	80	40.00
0.9953406262	1	0.50	81	40.50
0.9955340664	1	0.50	82	41.00
0.9958680581	1	0.50	83	41.50
0.9959987086	1	0.50	84	42.00
0.9960667141	1	0.50	85	42.50
0.9962828751	1	0.50	86	43.00

---------- n=150 r=2 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9964013725	1	0.50	87	43.50
0.9966299123	1	0.50	88	44.00
0.9968231576	1	0.50	89	44.50
0.9968409345	1	0.50	90	45.00
0.9970136028	1	0.50	91	45.50
0.9970256293	1	0.50	92	46.00
0.9975538366	1	0.50	93	46.50
0.9975810915	1	0.50	94	47.00
0.9978349684	1	0.50	95	47.50
0.997956125	1	0.50	96	48.00
0.9981199508	1	0.50	97	48.50
0.99821465	1	0.50	98	49.00
0.9982699308	1	0.50	99	49.50
0.9982907401	1	0.50	100	50.00
0.9983008502	1	0.50	101	50.50
0.9984246534	1	0.50	102	51.00
0.9984575543	1	0.50	103	51.50
0.9985469702	1	0.50	104	52.00
0.9987188717	1	0.50	105	52.50
0.9987438186	1	0.50	106	53.00
0.9991909794	1	0.50	107	53.50
0.9992538461	1	0.50	108	54.00
0.9994663397	1	0.50	109	54.50
1.000045637	1	0.50	110	55.00
1.0005253368	1	0.50	111	55.50
1.0005372283	1	0.50	112	56.00
1.0006187123	1	0.50	113	56.50
1.0007285436	1	0.50	114	57.00
1.0007300181	1	0.50	115	57.50
1.0007377401	1	0.50	116	58.00
1.0007627807	1	0.50	117	58.50
1.000792158	1	0.50	118	59.00
1.000914811	1	0.50	119	59.50
1.0009368742	1	0.50	120	60.00
1.0010452605	1	0.50	121	60.50
1.0017619014	1	0.50	122	61.00
1.0020137323	1	0.50	123	61.50
1.0020658518	1	0.50	124	62.00
1.0021086694	1	0.50	125	62.50
1.0026265549	1	0.50	126	63.00
1.0027081144	1	0.50	127	63.50
1.0027243373	1	0.50	128	64.00
1.0027286154	1	0.50	129	64.50

---------- n=150 r=2 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.0030694118	1	0.50	130	65.00
1.0034226068	1	0.50	131	65.50
1.0036449127	1	0.50	132	66.00
1.0037395863	1	0.50	133	66.50
1.0038159773	1	0.50	134	67.00
1.0040776586	1	0.50	135	67.50
1.0042126891	1	0.50	136	68.00
1.0043037874	1	0.50	137	68.50
1.0045445214	1	0.50	138	69.00
1.0047262843	1	0.50	139	69.50
1.0048400868	1	0.50	140	70.00
1.0052810753	1	0.50	141	70.50
1.0058845751	1	0.50	142	71.00
1.0059060504	1	0.50	143	71.50
1.0067692567	1	0.50	144	72.00
1.0070043299	1	0.50	145	72.50
1.0072364785	1	0.50	146	73.00
1.0073643071	1	0.50	147	73.50
1.0083608759	1	0.50	148	74.00
1.0085212585	1	0.50	149	74.50
1.0086220428	1	0.50	150	75.00
1.0086955943	1	0.50	151	75.50
1.008756627	1	0.50	152	76.00
1.009252357	1	0.50	153	76.50
1.0098549569	1	0.50	154	77.00
1.0102466246	1	0.50	155	77.50
1.0103019351	1	0.50	156	78.00
1.0108980924	1	0.50	157	78.50
1.011072332	1	0.50	158	79.00
1.011535112	1	0.50	159	79.50
1.011545324	1	0.50	160	80.00
1.011590035	1	0.50	161	80.50
1.0116723327	1	0.50	162	81.00
1.0118542666	1	0.50	163	81.50
1.0120420135	1	0.50	164	82.00
1.0122386192	1	0.50	165	82.50
1.0126384938	1	0.50	166	83.00
1.0126644809	1	0.50	167	83.50
1.0127877563	1	0.50	168	84.00
1.0129257116	1	0.50	169	84.50
1.0132400842	1	0.50	170	85.00
1.0133608081	1	0.50	171	85.50
1.0133627548	1	0.50	172	86.00

---------- n=150 r=2 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.0134498536	1	0.50	173	86.50
1.0134796398	1	0.50	174	87.00
1.0136530391	1	0.50	175	87.50
1.0137468426	1	0.50	176	88.00
1.0142586324	1	0.50	177	88.50
1.014620416	1	0.50	178	89.00
1.01494913	1	0.50	179	89.50
1.0153007692	1	0.50	180	90.00
1.015612225	1	0.50	181	90.50
1.0157866904	1	0.50	182	91.00
1.0160628729	1	0.50	183	91.50
1.0166507191	1	0.50	184	92.00
1.0167357829	1	0.50	185	92.50
1.016843552	1	0.50	186	93.00
1.0168787448	1	0.50	187	93.50
1.0179264501	1	0.50	188	94.00
1.0195488785	1	0.50	189	94.50
1.0196160983	1	0.50	190	95.00
1.0206988269	1	0.50	191	95.50
1.0211541153	1	0.50	192	96.00
1.0227518633	1	0.50	193	96.50
1.0229753875	1	0.50	194	97.00
1.0255312333	1	0.50	195	97.50
1.0267698978	1	0.50	196	98.00
1.0271058194	1	0.50	197	98.50
1.027984828	1	0.50	198	99.00
1.0289525702	1	0.50	199	99.50
1.0363132228	1	0.50	200	100.00

---------- n=150 r=3 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9551541812	1	0.50	1	0.50
0.9679143586	1	0.50	2	1.00
0.9682774148	1	0.50	3	1.50
0.9685937966	1	0.50	4	2.00
0.970109428	1	0.50	5	2.50
0.9721560553	1	0.50	6	3.00
0.9727393914	1	0.50	7	3.50
0.972901928	1	0.50	8	4.00
0.9739372808	1	0.50	9	4.50
0.9753980613	1	0.50	10	5.00
0.9762357712	1	0.50	11	5.50
0.977174899	1	0.50	12	6.00
0.9773715512	1	0.50	13	6.50
0.978211212	1	0.50	14	7.00
0.9791660036	1	0.50	15	7.50

---------- n=150 r=3 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9792809749	1	0.50	16	8.00
0.9793230184	1	0.50	17	8.50
0.9795206252	1	0.50	18	9.00
0.9795424745	1	0.50	19	9.50
0.9800526069	1	0.50	20	10.00
0.9801038097	1	0.50	21	10.50
0.9817036885	1	0.50	22	11.00
0.981826329	1	0.50	23	11.50
0.9818500852	1	0.50	24	12.00
0.9819112294	1	0.50	25	12.50
0.9822676817	1	0.50	26	13.00
0.9828034413	1	0.50	27	13.50
0.9837535991	1	0.50	28	14.00
0.9841359631	1	0.50	29	14.50
0.9842259779	1	0.50	30	15.00
0.9843872765	1	0.50	31	15.50
0.9847384665	1	0.50	32	16.00
0.9848105377	1	0.50	33	16.50
0.984817726	1	0.50	34	17.00
0.9850994714	1	0.50	35	17.50
0.9857258567	1	0.50	36	18.00
0.9863946622	1	0.50	37	18.50
0.986964631	1	0.50	38	19.00
0.9870906324	1	0.50	39	19.50
0.98723711	1	0.50	40	20.00
0.9875812147	1	0.50	41	20.50
0.9875902579	1	0.50	42	21.00
0.9876807055	1	0.50	43	21.50
0.9881472142	1	0.50	44	22.00
0.9883404264	1	0.50	45	22.50
0.988412691	1	0.50	46	23.00
0.9889575132	1	0.50	47	23.50
0.9889748352	1	0.50	48	24.00
0.9895423013	1	0.50	49	24.50
0.9896335705	1	0.50	50	25.00
0.9899586699	1	0.50	51	25.50
0.9901600757	1	0.50	52	26.00
0.9902887359	1	0.50	53	26.50
0.990367678	1	0.50	54	27.00
0.9906363143	1	0.50	55	27.50
0.9908835339	1	0.50	56	28.00
0.990901388	1	0.50	57	28.50
0.9910945182	1	0.50	58	29.00
0.9911092215	1	0.50	59	29.50
0.9911346347	1	0.50	60	30.00
0.9914037442	1	0.50	61	30.50
0.9923667662	1	0.50	62	31.00
0.9926616413	1	0.50	63	31.50
0.9926955832	1	0.50	64	32.00
0.9929157886	1	0.50	65	32.50

---------- n=150 r=3 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9930829338	1	0.50	66	33.00
0.9936172497	1	0.50	67	33.50
0.9939169851	1	0.50	68	34.00
0.9940427272	1	0.50	69	34.50
0.994384488	1	0.50	70	35.00
0.9943870706	1	0.50	71	35.50
0.9951961442	1	0.50	72	36.00
0.9952275673	1	0.50	73	36.50
0.9954895477	1	0.50	74	37.00
0.9957501017	1	0.50	75	37.50
0.9959214808	1	0.50	76	38.00
0.995924382	1	0.50	77	38.50
0.9961039063	1	0.50	78	39.00
0.9962033417	1	0.50	79	39.50
0.9962982883	1	0.50	80	40.00
0.9965281903	1	0.50	81	40.50
0.9965667193	1	0.50	82	41.00
0.9969466539	1	0.50	83	41.50
0.9969546501	1	0.50	84	42.00
0.9970850926	1	0.50	85	42.50
0.9971252075	1	0.50	86	43.00
0.9973028976	1	0.50	87	43.50
0.9973599	1	0.50	88	44.00
0.9974491997	1	0.50	89	44.50
0.9975284154	1	0.50	90	45.00
0.9978172637	1	0.50	91	45.50
0.9978562615	1	0.50	92	46.00
0.9979043647	1	0.50	93	46.50
0.998167752	1	0.50	94	47.00
0.9982284731	1	0.50	95	47.50
0.998506932	1	0.50	96	48.00
0.9991271769	1	0.50	97	48.50
0.9994058178	1	0.50	98	49.00
0.9995971907	1	0.50	99	49.50
0.9996474434	1	0.50	100	50.00
0.9996827201	1	0.50	101	50.50
1.0002044283	1	0.50	102	51.00
1.0003143553	1	0.50	103	51.50
1.0014543258	1	0.50	104	52.00
1.0018680232	1	0.50	105	52.50
1.0019207824	1	0.50	106	53.00
1.002104897	1	0.50	107	53.50
1.0021487992	1	0.50	108	54.00
1.0022925411	1	0.50	109	54.50
1.002554319	1	0.50	110	55.00
1.0025860092	1	0.50	111	55.50
1.0027010831	1	0.50	112	56.00
1.0027549329	1	0.50	113	56.50
1.0032095725	1	0.50	114	57.00
1.003545535	1	0.50	115	57.50
1.0038850345	1	0.50	116	58.00

---------- n=150 r=3 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.0039510034	1	0.50	117	58.50
1.0043349982	1	0.50	118	59.00
1.00475371	1	0.50	119	59.50
1.0050072624	1	0.50	120	60.00
1.0051704305	1	0.50	121	60.50
1.005317443	1	0.50	122	61.00
1.0057428739	1	0.50	123	61.50
1.0061624251	1	0.50	124	62.00
1.0063187258	1	0.50	125	62.50
1.0063589707	1	0.50	126	63.00
1.0067536505	1	0.50	127	63.50
1.0068127938	1	0.50	128	64.00
1.0068157512	1	0.50	129	64.50
1.0069217924	1	0.50	130	65.00
1.0070707478	1	0.50	131	65.50
1.0071290482	1	0.50	132	66.00
1.007751171	1	0.50	133	66.50
1.0080506973	1	0.50	134	67.00
1.0080753511	1	0.50	135	67.50
1.0084500835	1	0.50	136	68.00
1.0085717928	1	0.50	137	68.50
1.0085887987	1	0.50	138	69.00
1.0086976252	1	0.50	139	69.50
1.0091890068	1	0.50	140	70.00
1.0092014893	1	0.50	141	70.50
1.0093087573	1	0.50	142	71.00
1.0093289219	1	0.50	143	71.50
1.0093389792	1	0.50	144	72.00
1.0094778823	1	0.50	145	72.50
1.0098248424	1	0.50	146	73.00
1.0098548928	1	0.50	147	73.50
1.0098877099	1	0.50	148	74.00
1.0099691962	1	0.50	149	74.50
1.0100193829	1	0.50	150	75.00
1.0102954985	1	0.50	151	75.50
1.0104113328	1	0.50	152	76.00
1.0107198813	1	0.50	153	76.50
1.0110298272	1	0.50	154	77.00
1.0115218071	1	0.50	155	77.50
1.0116542597	1	0.50	156	78.00
1.0117280734	1	0.50	157	78.50
1.0122240907	1	0.50	158	79.00
1.0128935041	1	0.50	159	79.50
1.0130519968	1	0.50	160	80.00
1.0132426288	1	0.50	161	80.50
1.0133279664	1	0.50	162	81.00
1.0135180219	1	0.50	163	81.50
1.0136874435	1	0.50	164	82.00
1.0143333428	1	0.50	165	82.50
1.0148264877	1	0.50	166	83.00
1.0150495605	1	0.50	167	83.50

---------- n=150 r=3 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.0155878045	1	0.50	168	84.00
1.0157873999	1	0.50	169	84.50
1.0163907669	1	0.50	170	85.00
1.0165702657	1	0.50	171	85.50
1.016820524	1	0.50	172	86.00
1.0170185749	1	0.50	173	86.50
1.0189016493	1	0.50	174	87.00
1.0190320429	1	0.50	175	87.50
1.0197421417	1	0.50	176	88.00
1.0198939904	1	0.50	177	88.50
1.0203430504	1	0.50	178	89.00
1.0205826446	1	0.50	179	89.50
1.020592934	1	0.50	180	90.00
1.0207274795	1	0.50	181	90.50
1.0208591019	1	0.50	182	91.00
1.0213916962	1	0.50	183	91.50
1.0215389301	1	0.50	184	92.00
1.0215621859	1	0.50	185	92.50
1.0218701839	1	0.50	186	93.00
1.0222108398	1	0.50	187	93.50
1.0227090184	1	0.50	188	94.00
1.023522183	1	0.50	189	94.50
1.0238464383	1	0.50	190	95.00
1.0241474584	1	0.50	191	95.50
1.0254859624	1	0.50	192	96.00
1.0277256647	1	0.50	193	96.50
1.0281015294	1	0.50	194	97.00
1.0281175847	1	0.50	195	97.50
1.0289508738	1	0.50	196	98.00
1.0291381066	1	0.50	197	98.50
1.0310381071	1	0.50	198	99.00
1.0335203619	1	0.50	199	99.50
1.0337504369	1	0.50	200	100.00

---------- n=500 r=1 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9822844724	1	0.50	1	0.50
0.9844725079	1	0.50	2	1.00
0.985293855	1	0.50	3	1.50
0.9868503184	1	0.50	4	2.00
0.9869136514	1	0.50	5	2.50
0.9874738104	1	0.50	6	3.00
0.9886181566	1	0.50	7	3.50
0.9889134191	1	0.50	8	4.00
0.9891038579	1	0.50	9	4.50
0.9893178116	1	0.50	10	5.00
0.9895156841	1	0.50	11	5.50
0.99006596	1	0.50	12	6.00

---------- n=500 r=1 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9901636833	1	0.50	13	6.50
0.9906287617	1	0.50	14	7.00
0.9908896191	1	0.50	15	7.50
0.9914144375	1	0.50	16	8.00
0.9915749478	1	0.50	17	8.50
0.9917752361	1	0.50	18	9.00
0.9918122494	1	0.50	19	9.50
0.9918167478	1	0.50	20	10.00
0.9919802343	1	0.50	21	10.50
0.9920062945	1	0.50	22	11.00
0.9920804966	1	0.50	23	11.50
0.9922794066	1	0.50	24	12.00
0.9923154648	1	0.50	25	12.50
0.9923164435	1	0.50	26	13.00
0.9923798391	1	0.50	27	13.50
0.9924393811	1	0.50	28	14.00
0.9924959793	1	0.50	29	14.50
0.9925484554	1	0.50	30	15.00
0.992663394	1	0.50	31	15.50
0.9927666296	1	0.50	32	16.00
0.9928893683	1	0.50	33	16.50
0.9930640857	1	0.50	34	17.00
0.9931136809	1	0.50	35	17.50
0.9932824405	1	0.50	36	18.00
0.9933367326	1	0.50	37	18.50
0.9934008254	1	0.50	38	19.00
0.9934569639	1	0.50	39	19.50
0.9936753012	1	0.50	40	20.00
0.9937806561	1	0.50	41	20.50
0.9938150735	1	0.50	42	21.00
0.994079423	1	0.50	43	21.50
0.9944704762	1	0.50	44	22.00
0.9946694779	1	0.50	45	22.50
0.9948609074	1	0.50	46	23.00
0.9948656204	1	0.50	47	23.50
0.9948684392	1	0.50	48	24.00
0.994897275	1	0.50	49	24.50
0.9949258496	1	0.50	50	25.00
0.9950027178	1	0.50	51	25.50
0.9951347206	1	0.50	52	26.00
0.995258477	1	0.50	53	26.50
0.9953829183	1	0.50	54	27.00
0.9955242048	1	0.50	55	27.50
0.9957430799	1	0.50	56	28.00
0.9957943042	1	0.50	57	28.50
0.9959296606	1	0.50	58	29.00
0.99630339	1	0.50	59	29.50
0.9965095293	1	0.50	60	30.00
0.9966142767	1	0.50	61	30.50
0.9967449894	1	0.50	62	31.00
0.9968163657	1	0.50	63	31.50
0.9969493662	1	0.50	64	32.00

---------- n=500 r=1 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9969779488	1	0.50	65	32.50
0.9970255924	1	0.50	66	33.00
0.9971296416	1	0.50	67	33.50
0.9973996649	1	0.50	68	34.00
0.9974238285	1	0.50	69	34.50
0.9974531996	1	0.50	70	35.00
0.9975173106	1	0.50	71	35.50
0.9976119632	1	0.50	72	36.00
0.9979553367	1	0.50	73	36.50
0.9979708072	1	0.50	74	37.00
0.9980072903	1	0.50	75	37.50
0.9980669863	1	0.50	76	38.00
0.9984371479	1	0.50	77	38.50
0.9984773793	1	0.50	78	39.00
0.9985147586	1	0.50	79	39.50
0.9986151343	1	0.50	80	40.00
0.9986849949	1	0.50	81	40.50
0.9986926502	1	0.50	82	41.00
0.998692784	1	0.50	83	41.50
0.9987008331	1	0.50	84	42.00
0.9988062547	1	0.50	85	42.50
0.9990764947	1	0.50	86	43.00
0.9990835206	1	0.50	87	43.50
0.9991318199	1	0.50	88	44.00
0.9992849508	1	0.50	89	44.50
0.999286428	1	0.50	90	45.00
0.99938631	1	0.50	91	45.50
0.9994228651	1	0.50	92	46.00
0.9994795168	1	0.50	93	46.50
0.9994911042	1	0.50	94	47.00
0.9995220095	1	0.50	95	47.50
0.9995645356	1	0.50	96	48.00
0.9995690235	1	0.50	97	48.50
0.999704784	1	0.50	98	49.00
0.9997523077	1	0.50	99	49.50
0.9998043126	1	0.50	100	50.00
1.0000221476	1	0.50	101	50.50
1.0000900436	1	0.50	102	51.00
1.0001015638	1	0.50	103	51.50
1.0001925933	1	0.50	104	52.00
1.0004574207	1	0.50	105	52.50
1.0005652951	1	0.50	106	53.00
1.0005957749	1	0.50	107	53.50
1.0007842903	1	0.50	108	54.00
1.0007959834	1	0.50	109	54.50
1.0008605401	1	0.50	110	55.00
1.0009142797	1	0.50	111	55.50
1.0009498325	1	0.50	112	56.00
1.000968421	1	0.50	113	56.50
1.0011241334	1	0.50	114	57.00
1.0011282948	1	0.50	115	57.50

---------- n=500 r=1 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.0011552433	1	0.50	116	58.00
1.0011736343	1	0.50	117	58.50
1.0011835086	1	0.50	118	59.00
1.0012727579	1	0.50	119	59.50
1.0013708899	1	0.50	120	60.00
1.0014729557	1	0.50	121	60.50
1.0014895772	1	0.50	122	61.00
1.0015380893	1	0.50	123	61.50
1.0016527396	1	0.50	124	62.00
1.0017371489	1	0.50	125	62.50
1.0017633445	1	0.50	126	63.00
1.0018702716	1	0.50	127	63.50
1.0020859993	1	0.50	128	64.00
1.0021491612	1	0.50	129	64.50
1.0021998786	1	0.50	130	65.00
1.0024434838	1	0.50	131	65.50
1.0024992548	1	0.50	132	66.00
1.0025858042	1	0.50	133	66.50
1.0026030185	1	0.50	134	67.00
1.0026640865	1	0.50	135	67.50
1.0027240674	1	0.50	136	68.00
1.0028453598	1	0.50	137	68.50
1.0029881502	1	0.50	138	69.00
1.0029979893	1	0.50	139	69.50
1.0031458452	1	0.50	140	70.00
1.003163304	1	0.50	141	70.50
1.0032150784	1	0.50	142	71.00
1.0034029592	1	0.50	143	71.50
1.0034132829	1	0.50	144	72.00
1.0034848005	1	0.50	145	72.50
1.0035135198	1	0.50	146	73.00
1.0035946391	1	0.50	147	73.50
1.003686034	1	0.50	148	74.00
1.0038494513	1	0.50	149	74.50
1.0040488027	1	0.50	150	75.00
1.0040872205	1	0.50	151	75.50
1.0041852584	1	0.50	152	76.00
1.0044021355	1	0.50	153	76.50
1.0044766011	1	0.50	154	77.00
1.0047436791	1	0.50	155	77.50
1.0047450105	1	0.50	156	78.00
1.0049442419	1	0.50	157	78.50
1.0051404729	1	0.50	158	79.00
1.0052108985	1	0.50	159	79.50
1.0054334538	1	0.50	160	80.00
1.0055074039	1	0.50	161	80.50
1.0055235241	1	0.50	162	81.00
1.0058693121	1	0.50	163	81.50
1.0059893338	1	0.50	164	82.00
1.0060992275	1	0.50	165	82.50
1.0061693865	1	0.50	166	83.00

---------- n=500 r=1 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.0062234777	1	0.50	167	83.50
1.0062791621	1	0.50	168	84.00
1.0064274939	1	0.50	169	84.50
1.0064303764	1	0.50	170	85.00
1.0064831494	1	0.50	171	85.50
1.0065186028	1	0.50	172	86.00
1.0065203972	1	0.50	173	86.50
1.0069047855	1	0.50	174	87.00
1.0069981601	1	0.50	175	87.50
1.0073529381	1	0.50	176	88.00
1.0076337502	1	0.50	177	88.50
1.0080384862	1	0.50	178	89.00
1.0081841801	1	0.50	179	89.50
1.0084316802	1	0.50	180	90.00
1.0086533425	1	0.50	181	90.50
1.0092850403	1	0.50	182	91.00
1.0093368366	1	0.50	183	91.50
1.0099795778	1	0.50	184	92.00
1.0101072244	1	0.50	185	92.50
1.0107354396	1	0.50	186	93.00
1.0113303684	1	0.50	187	93.50
1.0117790697	1	0.50	188	94.00
1.0126391786	1	0.50	189	94.50
1.013231831	1	0.50	190	95.00
1.0136822758	1	0.50	191	95.50
1.0137086564	1	0.50	192	96.00
1.0137141982	1	0.50	193	96.50
1.0146324022	1	0.50	194	97.00
1.0148030531	1	0.50	195	97.50
1.0158105195	1	0.50	196	98.00
1.0163421145	1	0.50	197	98.50
1.0167298185	1	0.50	198	99.00
1.0190537544	1	0.50	199	99.50
1.0204587401	1	0.50	200	100.00

---------- n=500 r=2 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9799611331	1	0.50	1	0.50
0.9811794861	1	0.50	2	1.00
0.9812218923	1	0.50	3	1.50
0.9821106029	1	0.50	4	2.00
0.9824744986	1	0.50	5	2.50
0.9850936671	1	0.50	6	3.00
0.9856002135	1	0.50	7	3.50
0.9861392251	1	0.50	8	4.00
0.9861400977	1	0.50	9	4.50
0.9868089754	1	0.50	10	5.00
0.9869591938	1	0.50	11	5.50

---------- n=500 r=2 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9884857191	1	0.50	12	6.00
0.9885076556	1	0.50	13	6.50
0.9888949717	1	0.50	14	7.00
0.989056419	1	0.50	15	7.50
0.9893029628	1	0.50	16	8.00
0.9895294692	1	0.50	17	8.50
0.9899977691	1	0.50	18	9.00
0.9905268445	1	0.50	19	9.50
0.9907478782	1	0.50	20	10.00
0.9909742421	1	0.50	21	10.50
0.9910271158	1	0.50	22	11.00
0.9910947295	1	0.50	23	11.50
0.9911074865	1	0.50	24	12.00
0.9912623767	1	0.50	25	12.50
0.9913901684	1	0.50	26	13.00
0.9916399975	1	0.50	27	13.50
0.991665995	1	0.50	28	14.00
0.9916746537	1	0.50	29	14.50
0.9917353579	1	0.50	30	15.00
0.9920611994	1	0.50	31	15.50
0.9921476743	1	0.50	32	16.00
0.9926892166	1	0.50	33	16.50
0.9930439326	1	0.50	34	17.00
0.9936026001	1	0.50	35	17.50
0.9936557576	1	0.50	36	18.00
0.9937141153	1	0.50	37	18.50
0.9937243124	1	0.50	38	19.00
0.9939068958	1	0.50	39	19.50
0.9939574779	1	0.50	40	20.00
0.9941624915	1	0.50	41	20.50
0.9941685106	1	0.50	42	21.00
0.9942413867	1	0.50	43	21.50
0.9945554137	1	0.50	44	22.00
0.9945622664	1	0.50	45	22.50
0.9954917029	1	0.50	46	23.00
0.9956791237	1	0.50	47	23.50
0.9957305081	1	0.50	48	24.00
0.9957621547	1	0.50	49	24.50
0.9957662466	1	0.50	50	25.00
0.9958993827	1	0.50	51	25.50
0.9959487941	1	0.50	52	26.00
0.9961193535	1	0.50	53	26.50
0.9961217576	1	0.50	54	27.00
0.9963586164	1	0.50	55	27.50
0.9963739811	1	0.50	56	28.00
0.9966159704	1	0.50	57	28.50
0.9968668165	1	0.50	58	29.00
0.9969498463	1	0.50	59	29.50
0.9969590192	1	0.50	60	30.00
0.9970167176	1	0.50	61	30.50
0.9972220912	1	0.50	62	31.00

---------- n=500 r=2 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9973451004	1	0.50	63	31.50
0.9974023602	1	0.50	64	32.00
0.9974210961	1	0.50	65	32.50
0.9974698587	1	0.50	66	33.00
0.9975272111	1	0.50	67	33.50
0.9975383429	1	0.50	68	34.00
0.9977028546	1	0.50	69	34.50
0.9977231238	1	0.50	70	35.00
0.9979499413	1	0.50	71	35.50
0.9980967888	1	0.50	72	36.00
0.9981671376	1	0.50	73	36.50
0.9981717702	1	0.50	74	37.00
0.9981918739	1	0.50	75	37.50
0.9982527426	1	0.50	76	38.00
0.998368491	1	0.50	77	38.50
0.9984185286	1	0.50	78	39.00
0.9984287776	1	0.50	79	39.50
0.9984752211	1	0.50	80	40.00
0.9985658305	1	0.50	81	40.50
0.9986641251	1	0.50	82	41.00
0.9987195236	1	0.50	83	41.50
0.9987569554	1	0.50	84	42.00
0.9988413178	1	0.50	85	42.50
0.9988725313	1	0.50	86	43.00
0.9989026408	1	0.50	87	43.50
0.9991493822	1	0.50	88	44.00
0.9995321612	1	0.50	89	44.50
0.9995548279	1	0.50	90	45.00
0.9997690207	1	0.50	91	45.50
1.000026971	1	0.50	92	46.00
1.0001372678	1	0.50	93	46.50
1.0001457425	1	0.50	94	47.00
1.0001708477	1	0.50	95	47.50
1.0002102001	1	0.50	96	48.00
1.0002996668	1	0.50	97	48.50
1.0003463865	1	0.50	98	49.00
1.0004768805	1	0.50	99	49.50
1.0005362565	1	0.50	100	50.00
1.0006026293	1	0.50	101	50.50
1.0006750983	1	0.50	102	51.00
1.0007380645	1	0.50	103	51.50
1.0008062613	1	0.50	104	52.00
1.0008097785	1	0.50	105	52.50
1.0009061719	1	0.50	106	53.00
1.0009178316	1	0.50	107	53.50
1.0009206313	1	0.50	108	54.00
1.0009382786	1	0.50	109	54.50
1.0012307846	1	0.50	110	55.00
1.0012373561	1	0.50	111	55.50
1.0014349623	1	0.50	112	56.00
1.0014554194	1	0.50	113	56.50

---------- n=500 r=2 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.0015398914	1	0.50	114	57.00
1.0016805626	1	0.50	115	57.50
1.0017370977	1	0.50	116	58.00
1.0018197734	1	0.50	117	58.50
1.0018484263	1	0.50	118	59.00
1.0019814061	1	0.50	119	59.50
1.0021595702	1	0.50	120	60.00
1.0023597311	1	0.50	121	60.50
1.0023787068	1	0.50	122	61.00
1.0023843379	1	0.50	123	61.50
1.0023936969	1	0.50	124	62.00
1.0024010989	1	0.50	125	62.50
1.0024272515	1	0.50	126	63.00
1.0026711233	1	0.50	127	63.50
1.002712712	1	0.50	128	64.00
1.0027174047	1	0.50	129	64.50
1.002778809	1	0.50	130	65.00
1.0028837142	1	0.50	131	65.50
1.0029571581	1	0.50	132	66.00
1.0030315746	1	0.50	133	66.50
1.0030408313	1	0.50	134	67.00
1.0031087596	1	0.50	135	67.50
1.003165908	1	0.50	136	68.00
1.0032047374	1	0.50	137	68.50
1.003282603	1	0.50	138	69.00
1.0033119696	1	0.50	139	69.50
1.003358846	1	0.50	140	70.00
1.0033877721	1	0.50	141	70.50
1.0033986833	1	0.50	142	71.00
1.0035332234	1	0.50	143	71.50
1.0037190687	1	0.50	144	72.00
1.0037864437	1	0.50	145	72.50
1.0040554462	1	0.50	146	73.00
1.0043467637	1	0.50	147	73.50
1.0044269441	1	0.50	148	74.00
1.0044573689	1	0.50	149	74.50
1.0045443407	1	0.50	150	75.00
1.0045639254	1	0.50	151	75.50
1.0046214819	1	0.50	152	76.00
1.0046685379	1	0.50	153	76.50
1.0046818052	1	0.50	154	77.00
1.0050557916	1	0.50	155	77.50
1.0050626989	1	0.50	156	78.00
1.0053157031	1	0.50	157	78.50
1.0056257371	1	0.50	158	79.00
1.0058039469	1	0.50	159	79.50
1.0059188598	1	0.50	160	80.00
1.0059859713	1	0.50	161	80.50
1.005987723	1	0.50	162	81.00
1.0060861009	1	0.50	163	81.50
1.0060897955	1	0.50	164	82.00

---------- n=500 r=2 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.0061929258	1	0.50	165	82.50
1.0064188936	1	0.50	166	83.00
1.0064617688	1	0.50	167	83.50
1.0064873311	1	0.50	168	84.00
1.0068258792	1	0.50	169	84.50
1.0069820667	1	0.50	170	85.00
1.0073006385	1	0.50	171	85.50
1.007724876	1	0.50	172	86.00
1.0078818841	1	0.50	173	86.50
1.0079457543	1	0.50	174	87.00
1.0080740496	1	0.50	175	87.50
1.0084371699	1	0.50	176	88.00
1.0084585928	1	0.50	177	88.50
1.0087285298	1	0.50	178	89.00
1.0090993618	1	0.50	179	89.50
1.0092058615	1	0.50	180	90.00
1.0092065682	1	0.50	181	90.50
1.0096991563	1	0.50	182	91.00
1.0100846583	1	0.50	183	91.50
1.0101668317	1	0.50	184	92.00
1.0102865342	1	0.50	185	92.50
1.0108745751	1	0.50	186	93.00
1.0111614098	1	0.50	187	93.50
1.0114442436	1	0.50	188	94.00
1.0119087225	1	0.50	189	94.50
1.0122569345	1	0.50	190	95.00
1.0123256499	1	0.50	191	95.50
1.012781189	1	0.50	192	96.00
1.013027363	1	0.50	193	96.50
1.0134041651	1	0.50	194	97.00
1.013499946	1	0.50	195	97.50
1.0143608283	1	0.50	196	98.00
1.0158295357	1	0.50	197	98.50
1.0160852218	1	0.50	198	99.00
1.0165347729	1	0.50	199	99.50
1.0170125333	1	0.50	200	100.00

---------- n=500 r=3 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9718425225	1	0.50	1	0.50
0.9809698454	1	0.50	2	1.00
0.9821317345	1	0.50	3	1.50
0.9831181104	1	0.50	4	2.00
0.9838176693	1	0.50	5	2.50
0.9849619661	1	0.50	6	3.00
0.9855434212	1	0.50	7	3.50
0.9856006211	1	0.50	8	4.00
0.9857155587	1	0.50	9	4.50
0.9867421139	1	0.50	10	5.00

---------- n=500 r=3 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9878641809	1	0.50	11	5.50
0.987886383	1	0.50	12	6.00
0.9879059687	1	0.50	13	6.50
0.9879555189	1	0.50	14	7.00
0.9880806551	1	0.50	15	7.50
0.9886243956	1	0.50	16	8.00
0.9888779351	1	0.50	17	8.50
0.9891873389	1	0.50	18	9.00
0.989228151	1	0.50	19	9.50
0.9893968407	1	0.50	20	10.00
0.9894249301	1	0.50	21	10.50
0.9895351787	1	0.50	22	11.00
0.9897159226	1	0.50	23	11.50
0.9899451568	1	0.50	24	12.00
0.9905971923	1	0.50	25	12.50
0.9907335774	1	0.50	26	13.00
0.9910249745	1	0.50	27	13.50
0.9912458895	1	0.50	28	14.00
0.9912806736	1	0.50	29	14.50
0.9914205188	1	0.50	30	15.00
0.9916491078	1	0.50	31	15.50
0.9919760406	1	0.50	32	16.00
0.9920475269	1	0.50	33	16.50
0.9921512618	1	0.50	34	17.00
0.9921660338	1	0.50	35	17.50
0.9923819494	1	0.50	36	18.00
0.9923970332	1	0.50	37	18.50
0.9925371527	1	0.50	38	19.00
0.9925702081	1	0.50	39	19.50
0.992720831	1	0.50	40	20.00
0.9929641686	1	0.50	41	20.50
0.993048245	1	0.50	42	21.00
0.9931042421	1	0.50	43	21.50
0.9931593458	1	0.50	44	22.00
0.9935154287	1	0.50	45	22.50
0.9935806757	1	0.50	46	23.00
0.9936063059	1	0.50	47	23.50
0.9939870838	1	0.50	48	24.00
0.9944279508	1	0.50	49	24.50
0.9944689277	1	0.50	50	25.00
0.9946437297	1	0.50	51	25.50
0.9949134955	1	0.50	52	26.00
0.994937132	1	0.50	53	26.50
0.9951518151	1	0.50	54	27.00
0.99527371	1	0.50	55	27.50
0.9953131147	1	0.50	56	28.00
0.995336989	1	0.50	57	28.50
0.9955754209	1	0.50	58	29.00
0.9956720611	1	0.50	59	29.50
0.9957575723	1	0.50	60	30.00
0.9958258182	1	0.50	61	30.50

---------- n=500 r=3 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9959984091	1	0.50	62	31.00
0.9962852372	1	0.50	63	31.50
0.9963005635	1	0.50	64	32.00
0.996523861	1	0.50	65	32.50
0.9965318302	1	0.50	66	33.00
0.9965922198	1	0.50	67	33.50
0.9968149213	1	0.50	68	34.00
0.9968336037	1	0.50	69	34.50
0.9969192537	1	0.50	70	35.00
0.9970089755	1	0.50	71	35.50
0.9970842005	1	0.50	72	36.00
0.9971997327	1	0.50	73	36.50
0.9973287271	1	0.50	74	37.00
0.9975110452	1	0.50	75	37.50
0.9975125056	1	0.50	76	38.00
0.9976020409	1	0.50	77	38.50
0.9976384117	1	0.50	78	39.00
0.9976590161	1	0.50	79	39.50
0.9977215325	1	0.50	80	40.00
0.9977523235	1	0.50	81	40.50
0.9977821497	1	0.50	82	41.00
0.9977869653	1	0.50	83	41.50
0.9978967023	1	0.50	84	42.00
0.9979336024	1	0.50	85	42.50
0.9979742021	1	0.50	86	43.00
0.9979768188	1	0.50	87	43.50
0.9980078137	1	0.50	88	44.00
0.9980231976	1	0.50	89	44.50
0.9981431649	1	0.50	90	45.00
0.9984302707	1	0.50	91	45.50
0.9985983628	1	0.50	92	46.00
0.9986260034	1	0.50	93	46.50
0.9986347235	1	0.50	94	47.00
0.9986942356	1	0.50	95	47.50
0.9987539055	1	0.50	96	48.00
0.9988941037	1	0.50	97	48.50
0.9990389771	1	0.50	98	49.00
0.9992398995	1	0.50	99	49.50
0.9993387829	1	0.50	100	50.00
0.9996063884	1	0.50	101	50.50
0.9996707786	1	0.50	102	51.00
0.9996915361	1	0.50	103	51.50
0.9997287515	1	0.50	104	52.00
0.9998001756	1	0.50	105	52.50
0.9998892741	1	0.50	106	53.00
1.0000501458	1	0.50	107	53.50
1.0001757966	1	0.50	108	54.00
1.0001849982	1	0.50	109	54.50
1.0005868435	1	0.50	110	55.00
1.0006723225	1	0.50	111	55.50
1.0007984583	1	0.50	112	56.00

---------- n=500 r=3 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.0008085249	1	0.50	113	56.50
1.0008154713	1	0.50	114	57.00
1.0008961495	1	0.50	115	57.50
1.000960098	1	0.50	116	58.00
1.0009718673	1	0.50	117	58.50
1.0012973946	1	0.50	118	59.00
1.001305662	1	0.50	119	59.50
1.0013653862	1	0.50	120	60.00
1.0015704118	1	0.50	121	60.50
1.0016763749	1	0.50	122	61.00
1.0016959438	1	0.50	123	61.50
1.0017630978	1	0.50	124	62.00
1.0017710018	1	0.50	125	62.50
1.0022498859	1	0.50	126	63.00
1.0022739702	1	0.50	127	63.50
1.0023151444	1	0.50	128	64.00
1.0023755489	1	0.50	129	64.50
1.0024411357	1	0.50	130	65.00
1.0025861599	1	0.50	131	65.50
1.0026580366	1	0.50	132	66.00
1.0029895864	1	0.50	133	66.50
1.0030485352	1	0.50	134	67.00
1.0030709461	1	0.50	135	67.50
1.0031144461	1	0.50	136	68.00
1.0032725994	1	0.50	137	68.50
1.0033565725	1	0.50	138	69.00
1.0033740973	1	0.50	139	69.50
1.003460651	1	0.50	140	70.00
1.0036444666	1	0.50	141	70.50
1.0037776642	1	0.50	142	71.00
1.0038753996	1	0.50	143	71.50
1.0038876258	1	0.50	144	72.00
1.0039044737	1	0.50	145	72.50
1.0039542364	1	0.50	146	73.00
1.0039602023	1	0.50	147	73.50
1.0040381915	1	0.50	148	74.00
1.0040866042	1	0.50	149	74.50
1.0042176129	1	0.50	150	75.00
1.0043531322	1	0.50	151	75.50
1.0045328316	1	0.50	152	76.00
1.0045605144	1	0.50	153	76.50
1.0052252605	1	0.50	154	77.00
1.0052449824	1	0.50	155	77.50
1.0052748099	1	0.50	156	78.00
1.0053351822	1	0.50	157	78.50
1.0054620682	1	0.50	158	79.00
1.0055922607	1	0.50	159	79.50
1.0057809328	1	0.50	160	80.00
1.0058817586	1	0.50	161	80.50
1.0060513048	1	0.50	162	81.00
1.0060744718	1	0.50	163	81.50

---------- n=500 r=3 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.0063069442	1	0.50	164	82.00
1.0063575213	1	0.50	165	82.50
1.0066548837	1	0.50	166	83.00
1.0069025409	1	0.50	167	83.50
1.0069760666	1	0.50	168	84.00
1.0072914095	1	0.50	169	84.50
1.007490473	1	0.50	170	85.00
1.0075375141	1	0.50	171	85.50
1.0076929612	1	0.50	172	86.00
1.0081621189	1	0.50	173	86.50
1.008209407	1	0.50	174	87.00
1.0084189351	1	0.50	175	87.50
1.008823633	1	0.50	176	88.00
1.0091328339	1	0.50	177	88.50
1.0093413375	1	0.50	178	89.00
1.0095120624	1	0.50	179	89.50
1.0095186718	1	0.50	180	90.00
1.0099283375	1	0.50	181	90.50
1.0105387311	1	0.50	182	91.00
1.0108208645	1	0.50	183	91.50
1.0112076966	1	0.50	184	92.00
1.0113086015	1	0.50	185	92.50
1.0114287078	1	0.50	186	93.00
1.0116009964	1	0.50	187	93.50
1.0121431596	1	0.50	188	94.00
1.0133736335	1	0.50	189	94.50
1.0137880578	1	0.50	190	95.00
1.0140198725	1	0.50	191	95.50
1.0142510249	1	0.50	192	96.00
1.0145207241	1	0.50	193	96.50
1.0146346521	1	0.50	194	97.00
1.0157199572	1	0.50	195	97.50
1.015995445	1	0.50	196	98.00
1.016275387	1	0.50	197	98.50
1.0166318903	1	0.50	198	99.00
1.0170539082	1	0.50	199	99.50
1.0285864265	1	0.50	200	100.00

---------- n=1000 r=1 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9864496336	1	0.50	1	0.50
0.9875452285	1	0.50	2	1.00
0.9889723886	1	0.50	3	1.50
0.9899105738	1	0.50	4	2.00
0.9899980531	1	0.50	5	2.50
0.9901227189	1	0.50	6	3.00
0.9907075045	1	0.50	7	3.50
0.9913715894	1	0.50	8	4.00
0.9915127214	1	0.50	9	4.50
0.9916027769	1	0.50	10	5.00
0.9920849612	1	0.50	11	5.50
0.9921879705	1	0.50	12	6.00
0.9924106303	1	0.50	13	6.50
0.9924859034	1	0.50	14	7.00
0.9927660912	1	0.50	15	7.50
0.9928174171	1	0.50	16	8.00
0.9930654591	1	0.50	17	8.50
0.9933175559	1	0.50	18	9.00
0.9937468923	1	0.50	19	9.50
0.9937908352	1	0.50	20	10.00
0.9938212882	1	0.50	21	10.50
0.993892662	1	0.50	22	11.00
0.994025513	1	0.50	23	11.50
0.9942123833	1	0.50	24	12.00
0.9942840976	1	0.50	25	12.50
0.9943595981	1	0.50	26	13.00
0.9943693687	1	0.50	27	13.50
0.9944408918	1	0.50	28	14.00
0.9945068388	1	0.50	29	14.50
0.9946028239	1	0.50	30	15.00
0.994639761	1	0.50	31	15.50
0.994705233	1	0.50	32	16.00
0.9947449806	1	0.50	33	16.50
0.9948874904	1	0.50	34	17.00
0.9949683769	1	0.50	35	17.50
0.9950213775	1	0.50	36	18.00
0.99502635	1	0.50	37	18.50
0.9950427042	1	0.50	38	19.00
0.9950933688	1	0.50	39	19.50
0.9951570866	1	0.50	40	20.00
0.9951874478	1	0.50	41	20.50
0.9952009567	1	0.50	42	21.00
0.995725853	1	0.50	43	21.50

---------- n=1000 r=1 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9958931611	1	0.50	44	22.00
0.9959263878	1	0.50	45	22.50
0.9963105362	1	0.50	46	23.00
0.9963229292	1	0.50	47	23.50
0.9963724929	1	0.50	48	24.00
0.996494167	1	0.50	49	24.50
0.9964972999	1	0.50	50	25.00
0.9965555088	1	0.50	51	25.50
0.9966144769	1	0.50	52	26.00
0.9966992482	1	0.50	53	26.50
0.9967276356	1	0.50	54	27.00
0.996730341	1	0.50	55	27.50
0.9969816075	1	0.50	56	28.00
0.9970120164	1	0.50	57	28.50
0.9971758459	1	0.50	58	29.00
0.9972873588	1	0.50	59	29.50
0.9973192512	1	0.50	60	30.00
0.9973757014	1	0.50	61	30.50
0.997387585	1	0.50	62	31.00
0.9974445677	1	0.50	63	31.50
0.9975259594	1	0.50	64	32.00
0.9975730548	1	0.50	65	32.50
0.9975892256	1	0.50	66	33.00
0.9975927962	1	0.50	67	33.50
0.9976112537	1	0.50	68	34.00
0.9978098429	1	0.50	69	34.50
0.9978449622	1	0.50	70	35.00
0.9979447558	1	0.50	71	35.50
0.9980070941	1	0.50	72	36.00
0.9980332655	1	0.50	73	36.50
0.9980380103	1	0.50	74	37.00
0.9980481384	1	0.50	75	37.50
0.9982937588	1	0.50	76	38.00
0.9983776558	1	0.50	77	38.50
0.9983829637	1	0.50	78	39.00
0.9985161069	1	0.50	79	39.50
0.9985357894	1	0.50	80	40.00
0.9986728491	1	0.50	81	40.50
0.9986738231	1	0.50	82	41.00
0.9988722344	1	0.50	83	41.50
0.9988870633	1	0.50	84	42.00
0.9989251987	1	0.50	85	42.50
0.9989943376	1	0.50	86	43.00

---------- n=1000 r=1 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.999021566	1	0.50	87	43.50
0.9990439234	1	0.50	88	44.00
0.9991396339	1	0.50	89	44.50
0.9992468539	1	0.50	90	45.00
0.9992745548	1	0.50	91	45.50
0.9993219717	1	0.50	92	46.00
0.9993389445	1	0.50	93	46.50
0.9993767784	1	0.50	94	47.00
0.9996212571	1	0.50	95	47.50
0.9996619892	1	0.50	96	48.00
0.9996869854	1	0.50	97	48.50
0.9997183137	1	0.50	98	49.00
0.9997248773	1	0.50	99	49.50
0.9997477665	1	0.50	100	50.00
0.9999823002	1	0.50	101	50.50
1.0000990775	1	0.50	102	51.00
1.0001865499	1	0.50	103	51.50
1.0003337477	1	0.50	104	52.00
1.0003954494	1	0.50	105	52.50
1.0003980088	1	0.50	106	53.00
1.0004711458	1	0.50	107	53.50
1.0005142807	1	0.50	108	54.00
1.0006198378	1	0.50	109	54.50
1.0006495783	1	0.50	110	55.00
1.0008054301	1	0.50	111	55.50
1.0009801875	1	0.50	112	56.00
1.0010531229	1	0.50	113	56.50
1.0010758136	1	0.50	114	57.00
1.0011789939	1	0.50	115	57.50
1.0012394284	1	0.50	116	58.00
1.0012666406	1	0.50	117	58.50
1.0012743533	1	0.50	118	59.00
1.0013620266	1	0.50	119	59.50
1.0014087127	1	0.50	120	60.00
1.0015470401	1	0.50	121	60.50
1.0016751848	1	0.50	122	61.00
1.0017697272	1	0.50	123	61.50
1.0017722745	1	0.50	124	62.00
1.0018831166	1	0.50	125	62.50
1.0019595507	1	0.50	126	63.00
1.00210785	1	0.50	127	63.50
1.0022333807	1	0.50	128	64.00
1.0022870722	1	0.50	129	64.50

---------- n=1000 r=1 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.0023747067	1	0.50	130	65.00
1.0023887024	1	0.50	131	65.50
1.002399859	1	0.50	132	66.00
1.0024214095	1	0.50	133	66.50
1.002502634	1	0.50	134	67.00
1.0025741081	1	0.50	135	67.50
1.0025870105	1	0.50	136	68.00
1.0026019927	1	0.50	137	68.50
1.0026262988	1	0.50	138	69.00
1.0026959268	1	0.50	139	69.50
1.00269733	1	0.50	140	70.00
1.0027017942	1	0.50	141	70.50
1.0027442723	1	0.50	142	71.00
1.0027952877	1	0.50	143	71.50
1.0029044599	1	0.50	144	72.00
1.0029412838	1	0.50	145	72.50
1.0029991901	1	0.50	146	73.00
1.0030090596	1	0.50	147	73.50
1.0030114704	1	0.50	148	74.00
1.0030574797	1	0.50	149	74.50
1.0030683682	1	0.50	150	75.00
1.003086668	1	0.50	151	75.50
1.0031360724	1	0.50	152	76.00
1.0032594238	1	0.50	153	76.50
1.0032801045	1	0.50	154	77.00
1.0033152037	1	0.50	155	77.50
1.003600757	1	0.50	156	78.00
1.0036075903	1	0.50	157	78.50
1.003706587	1	0.50	158	79.00
1.0038778156	1	0.50	159	79.50
1.0039242761	1	0.50	160	80.00
1.003997337	1	0.50	161	80.50
1.0040005329	1	0.50	162	81.00
1.0040755787	1	0.50	163	81.50
1.0041595539	1	0.50	164	82.00
1.0041620769	1	0.50	165	82.50
1.0041756821	1	0.50	166	83.00
1.0041866491	1	0.50	167	83.50
1.0043127894	1	0.50	168	84.00
1.0044829937	1	0.50	169	84.50
1.0046137669	1	0.50	170	85.00
1.0046599271	1	0.50	171	85.50
1.0049007617	1	0.50	172	86.00

---------- n=1000 r=1 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.0049126893	1	0.50	173	86.50
1.0058145953	1	0.50	174	87.00
1.0058560068	1	0.50	175	87.50
1.0063279435	1	0.50	176	88.00
1.0063344446	1	0.50	177	88.50
1.0063469773	1	0.50	178	89.00
1.006388202	1	0.50	179	89.50
1.0064298036	1	0.50	180	90.00
1.0065725907	1	0.50	181	90.50
1.0067611174	1	0.50	182	91.00
1.0068188575	1	0.50	183	91.50
1.0068386973	1	0.50	184	92.00
1.0071515411	1	0.50	185	92.50
1.007216533	1	0.50	186	93.00
1.0073933341	1	0.50	187	93.50
1.0075385257	1	0.50	188	94.00
1.0077494138	1	0.50	189	94.50
1.007899174	1	0.50	190	95.00
1.0081306333	1	0.50	191	95.50
1.0081652352	1	0.50	192	96.00
1.0084648762	1	0.50	193	96.50
1.0089769219	1	0.50	194	97.00
1.00957204	1	0.50	195	97.50
1.0097600418	1	0.50	196	98.00
1.0098678771	1	0.50	197	98.50
1.010445066	1	0.50	198	99.00
1.0117324263	1	0.50	199	99.50
1.0129983521	1	0.50	200	100.00

---------- n=1000 r=2 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9870407446	1	0.50	1	0.50
0.9871813951	1	0.50	2	1.00
0.987274144	1	0.50	3	1.50
0.9878245402	1	0.50	4	2.00
0.9884312397	1	0.50	5	2.50
0.9890539289	1	0.50	6	3.00
0.9904392726	1	0.50	7	3.50
0.9909466369	1	0.50	8	4.00
0.991266612	1	0.50	9	4.50
0.9917371286	1	0.50	10	5.00
0.9918766687	1	0.50	11	5.50
0.9919691574	1	0.50	12	6.00
0.9922045131	1	0.50	13	6.50
0.9923914709	1	0.50	14	7.00
0.9926087613	1	0.50	15	7.50

---------- n=500 r=3 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9926455924	1	0.50	16	8.00
0.9927804981	1	0.50	17	8.50
0.993039792	1	0.50	18	9.00
0.9930603933	1	0.50	19	9.50
0.993119514	1	0.50	20	10.00
0.993195013	1	0.50	21	10.50
0.9934455206	1	0.50	22	11.00
0.9935499689	1	0.50	23	11.50
0.9936057883	1	0.50	24	12.00
0.9937081299	1	0.50	25	12.50
0.9938749631	1	0.50	26	13.00
0.9939338159	1	0.50	27	13.50
0.9939650297	1	0.50	28	14.00
0.9939756333	1	0.50	29	14.50
0.9940711166	1	0.50	30	15.00
0.9942191567	1	0.50	31	15.50
0.9942848602	1	0.50	32	16.00
0.9943701601	1	0.50	33	16.50
0.9944499179	1	0.50	34	17.00
0.9945463396	1	0.50	35	17.50
0.9946477268	1	0.50	36	18.00
0.994951837	1	0.50	37	18.50
0.9949752336	1	0.50	38	19.00
0.9950268783	1	0.50	39	19.50
0.9952084322	1	0.50	40	20.00
0.9954661507	1	0.50	41	20.50
0.995562358	1	0.50	42	21.00
0.9956392686	1	0.50	43	21.50
0.9956661115	1	0.50	45	22.50
0.9957840183	1	0.50	46	23.00
0.9959368997	1	0.50	47	23.50
0.9959925124	1	0.50	48	24.00
0.9960385624	1	0.50	49	24.50
0.9960628231	1	0.50	50	25.00
0.9962537005	1	0.50	51	25.50
0.9962571393	1	0.50	52	26.00
0.9963726056	1	0.50	53	26.50
0.9964554514	1	0.50	54	27.00
0.9966751281	1	0.50	55	27.50
0.9968157934	1	0.50	56	28.00
0.9969103529	1	0.50	57	28.50
0.9969676883	1	0.50	58	29.00
0.997000926	1	0.50	59	29.50
0.9970829365	1	0.50	60	30.00
0.997131193	1	0.50	61	30.50
0.9971652366	1	0.50	62	31.00
0.997230117	1	0.50	63	31.50
0.9972866961	1	0.50	64	32.00
0.9973347543	1	0.50	65	32.50
0.9973921479	1	0.50	66	33.00
0.997409967	1	0.50	67	33.50

---------- n=1000 r=2 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.997409967	1	0.50	67	33.50
0.9974695281	1	0.50	68	34.00
0.9974911773	1	0.50	69	34.50
0.9978012514	1	0.50	70	35.00
0.9978460796	1	0.50	71	35.50
0.9978577749	1	0.50	72	36.00
0.9978602903	1	0.50	73	36.50
0.9979037593	1	0.50	74	37.00
0.9979523841	1	0.50	75	37.50
0.9979650636	1	0.50	76	38.00
0.9981364154	1	0.50	77	38.50
0.9981405472	1	0.50	78	39.00
0.9981571353	1	0.50	79	39.50
0.9981970507	1	0.50	80	40.00
0.9983011009	1	0.50	81	40.50
0.998387948	1	0.50	82	41.00
0.998500544	1	0.50	83	41.50
0.9985477943	1	0.50	84	42.00
0.9985512834	1	0.50	85	42.50
0.9986994665	1	0.50	86	43.00
0.9988794274	1	0.50	88	44.00
0.9988881052	1	0.50	89	44.50
0.9989130067	1	0.50	90	45.00
0.9989198871	1	0.50	91	45.50
0.9989324964	1	0.50	92	46.00
0.9989709227	1	0.50	93	46.50
0.998987574	1	0.50	94	47.00
0.9991168422	1	0.50	95	47.50
0.9991263734	1	0.50	96	48.00
0.9991467828	1	0.50	97	48.50
0.9991750277	1	0.50	98	49.00
0.9992474295	1	0.50	99	49.50
0.9993575919	1	0.50	100	50.00
0.9994539502	1	0.50	101	50.50
0.9996150531	1	0.50	102	51.00
0.9996995683	1	0.50	103	51.50
0.9997018488	1	0.50	104	52.00
0.9997701647	1	0.50	105	52.50
0.9998248053	1	0.50	106	53.00
0.9998841235	1	0.50	107	53.50
0.9999261175	1	0.50	108	54.00
1.0000239818	1	0.50	109	54.50
1.0000483242	1	0.50	110	55.00
1.0001800863	1	0.50	111	55.50
1.0002704027	1	0.50	112	56.00
1.0003878549	1	0.50	113	56.50
1.0004724313	1	0.50	114	57.00
1.000532218	1	0.50	115	57.50
1.000555344	1	0.50	116	58.00
1.0006782471	1	0.50	117	58.50
1.000734566	1	0.50	118	59.00

---------- n=1000 r=2 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.0008257718	1	0.50	119	59.50
1.00083235	1	0.50	120	60.00
1.0008556564	1	0.50	121	60.50
1.0009381557	1	0.50	122	61.00
1.0010298278	1	0.50	123	61.50
1.001034251	1	0.50	124	62.00
1.0011305279	1	0.50	125	62.50
1.0011307428	1	0.50	126	63.00
1.0012159655	1	0.50	127	63.50
1.0012808761	1	0.50	128	64.00
1.0013182538	1	0.50	129	64.50
1.0013656692	1	0.50	130	65.00
1.0013814815	1	0.50	131	65.50
1.0014203588	1	0.50	132	66.00
1.0014937016	1	0.50	133	66.50
1.0016172893	1	0.50	134	67.00
1.0016342624	1	0.50	135	67.50
1.001714462	1	0.50	136	68.00
1.0017618522	1	0.50	137	68.50
1.00178947	1	0.50	138	69.00
1.0018712328	1	0.50	139	69.50
1.0019220151	1	0.50	140	70.00
1.0019525653	1	0.50	141	70.50
1.0020487527	1	0.50	142	71.00
1.0022498461	1	0.50	143	71.50
1.0022599855	1	0.50	144	72.00
1.0024329235	1	0.50	145	72.50
1.0024823428	1	0.50	146	73.00
1.0025315627	1	0.50	147	73.50
1.0025642089	1	0.50	148	74.00
1.0026426927	1	0.50	149	74.50
1.0028046489	1	0.50	150	75.00
1.0028074071	1	0.50	151	75.50
1.0029156933	1	0.50	152	76.00
1.0029466993	1	0.50	153	76.50
1.0029945767	1	0.50	154	77.00
1.0030337161	1	0.50	155	77.50
1.0030908765	1	0.50	156	78.00
1.0031102807	1	0.50	157	78.50
1.003116319	1	0.50	158	79.00
1.0032065996	1	0.50	159	79.50
1.0033523017	1	0.50	160	80.00
1.0036332021	1	0.50	161	80.50
1.0036657004	1	0.50	162	81.00
1.0037013454	1	0.50	163	81.50
1.0038101482	1	0.50	164	82.00
1.0040146128	1	0.50	165	82.50
1.0040452088	1	0.50	166	83.00
1.0040759922	1	0.50	167	83.50
1.0044942655	1	0.50	168	84.00
1.0046011104	1	0.50	169	84.50

---------- n=1000 r=2 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.0049243833	1	0.50	170	85.00
1.0051649745	1	0.50	171	85.50
1.005281523	1	0.50	172	86.00
1.0053600859	1	0.50	173	86.50
1.0053642827	1	0.50	174	87.00
1.0055408353	1	0.50	175	87.50
1.0057714449	1	0.50	176	88.00
1.0058217032	1	0.50	177	88.50
1.0059554837	1	0.50	178	89.00
1.0060151762	1	0.50	179	89.50
1.0060805939	1	0.50	180	90.00
1.0062935028	1	0.50	181	90.50
1.0064299824	1	0.50	182	91.00
1.0064538609	1	0.50	183	91.50
1.0065507567	1	0.50	184	92.00
1.0069374643	1	0.50	185	92.50
1.0069864922	1	0.50	186	93.00
1.0069873141	1	0.50	187	93.50
1.0071938453	1	0.50	188	94.00
1.0079687249	1	0.50	189	94.50
1.0080043459	1	0.50	190	95.00
1.00826471	1	0.50	191	95.50
1.0088216472	1	0.50	192	96.00
1.0092938817	1	0.50	193	96.50
1.0095315409	1	0.50	194	97.00
1.0102114384	1	0.50	195	97.50
1.0106725269	1	0.50	196	98.00
1.011306304	1	0.50	197	98.50
1.0116148361	1	0.50	198	99.00
1.0123901485	1	0.50	199	99.50
1.0130368915	1	0.50	200	100.00

---------- n=1000 r=3 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9875134604	1	0.50	1	0.50
0.9875915566	1	0.50	2	1.00
0.9884636424	1	0.50	3	1.50
0.9888418798	1	0.50	4	2.00
0.9889708884	1	0.50	5	2.50
0.9904213041	1	0.50	6	3.00
0.9904538729	1	0.50	7	3.50
0.9904767344	1	0.50	8	4.00
0.9906902723	1	0.50	9	4.50
0.9909414884	1	0.50	10	5.00
0.9909698347	1	0.50	11	5.50
0.991081634	1	0.50	12	6.00
0.9914777745	1	0.50	13	6.50
0.9915405475	1	0.50	14	7.00

---------- n=1000 r=3 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9917178131	1	0.50	15	7.50
0.9919017596	1	0.50	16	8.00
0.9919666196	1	0.50	17	8.50
0.9922105219	1	0.50	18	9.00
0.9922438354	1	0.50	19	9.50
0.9923614239	1	0.50	20	10.00
0.9924870646	1	0.50	21	10.50
0.9925534267	1	0.50	22	11.00
0.9925721622	1	0.50	23	11.50
0.9925999951	1	0.50	24	12.00
0.992624491	1	0.50	25	12.50
0.9926410922	1	0.50	26	13.00
0.992796534	1	0.50	27	13.50
0.9928461001	1	0.50	28	14.00
0.9928837892	1	0.50	29	14.50
0.9929646932	1	0.50	30	15.00
0.9933282672	1	0.50	31	15.50
0.9937426506	1	0.50	32	16.00
0.9938544264	1	0.50	33	16.50
0.9939378531	1	0.50	34	17.00
0.9940849912	1	0.50	35	17.50
0.9941008445	1	0.50	36	18.00
0.9941050182	1	0.50	37	18.50
0.9941174366	1	0.50	38	19.00
0.9941402352	1	0.50	39	19.50
0.9946534469	1	0.50	40	20.00
0.994678649	1	0.50	41	20.50
0.9948353146	1	0.50	42	21.00
0.994836303	1	0.50	43	21.50
0.9948364192	1	0.50	44	22.00
0.9948617488	1	0.50	45	22.50
0.9948900598	1	0.50	46	23.00
0.9949026639	1	0.50	47	23.50
0.9949808706	1	0.50	48	24.00
0.9950295298	1	0.50	49	24.50
0.995254809	1	0.50	50	25.00
0.9952751645	1	0.50	51	25.50
0.9953989542	1	0.50	52	26.00
0.9957395099	1	0.50	53	26.50
0.9958137417	1	0.50	54	27.00
0.9959901983	1	0.50	55	27.50
0.9960026273	1	0.50	56	28.00
0.9960297586	1	0.50	57	28.50
0.9960651165	1	0.50	58	29.00
0.9960872106	1	0.50	59	29.50
0.9962170683	1	0.50	60	30.00
0.9967066975	1	0.50	61	30.50
0.9970296819	1	0.50	62	31.00
0.99707686	1	0.50	63	31.50
0.9970918236	1	0.50	64	32.00
0.9971375198	1	0.50	65	32.50
0.9971847709	1	0.50	66	33.00

---------- n=1000 r=3 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9972264202	1	0.50	67	33.50
0.9972433099	1	0.50	68	34.00
0.9974041814	1	0.50	69	34.50
0.9974734895	1	0.50	70	35.00
0.9975183023	1	0.50	71	35.50
0.9975520495	1	0.50	72	36.00
0.9975636023	1	0.50	73	36.50
0.9975865558	1	0.50	74	37.00
0.9977153355	1	0.50	75	37.50
0.9977250949	1	0.50	76	38.00
0.9978082936	1	0.50	77	38.50
0.9979163584	1	0.50	78	39.00
0.9979304843	1	0.50	79	39.50
0.998014852	1	0.50	80	40.00
0.9980306628	1	0.50	81	40.50
0.9980356127	1	0.50	82	41.00
0.9980366533	1	0.50	83	41.50
0.9981594463	1	0.50	84	42.00
0.9982139533	1	0.50	85	42.50
0.9982157157	1	0.50	86	43.00
0.9983066224	1	0.50	87	43.50
0.9984238146	1	0.50	88	44.00
0.9984254762	1	0.50	89	44.50
0.998674688	1	0.50	90	45.00
0.9987213468	1	0.50	91	45.50
0.9987439069	1	0.50	92	46.00
0.9987497568	1	0.50	93	46.50
0.9988081754	1	0.50	94	47.00
0.9988626711	1	0.50	95	47.50
0.9988872582	1	0.50	96	48.00
0.9989023869	1	0.50	97	48.50
0.9990965269	1	0.50	98	49.00
0.9991947317	1	0.50	99	49.50
0.9992685705	1	0.50	100	50.00
0.9992954173	1	0.50	101	50.50
0.9993098869	1	0.50	102	51.00
0.9993157335	1	0.50	103	51.50
0.9993383149	1	0.50	104	52.00
0.9993809088	1	0.50	105	52.50
0.999407369	1	0.50	106	53.00
0.9997096401	1	0.50	107	53.50
0.9997658131	1	0.50	108	54.00
0.9997864503	1	0.50	109	54.50
0.9998141278	1	0.50	110	55.00
0.9998638638	1	0.50	111	55.50
0.9999038102	1	0.50	112	56.00
0.9999167445	1	0.50	113	56.50
0.9999226111	1	0.50	114	57.00
0.999927527	1	0.50	115	57.50
0.9999312251	1	0.50	116	58.00
0.9999838359	1	0.50	117	58.50
1.0001109238	1	0.50	118	59.00

---------- n=1000 r=3 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.0002700163	1	0.50	119	59.50
1.0003729428	1	0.50	120	60.00
1.0006030256	1	0.50	121	60.50
1.0007168659	1	0.50	122	61.00
1.0007182602	1	0.50	123	61.50
1.0008394071	1	0.50	124	62.00
1.0008878321	1	0.50	125	62.50
1.0009244206	1	0.50	126	63.00
1.0010457354	1	0.50	127	63.50
1.0010527308	1	0.50	128	64.00
1.0010868061	1	0.50	129	64.50
1.0012408573	1	0.50	130	65.00
1.0012539349	1	0.50	131	65.50
1.0012649551	1	0.50	132	66.00
1.001265606	1	0.50	133	66.50
1.0014092088	1	0.50	134	67.00
1.0014174228	1	0.50	135	67.50
1.0016108086	1	0.50	136	68.00
1.001673715	1	0.50	137	68.50
1.001676357	1	0.50	138	69.00
1.0017461847	1	0.50	139	69.50
1.0019194234	1	0.50	140	70.00
1.0020248687	1	0.50	141	70.50
1.0022813427	1	0.50	142	71.00
1.0024639925	1	0.50	143	71.50
1.0024956318	1	0.50	144	72.00
1.0026019482	1	0.50	145	72.50
1.0026146028	1	0.50	146	73.00
1.0026525409	1	0.50	147	73.50
1.0029067454	1	0.50	148	74.00
1.0029210656	1	0.50	149	74.50
1.0034787797	1	0.50	150	75.00
1.0035527276	1	0.50	151	75.50
1.0035819034	1	0.50	152	76.00
1.0038367738	1	0.50	153	76.50
1.003841476	1	0.50	154	77.00
1.0038984429	1	0.50	155	77.50
1.0039746538	1	0.50	156	78.00
1.0041563771	1	0.50	157	78.50
1.0042940568	1	0.50	158	79.00
1.0043139002	1	0.50	159	79.50
1.0043980292	1	0.50	160	80.00
1.0044857233	1	0.50	161	80.50
1.0045561569	1	0.50	162	81.00
1.0046099034	1	0.50	163	81.50
1.0048106838	1	0.50	164	82.00
1.0049601015	1	0.50	165	82.50
1.0049679911	1	0.50	166	83.00
1.0049956197	1	0.50	167	83.50
1.005214178	1	0.50	168	84.00
1.0052283227	1	0.50	169	84.50
1.0053186405	1	0.50	170	85.00

---------- n=1000 r=3 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.0053202837	1	0.50	171	85.50
1.0053656776	1	0.50	172	86.00
1.0054080042	1	0.50	173	86.50
1.0054805839	1	0.50	174	87.00
1.0055825466	1	0.50	175	87.50
1.0056437871	1	0.50	176	88.00
1.0057399387	1	0.50	177	88.50
1.006056922	1	0.50	178	89.00
1.006082487	1	0.50	179	89.50
1.006224013	1	0.50	180	90.00
1.0063283913	1	0.50	181	90.50
1.0063910828	1	0.50	182	91.00
1.0064025384	1	0.50	183	91.50
1.0066954766	1	0.50	184	92.00
1.0068119485	1	0.50	185	92.50
1.0069130916	1	0.50	186	93.00
1.0069785327	1	0.50	187	93.50
1.0070162581	1	0.50	188	94.00
1.0073668693	1	0.50	189	94.50
1.0074458755	1	0.50	190	95.00
1.0078788006	1	0.50	191	95.50
1.0087409328	1	0.50	192	96.00
1.0092187659	1	0.50	193	96.50
1.0094066401	1	0.50	194	97.00
1.0097502013	1	0.50	195	97.50
1.0105566432	1	0.50	196	98.00
1.0115340062	1	0.50	197	98.50
1.012367958	1	0.50	198	99.00
1.0127687787	1	0.50	199	99.50
1.0158526579	1	0.50	200	100.00

---------- n=2000 r=1 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.990449779	1	0.50	1	0.50
0.9917876117	1	0.50	2	1.00
0.9925893918	1	0.50	3	1.50
0.9927967984	1	0.50	4	2.00
0.9929330038	1	0.50	5	2.50
0.9930183273	1	0.50	6	3.00
0.9931801943	1	0.50	7	3.50
0.9933706763	1	0.50	8	4.00
0.9934122076	1	0.50	9	4.50
0.9937974569	1	0.50	10	5.00
0.9940091367	1	0.50	11	5.50
0.9943722986	1	0.50	12	6.00
0.9946268997	1	0.50	13	6.50
0.9946950079	1	0.50	14	7.00

---------- n=2000 r=1 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9947063453	1	0.50	15	7.50
0.9947673509	1	0.50	16	8.00
0.9951479006	1	0.50	17	8.50
0.9952806287	1	0.50	18	9.00
0.9953281788	1	0.50	19	9.50
0.9953672288	1	0.50	20	10.00
0.9955778156	1	0.50	21	10.50
0.9955949146	1	0.50	22	11.00
0.9958119306	1	0.50	23	11.50
0.9958376896	1	0.50	24	12.00
0.9958673457	1	0.50	25	12.50
0.9959613401	1	0.50	26	13.00
0.9960369428	1	0.50	27	13.50
0.9960850775	1	0.50	28	14.00
0.9962469602	1	0.50	29	14.50
0.9964199028	1	0.50	30	15.00
0.9964425433	1	0.50	31	15.50
0.9964714174	1	0.50	32	16.00
0.9966156723	1	0.50	33	16.50
0.9966327797	1	0.50	34	17.00
0.9966557133	1	0.50	35	17.50
0.9967035552	1	0.50	36	18.00
0.9967086739	1	0.50	37	18.50
0.9967625534	1	0.50	38	19.00
0.9968811939	1	0.50	39	19.50
0.996887047	1	0.50	40	20.00
0.9969327969	1	0.50	41	20.50
0.9969338641	1	0.50	42	21.00
0.9972112991	1	0.50	43	21.50
0.9972505529	1	0.50	44	22.00
0.9973089366	1	0.50	45	22.50
0.9973734748	1	0.50	46	23.00
0.9973887215	1	0.50	47	23.50
0.9975326219	1	0.50	48	24.00
0.9975441745	1	0.50	49	24.50
0.9975706411	1	0.50	50	25.00
0.9975882159	1	0.50	51	25.50
0.9976252682	1	0.50	52	26.00
0.9976832594	1	0.50	53	26.50
0.9977211137	1	0.50	54	27.00
0.9977572968	1	0.50	55	27.50
0.9978047898	1	0.50	56	28.00
0.9978153066	1	0.50	57	28.50
0.9978854629	1	0.50	58	29.00
0.9979595847	1	0.50	59	29.50
0.9979795208	1	0.50	60	30.00
0.9980001895	1	0.50	61	30.50
0.9981533592	1	0.50	62	31.00
0.9982418909	1	0.50	63	31.50
0.9983104124	1	0.50	64	32.00
0.9983440865	1	0.50	65	32.50
0.998364474	1	0.50	66	33.00

---------- n=2000 r=1 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9984089289	1	0.50	67	33.50
0.9984158166	1	0.50	68	34.00
0.9984298381	1	0.50	69	34.50
0.9984763737	1	0.50	70	35.00
0.9985355683	1	0.50	71	35.50
0.9987718229	1	0.50	72	36.00
0.9988087338	1	0.50	73	36.50
0.9988282671	1	0.50	74	37.00
0.9988507633	1	0.50	75	37.50
0.9990646536	1	0.50	76	38.00
0.999078304	1	0.50	77	38.50
0.999124772	1	0.50	78	39.00
0.999200321	1	0.50	79	39.50
0.9993130164	1	0.50	80	40.00
0.9993633466	1	0.50	81	40.50
0.9995504976	1	0.50	82	41.00
0.9995580868	1	0.50	83	41.50
0.9995785756	1	0.50	84	42.00
0.9995964525	1	0.50	85	42.50
0.9996626211	1	0.50	86	43.00
0.9996800732	1	0.50	87	43.50
0.9997010159	1	0.50	88	44.00
0.9997144135	1	0.50	89	44.50
0.999720059	1	0.50	90	45.00
0.999736934	1	0.50	91	45.50
0.9997724496	1	0.50	92	46.00
0.9997970345	1	0.50	93	46.50
0.9998292366	1	0.50	94	47.00
0.9998302464	1	0.50	95	47.50
0.9999386925	1	0.50	96	48.00
1.0000704113	1	0.50	97	48.50
1.0001180565	1	0.50	98	49.00
1.0001372514	1	0.50	99	49.50
1.0001392881	1	0.50	100	50.00
1.0001537926	1	0.50	101	50.50
1.0002284206	1	0.50	102	51.00
1.0002619185	1	0.50	103	51.50
1.0003148401	1	0.50	104	52.00
1.0003730622	1	0.50	105	52.50
1.0003933062	1	0.50	106	53.00
1.0004107119	1	0.50	107	53.50
1.0004985357	1	0.50	108	54.00
1.000540695	1	0.50	109	54.50
1.0005449138	1	0.50	110	55.00
1.0006292	1	0.50	111	55.50
1.0006979274	1	0.50	112	56.00
1.0007535662	1	0.50	113	56.50
1.0008126253	1	0.50	114	57.00
1.0008162816	1	0.50	115	57.50
1.0008186371	1	0.50	116	58.00
1.0008584604	1	0.50	117	58.50
1.0010170975	1	0.50	118	59.00

---------- n=2000 r=1 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.0010553876	1	0.50	119	59.50
1.0011168162	1	0.50	120	60.00
1.0011320453	1	0.50	121	60.50
1.001135917	1	0.50	122	61.00
1.0011446381	1	0.50	123	61.50
1.0012014468	1	0.50	124	62.00
1.0012848123	1	0.50	125	62.50
1.0013649319	1	0.50	126	63.00
1.0014562526	1	0.50	127	63.50
1.0014857108	1	0.50	128	64.00
1.0015011624	1	0.50	129	64.50
1.0015235311	1	0.50	130	65.00
1.0015307784	1	0.50	131	65.50
1.0015868407	1	0.50	132	66.00
1.0017334409	1	0.50	133	66.50
1.001797197	1	0.50	134	67.00
1.0018886116	1	0.50	135	67.50
1.0019514667	1	0.50	136	68.00
1.0019768011	1	0.50	137	68.50
1.0020113772	1	0.50	138	69.00
1.0020231365	1	0.50	139	69.50
1.0021807639	1	0.50	140	70.00
1.0022567319	1	0.50	141	70.50
1.002271286	1	0.50	142	71.00
1.0023079455	1	0.50	143	71.50
1.0023087475	1	0.50	144	72.00
1.0023494042	1	0.50	145	72.50
1.002421816	1	0.50	146	73.00
1.0024329118	1	0.50	147	73.50
1.0024599126	1	0.50	148	74.00
1.0025849033	1	0.50	149	74.50
1.0025986625	1	0.50	150	75.00
1.002634959	1	0.50	151	75.50
1.0026671946	1	0.50	152	76.00
1.0027338077	1	0.50	153	76.50
1.0027650959	1	0.50	154	77.00
1.0031173591	1	0.50	155	77.50
1.0031359085	1	0.50	156	78.00
1.0031546656	1	0.50	157	78.50
1.0032467159	1	0.50	158	79.00
1.0035031351	1	0.50	159	79.50
1.0035481678	1	0.50	160	80.00
1.0036022842	1	0.50	161	80.50
1.0036146555	1	0.50	162	81.00
1.0036616508	1	0.50	163	81.50
1.0038401411	1	0.50	164	82.00
1.0039157078	1	0.50	165	82.50
1.0041098964	1	0.50	166	83.00
1.0041375105	1	0.50	167	83.50
1.004147377	1	0.50	168	84.00
1.0043178101	1	0.50	169	84.50
1.0043967389	1	0.50	170	85.00

n=2000 r=1 p=1

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.004440486	1	0.50	171	85.50
1.0044667659	1	0.50	172	86.00
1.004552786	1	0.50	173	86.50
1.0047400561	1	0.50	174	87.00
1.0048662243	1	0.50	175	87.50
1.0048734639	1	0.50	176	88.00
1.0049674959	1	0.50	177	88.50
1.0050493935	1	0.50	178	89.00
1.0050552036	1	0.50	179	89.50
1.0052210842	1	0.50	180	90.00
1.0053355855	1	0.50	181	90.50
1.0054376352	1	0.50	182	91.00
1.0055077159	1	0.50	183	91.50
1.0056118942	1	0.50	184	92.00
1.0056577646	1	0.50	185	92.50
1.0057826161	1	0.50	186	93.00
1.0058165681	1	0.50	187	93.50
1.0060728381	1	0.50	188	94.00
1.0061983585	1	0.50	189	94.50
1.0064497658	1	0.50	190	95.00
1.0065923263	1	0.50	191	95.50
1.0068571612	1	0.50	192	96.00
1.0069005405	1	0.50	193	96.50
1.0074503078	1	0.50	194	97.00
1.0080737635	1	0.50	195	97.50
1.008523602	1	0.50	196	98.00
1.0085758976	1	0.50	197	98.50
1.0086546972	1	0.50	198	99.00
1.009019923	1	0.50	199	99.50
1.0094694736	1	0.50	200	100.00

n=2000 r=2 p=1

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9898631501	1	0.50	1	0.50
0.991059546	1	0.50	2	1.00
0.9920320863	1	0.50	3	1.50
0.9921379633	1	0.50	4	2.00
0.9921517963	1	0.50	5	2.50
0.992642297	1	0.50	6	3.00
0.9933690133	1	0.50	7	3.50
0.9937145342	1	0.50	8	4.00
0.9938867653	1	0.50	9	4.50
0.9939230156	1	0.50	10	5.00
0.9943177809	1	0.50	11	5.50
0.9945752754	1	0.50	12	6.00
0.9946128236	1	0.50	13	6.50
0.9952757026	1	0.50	14	7.00
0.995367391	1	0.50	15	7.50
0.995416398	1	0.50	16	8.00

---------- n=2000 r=2 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9955964689	1	0.50	17	8.50
0.9957008976	1	0.50	19	9.50
0.9957861114	1	0.50	20	10.0
0.9958461679	1	0.50	21	10.50
0.9958993731	1	0.50	22	11.00
0.9959000828	1	0.50	23	11.50
0.9959759095	1	0.50	24	12.00
0.9962551229	1	0.50	25	12.50
0.9962920693	1	0.50	26	13.00
0.9963295589	1	0.50	27	13.50
0.9965230317	1	0.50	28	14.00
0.9966976962	1	0.50	29	14.50
0.9967383808	1	0.50	30	15.00
0.9967506069	1	0.50	31	15.50
0.9967578155	1	0.50	32	16.00
0.99680311	1	0.50	33	16.50
0.9968198959	1	0.50	34	17.00
0.9969580826	1	0.50	35	17.50
0.9969746198	1	0.50	36	18.00
0.9969995541	1	0.50	37	18.50
0.9970648844	1	0.50	38	19.00
0.9971459249	1	0.50	39	19.50
0.9971792443	1	0.50	40	20.00
0.9972133215	1	0.50	41	20.50
0.9972492955	1	0.50	42	21.00
0.9973049043	1	0.50	43	21.50
0.9973106679	1	0.50	44	22.00
0.9973506968	1	0.50	45	22.50
0.9973953723	1	0.50	46	23.00
0.9975116468	1	0.50	47	23.50
0.9975234553	1	0.50	48	24.00
0.997605355	1	0.50	49	24.50
0.9976908902	1	0.50	50	25.00
0.9976915749	1	0.50	51	25.50
0.9977157746	1	0.50	52	26.00
0.9977416594	1	0.50	53	26.50
0.9977944004	1	0.50	54	27.00
0.9978508875	1	0.50	55	27.50
0.9978909995	1	0.50	56	28.00
0.9979655167	1	0.50	57	28.50
0.9980057411	1	0.50	58	29.00
0.9980898286	1	0.50	59	29.50
0.9980908837	1	0.50	60	30.00
0.9981337769	1	0.50	61	30.50
0.9981375979	1	0.50	62	31.00
0.9981729148	1	0.50	63	31.50
0.9982478269	1	0.50	64	32.00
0.9982741285	1	0.50	65	32.50
0.9983986199	1	0.50	66	33.00
0.9985343017	1	0.50	67	33.50
0.9985532031	1	0.50	68	34.00
0.998618458	1	0.50	69	34.50

---------- n=2000 r=2 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.998651546	1	0.50	70	35.00
0.9986747326	1	0.50	71	35.50
0.9987173091	1	0.50	72	36.00
0.9987189068	1	0.50	73	36.50
0.9987551859	1	0.50	74	37.00
0.9987699711	1	0.50	75	37.50
0.9987970253	1	0.50	76	38.00
0.9988230939	1	0.50	77	38.50
0.9989619705	1	0.50	78	39.00
0.9990956193	1	0.50	79	39.50
0.9991084842	1	0.50	80	40.00
0.999130998	1	0.50	81	40.50
0.9991358715	1	0.50	82	41.00
0.9992863192	1	0.50	83	41.50
0.999299827	1	0.50	84	42.00
0.9993124895	1	0.50	85	42.50
0.9993682986	1	0.50	86	43.00
0.9993697303	1	0.50	87	43.50
0.999431257	1	0.50	88	44.00
0.9994494825	1	0.50	89	44.50
0.9994912016	1	0.50	90	45.00
0.9995354751	1	0.50	91	45.50
0.9996017502	1	0.50	92	46.00
0.999693077	1	0.50	93	46.50
0.9997035465	1	0.50	94	47.00
0.999732514	1	0.50	95	47.50
0.999804058	1	0.50	96	48.00
0.9998091757	1	0.50	97	48.50
0.9998494518	1	0.50	98	49.00
0.9998520825	1	0.50	99	49.50
0.9998618889	1	0.50	100	50.00
0.9999084238	1	0.50	101	50.50
0.9999143938	1	0.50	102	51.00
0.9999147495	1	0.50	103	51.50
0.9999491397	1	0.50	104	52.00
1.0000393377	1	0.50	105	52.50
1.0000432348	1	0.50	106	53.00
1.0000726022	1	0.50	107	53.50
1.000088736	1	0.50	108	54.00
1.0000969046	1	0.50	109	54.50
1.000102485	1	0.50	110	55.00
1.0001397145	1	0.50	111	55.50
1.0001489686	1	0.50	112	56.00
1.0002051336	1	0.50	113	56.50
1.0002122439	1	0.50	114	57.00
1.0002455059	1	0.50	115	57.50
1.0003115178	1	0.50	116	58.00
1.000317772	1	0.50	117	58.50
1.0003349022	1	0.50	118	59.00
1.0003395032	1	0.50	119	59.50
1.0003492159	1	0.50	120	60.00

---------- n=2000 r=2 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.0003744101	1	0.50	121	60.50
1.0004181493	1	0.50	122	61.00
1.0004423083	1	0.50	123	61.50
1.0004431405	1	0.50	124	62.00
1.0004496066	1	0.50	125	62.50
1.0006977135	1	0.50	126	63.00
1.0007922854	1	0.50	127	63.50
1.0008015467	1	0.50	128	64.00
1.0008389836	1	0.50	129	64.50
1.0008410173	1	0.50	130	65.00
1.0008675649	1	0.50	131	65.50
1.0010374654	1	0.50	132	66.00
1.001194691	1	0.50	133	66.50
1.0012155102	1	0.50	134	67.00
1.0013203789	1	0.50	135	67.50
1.0014289479	1	0.50	136	68.00
1.0014549724	1	0.50	137	68.50
1.0014642044	1	0.50	138	69.00
1.0014855708	1	0.50	139	69.50
1.0015305896	1	0.50	140	70.00
1.0015951787	1	0.50	141	70.50
1.001622548	1	0.50	142	71.00
1.0016359616	1	0.50	143	71.50
1.0016632608	1	0.50	144	72.00
1.0017929544	1	0.50	145	72.50
1.0018175185	1	0.50	146	73.00
1.0018635287	1	0.50	147	73.50
1.0019490305	1	0.50	148	74.00
1.0019495938	1	0.50	149	74.50
1.0020889688	1	0.50	150	75.00
1.002238407	1	0.50	151	75.50
1.0022387887	1	0.50	152	76.00
1.0022793275	1	0.50	153	76.50
1.0023798543	1	0.50	154	77.00
1.0024027425	1	0.50	155	77.50
1.0024136331	1	0.50	156	78.00
1.002449569	1	0.50	157	78.50
1.0024686177	1	0.50	158	79.00
1.0025351908	1	0.50	159	79.50
1.002662424	1	0.50	160	80.00
1.0026711986	1	0.50	161	80.50
1.0028027996	1	0.50	162	81.00
1.0028898061	1	0.50	163	81.50
1.0028943331	1	0.50	164	82.00
1.0029069704	1	0.50	165	82.50
1.0029631451	1	0.50	166	83.00
1.0031495669	1	0.50	167	83.50
1.0031762986	1	0.50	168	84.00
1.0035542298	1	0.50	169	84.50
1.0036717644	1	0.50	170	85.00
1.003680884	1	0.50	171	85.50

---------- n=2000 r=2 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.0037123878	1	0.50	172	86.00
1.0037375572	1	0.50	173	86.50
1.0038329568	1	0.50	174	87.00
1.0038935	1	0.50	175	87.50
1.0039214088	1	0.50	176	88.00
1.0040089981	1	0.50	177	88.50
1.0040480702	1	0.50	178	89.00
1.004117633	1	0.50	179	89.50
1.0043679159	1	0.50	180	90.00
1.0044228881	1	0.50	181	90.50
1.0046431515	1	0.50	182	91.00
1.0046663363	1	0.50	183	91.50
1.0046753057	1	0.50	184	92.00
1.0047205818	1	0.50	185	92.50
1.0048802769	1	0.50	186	93.00
1.0049259139	1	0.50	187	93.50
1.0051289323	1	0.50	188	94.00
1.005410934	1	0.50	189	94.50
1.005433082	1	0.50	190	95.00
1.005627882	1	0.50	191	95.50
1.005869275	1	0.50	192	96.00
1.0061778034	1	0.50	193	96.50
1.0063392055	1	0.50	194	97.00
1.0063860652	1	0.50	195	97.50
1.0077219612	1	0.50	196	98.00
1.0079545188	1	0.50	197	98.50
1.0080428481	1	0.50	198	99.00
1.0081310105	1	0.50	199	99.50
1.0088246386	1	0.50	200	100.00

---------- n=2000 r=3 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9880429271	1	0.50	1	0.50
0.9898123879	1	0.50	2	1.00
0.9899381171	1	0.50	3	1.50
0.9908661077	1	0.50	4	2.00
0.9912778788	1	0.50	5	2.50
0.9914386645	1	0.50	6	3.00
0.9918826204	1	0.50	7	3.50
0.9923215841	1	0.50	8	4.00
0.9923536841	1	0.50	9	4.50
0.9924243755	1	0.50	10	5.00
0.9926392636	1	0.50	11	5.50
0.9930606744	1	0.50	12	6.00
0.9934082874	1	0.50	13	6.50
0.9936235576	1	0.50	14	7.00
0.9939374237	1	0.50	15	7.50
0.9941375737	1	0.50	16	8.00
0.9941876941	1	0.50	17	8.50

---------- n=2000 r=3 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9942125127	1	0.50	18	9.00
0.9943329833	1	0.50	19	9.50
0.9943545717	1	0.50	20	10.00
0.9944542999	1	0.50	21	10.50
0.9944940274	1	0.50	22	11.00
0.9949268333	1	0.50	23	11.50
0.9950214823	1	0.50	24	12.00
0.9952492155	1	0.50	25	12.50
0.9952730307	1	0.50	26	13.00
0.9952835785	1	0.50	27	13.50
0.9952837793	1	0.50	28	14.00
0.9958268167	1	0.50	29	14.50
0.9959164393	1	0.50	30	15.00
0.996044781	1	0.50	31	15.50
0.9961330317	1	0.50	32	16.00
0.9961666601	1	0.50	33	16.50
0.9961839197	1	0.50	34	17.00
0.996228283	1	0.50	35	17.50
0.9962294259	1	0.50	36	18.00
0.9962598654	1	0.50	37	18.50
0.996428292	1	0.50	38	19.00
0.9964970991	1	0.50	39	19.50
0.9965462469	1	0.50	40	20.00
0.9966729188	1	0.50	41	20.50
0.9968887635	1	0.50	42	21.00
0.9969156712	1	0.50	43	21.50
0.9969322545	1	0.50	44	22.00
0.9969330668	1	0.50	45	22.50
0.9970814824	1	0.50	46	23.00
0.9971135526	1	0.50	47	23.50
0.9972483646	1	0.50	48	24.00
0.9972996703	1	0.50	49	24.50
0.9973227138	1	0.50	50	25.00
0.9973368044	1	0.50	51	25.50
0.9975324337	1	0.50	52	26.00
0.9976373575	1	0.50	53	26.50
0.997680307	1	0.50	54	27.00
0.9977018773	1	0.50	55	27.50
0.9977074128	1	0.50	56	28.00
0.9978208973	1	0.50	57	28.50
0.9978230962	1	0.50	58	29.00
0.9978255955	1	0.50	59	29.50
0.9980471311	1	0.50	60	30.00
0.9980976878	1	0.50	61	30.50
0.9981049316	1	0.50	62	31.00
0.9981767567	1	0.50	63	31.50
0.9982171262	1	0.50	64	32.00
0.9984851499	1	0.50	65	32.50
0.9986439017	1	0.50	66	33.00
0.9986672721	1	0.50	67	33.50
0.9986676457	1	0.50	68	34.00
0.9987135105	1	0.50	69	34.50

---------- n=2000 r=3 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9987869683	1	0.50	70	35.00
0.9988301148	1	0.50	71	35.50
0.9988892996	1	0.50	72	36.00
0.9988960296	1	0.50	73	36.50
0.9989091089	1	0.50	74	37.00
0.9989241798	1	0.50	75	37.50
0.9989829393	1	0.50	76	38.00
0.998999313	1	0.50	77	38.50
0.9990172624	1	0.50	78	39.00
0.9990340384	1	0.50	79	39.50
0.9990960731	1	0.50	80	40.00
0.9991334107	1	0.50	81	40.50
0.9991941648	1	0.50	82	41.00
0.999319938	1	0.50	83	41.50
0.9993328758	1	0.50	84	42.00
0.9994039066	1	0.50	85	42.50
0.9994227474	1	0.50	86	43.00
0.9994419327	1	0.50	87	43.50
0.9995066289	1	0.50	88	44.00
0.9995466165	1	0.50	89	44.50
0.9995563282	1	0.50	90	45.00
0.9996589754	1	0.50	91	45.50
0.9997635511	1	0.50	92	46.00
0.9997877369	1	0.50	93	46.50
0.9998050053	1	0.50	94	47.00
0.9998751715	1	0.50	95	47.50
0.9998954189	1	0.50	96	48.00
1.0000564829	1	0.50	97	48.50
1.0000756971	1	0.50	98	49.00
1.0003323676	1	0.50	99	49.50
1.0003597898	1	0.50	100	50.00
1.0003829209	1	0.50	101	50.50
1.0004054463	1	0.50	102	51.00
1.0005069254	1	0.50	103	51.50
1.0005610284	1	0.50	104	52.00
1.0006283924	1	0.50	105	52.50
1.0006847482	1	0.50	106	53.00
1.0007007318	1	0.50	107	53.50
1.0007107991	1	0.50	108	54.00
1.0008495771	1	0.50	109	54.50
1.0008557225	1	0.50	110	55.00
1.0008712919	1	0.50	111	55.50
1.0008756873	1	0.50	112	56.00
1.0009153758	1	0.50	113	56.50
1.0009228326	1	0.50	114	57.00
1.0009554474	1	0.50	115	57.50
1.0009698407	1	0.50	116	58.00
1.0009949917	1	0.50	117	58.50
1.0011150404	1	0.50	118	59.00
1.0011469249	1	0.50	119	59.50
1.0011905987	1	0.50	120	60.00

---------- n=2000 r=3 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.0012121314	1	0.50	121	60.50
1.0012385454	1	0.50	122	61.00
1.0012650304	1	0.50	123	61.50
1.0013313942	1	0.50	124	62.00
1.0013776249	1	0.50	125	62.50
1.0014297905	1	0.50	126	63.00
1.0014526526	1	0.50	127	63.50
1.001475346	1	0.50	128	64.00
1.0014981177	1	0.50	129	64.50
1.0015167245	1	0.50	131	65.50
1.0017754399	1	0.50	132	66.00
1.0017963818	1	0.50	133	66.50
1.0019332004	1	0.50	134	67.00
1.0019603063	1	0.50	135	67.50
1.0019935996	1	0.50	136	68.00
1.0020376504	1	0.50	137	68.50
1.0021008956	1	0.50	138	69.00
1.0021011628	1	0.50	139	69.50
1.0022112048	1	0.50	140	70.00
1.0022800152	1	0.50	141	70.50
1.0022915349	1	0.50	142	71.00
1.002293258	1	0.50	143	71.50
1.0023775823	1	0.50	144	72.00
1.0024553996	1	0.50	145	72.50
1.0025027802	1	0.50	146	73.00
1.0025549692	1	0.50	147	73.50
1.0025861917	1	0.50	148	74.00
1.0026089838	1	0.50	149	74.50
1.0026368689	1	0.50	150	75.00
1.002719522	1	0.50	151	75.50
1.002780757	1	0.50	152	76.00
1.0028535265	1	0.50	153	76.50
1.0028688334	1	0.50	154	77.00
1.0029099072	1	0.50	155	77.50
1.0029211068	1	0.50	156	78.00
1.0029544105	1	0.50	157	78.50
1.0031699852	1	0.50	158	79.00
1.0032424543	1	0.50	159	79.50
1.0033322911	1	0.50	160	80.00
1.0033692489	1	0.50	161	80.50
1.0033968341	1	0.50	162	81.00
1.003483227	1	0.50	163	81.50
1.0035251492	1	0.50	164	82.00
1.0037114971	1	0.50	165	82.50
1.0037208724	1	0.50	166	83.00
1.003748625	1	0.50	167	83.50
1.0037840877	1	0.50	168	84.00
1.0038868284	1	0.50	169	84.50
1.003989758	1	0.50	170	85.00
1.0040154035	1	0.50	171	85.50
1.004100416	1	0.50	172	86.00

---------- n=2000 r=3 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.0041452518	1	0.50	173	86.50
1.004250581	1	0.50	174	87.00
1.0042673972	1	0.50	175	87.50
1.0043267407	1	0.50	176	88.00
1.0043681891	1	0.50	177	88.50
1.0044771169	1	0.50	178	89.00
1.0046767908	1	0.50	179	89.50
1.0050986747	1	0.50	180	90.00
1.005129786	1	0.50	181	90.50
1.0051944664	1	0.50	182	91.00
1.0053167406	1	0.50	183	91.50
1.0054078308	1	0.50	184	92.00
1.0054286322	1	0.50	185	92.50
1.0055572257	1	0.50	186	93.00
1.0056733088	1	0.50	187	93.50
1.0061595126	1	0.50	188	94.00
1.0061611924	1	0.50	189	94.50
1.0061964026	1	0.50	190	95.00
1.0064119181	1	0.50	191	95.50
1.0064866033	1	0.50	192	96.00
1.0065188595	1	0.50	193	96.50
1.0066410078	1	0.50	194	97.00
1.0068386474	1	0.50	195	97.50
1.0078855268	1	0.50	196	98.00
1.0080101608	1	0.50	197	98.50
1.0085651433	1	0.50	198	99.00
1.0100059273	1	0.50	199	99.50
1.0126650029	1	0.50	200	100.00

Critical Values – Simulated Data

	150	500	1000	2000
r = 0.2	1.0213788742	1.0132318310	1.0078991740	1.0064497658
r = 0.4	1.0196160983	1.0122569345	1.0080043459	1.0054330820
r = 0.6	1.0238464383	1.0137880578	1.0074458755	1.0061964026

Appendix I. SAS code for Logistic Regression

```
data d1;
 input n1 deviance1 parameters1 df1 pc1
  n2 deviance2 parameters2 df2 pc2
  f n r p cell iter;
nsq=n*n;
r2=.2;
if r=2 then r2=.4;
if r=3 then r2=.6;
p2=2;
if p = 3 then p2=4;
if p = 4 then p2=6;
reject=0;
 if n=150 then do;
 if r=1 then do;
  ucv=1.0213788742;
 end;
 if r=2 then do;
  ucv=1.0196160983;
 end;
 if r=3 then do;
  ucv=1.0238464383;
 end;
end;
if n=500 then do;
 if r=1 then do;
  ucv=1.013231831;
 end;
 if r=2 then do;
  ucv=1.0122569345;
 end;
 if r=3 then do;
  ucv=1.0140198725;
 end;
end;
if n=1000 then do;
 if r=1 then do;
  ucv=1.007899174;
 end;
 if r=2 then do;
  ucv=1.0080043459;
 end;
 if r=3 then do;
  ucv=1.0074458755;
 end;
end;
if n=2000 then do;
 if r=1 then do;
  ucv=1.0064497658;
 end;
 if r=2 then do;
  ucv=1.005433082;
 end;
```

```
 if r=3 then do;
  ucv=1.0061964026;
 end;
end;
reject=0;
if f gt ucv then reject=1;
cards;
150 6797.452 30 120 56.645433333 150 6929.539 30 120 57.746158333
1.01943184 150 2 2 122 1
.
.
.
2000 27966.475 31 1969 14.203390046 2000 28881.204 31 1969 14.667955307
1.0327080549 2000 3 4 434 50
;
run;

proc means noprint nway data=d1;
 class n;
 var reject n;
 output out=ndat sum(reject)=reject;
run;

data ndat;
 set ndat;
 n_logit = log((reject + 1) / (_freq_ - reject + 1));
run;

proc plot;
 plot n_logit*n / vaxis = -3 to 3 by 1;
run;

proc means noprint nway data=d1;
 class p2;
 var reject p2;
 output out=pdat sum(reject)=reject;
run;

data pdat;
 set pdat;
 p_logit = log((reject + 1) / (_freq_ - reject + 1));
run;

proc plot;
 plot p_logit*p2 / vaxis = -3 to 3 by 1;
run;

proc means noprint nway data=d1;
 class r2;
 var reject r2;
 output out=rdat sum(reject)=reject;
run;

data rdat;
 set rdat;
 r_logit = log((reject + 1) / (_freq_ - reject + 1));
run;
```

```
proc plot;
 plot r_logit*r2 / vaxis = -3 to 3 by 1;
run;

proc logistic data=d1;
 title 'THREE-WAY INTERACTION MODEL WITH N QUADRATIC';
 model reject=p2 r2 n nsq p2*r2 p2*n p2*nsq r2*n r2*nsq r2*p2*n
r2*p2*nsq;
run;

proc logistic data=d1;
 title 'THREE-WAY INTERACTION MODEL WITHOUT N QUADRATIC';
 model reject=p2 r2 n nsq p2*r2 p2*n p2*nsq r2*n r2*nsq r2*p2*n;
run;

proc logistic data=d1;
 title 'ALL TWO-WAY INTERACTIONS MODEL WITH N QUADRATIC';
 model reject=p2 r2 n nsq p2*r2 p2*n p2*nsq r2*n r2*nsq;
run;

proc logistic data=d1;
 title 'TWO-WAY INTERACTIONS MODEL WITH N QUADRATIC (REMOVING R*N^2
TERM)';
 model reject=p2 r2 n nsq p2*r2 p2*n p2*nsq r2*n;
run;

proc logistic data=d1;
 title 'TWO-WAY INTERACTIONS MODEL WITH N QUADRATIC (REMOVING R*N^2 &
P*R TERMS)';
 model reject=p2 r2 n nsq p2*n p2*nsq r2*n;
run;

proc logistic data=d1;
 title 'MAIN EFFECT MODEL WITH N QUADRATIC';
 model reject=p2 r2 n nsq;
run;
```

Appendix J. Results of Logistic Regression

THREE-WAY INTERACTION MODEL WITH N QUADRATIC

The LOGISTIC Procedure
Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.6785	1.0795	2.4178	0.1200
p2	1	-0.0761	0.2528	0.0906	0.7634
r2	1	-1.3530	2.6214	0.2664	0.6058
n	1	0.000654	0.00258	0.0643	0.7998
nsq	1	-2.18E-6	1.25E-6	3.0310	0.0817
p2*r2	1	0.5978	0.6273	0.9081	0.3406
p2*n	1	-0.00055	0.000601	0.8277	0.3629
p2*nsq	1	3.343E-7	2.884E-7	1.3437	0.2464
r2*n	1	0.00333	0.00622	0.2858	0.5929
r2*nsq	1	5.956E-7	2.862E-6	0.0433	0.8352
p2*r2*n	1	-0.00001	0.00146	0.0001	0.9942
p2*r2*nsq	1	-1.68E-7	6.65E-7	0.0637	0.8008

THREE-WAY INTERACTION MODEL WITHOUT N QUADRATIC

The LOGISTIC Procedure
Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.8583	0.8125	5.2305	0.0222
p2	1	-0.1217	0.1765	0.4755	0.4904
r2	1	-1.8160	1.8717	0.9413	0.3319
n	1	0.000102	0.00137	0.0056	0.9405
nsq	1	-1.9E-6	6.207E-7	9.3929	0.0022
p2*r2	1	0.7168	0.4138	3.0011	0.0832
p2*n	1	-0.00041	0.000245	2.7853	0.0951
p2*nsq	1	2.663E-7	1.017E-7	6.8511	0.0089
r2*n	1	0.00473	0.00282	2.8051	0.0940
r2*nsq	1	-7.3E-8	1.081E-6	0.0046	0.9462
p2*r2*n	1	-0.00036	0.000415	0.7728	0.3793

ALL TWO-WAY INTERACTIONS MODEL WITH N QUADRATIC

The LOGISTIC Procedure
Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.4586	0.6675	4.7751	0.0289
p2	1	-0.0190	0.1324	0.0207	0.8856
r2	1	-0.6952	1.3766	0.2550	0.6136
n	1	0.000486	0.00129	0.1420	0.7063
nsq	1	-1.77E-6	5.935E-7	8.9431	0.0028
p2*r2	1	0.4234	0.2432	3.0308	0.0817
p2*n	1	-0.00051	0.000215	5.6137	0.0178
p2*nsq	1	2.372E-7	9.623E-8	6.0739	0.0137
r2*n	1	0.00337	0.00238	2.0071	0.1566
r2*nsq	1	-1.18E-7	1.081E-6	0.0119	0.9132

TWO-WAY INTERACTIONS MODEL WITH N QUADRATIC (REMOVING R*N^**2** TERM)

The LOGISTIC Procedure
Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.4263	0.5977	5.6951	0.0170
p2	1	-0.0190	0.1322	0.0207	0.8856
r2	1	-0.6150	1.1649	0.2787	0.5976
n	1	0.000586	0.000914	0.4101	0.5219
nsq	1	-1.82E-6	3.996E-7	20.8042	<.0001
p2*r2	1	0.4243	0.2431	3.0476	0.0809
p2*n	1	-0.00051	0.000215	5.6315	0.0176
p2*nsq	1	2.372E-7	9.618E-8	6.0805	0.0137
r2*n	1	0.00312	0.000674	21.4617	<.0001

TWO-WAY INTERACTIONS MODEL WITH N QUADRATIC (REMOVING R*N^2 & P*R TERMS)

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.7719	0.4557	2.8691	0.0903
p2	1	0.1448	0.0932	2.4148	0.1202
r2	1	1.0552	0.6682	2.4937	0.1143
n	1	0.000728	0.000905	0.6478	0.4209
nsq	1	-1.95E-6	3.967E-7	24.0891	<.0001
p2*n	1	-0.00055	0.000214	6.5401	0.0105
p2*nsq	1	2.669E-7	9.428E-8	8.0143	0.0046
r2*n	1	0.00316	0.000670	22.2057	<.0001

MAIN EFFECT MODEL WITH N QUADRATIC

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.4684	0.2413	3.7680	0.0522
p2	1	-0.0189	0.0345	0.2991	0.5844
r2	1	3.7495	0.3883	93.2531	<.0001
n	1	-0.00047	0.000338	1.9545	0.1621
nsq	1	-7.04E-7	1.483E-7	22.5319	<.0001

Appendix K. Statistical Power* for W-index Procedure by Number of Items Lacking Equivalence

n	150			500			1000			2000		
r	.2	.4	.6	.2	.4	.6	.2	.4	.6	.2	.4	.6
p												
2	.12	.10	.06	.16	.16	.12	.52	.26	.22	1.00	.92	.60
4	.24	.16	.06	.38	.28	.08	.62	.38	.26	.90	.92	.60
6	.28	.18	.10	.22	.20	.12	.68	.30	.22	1.00	.90	.64

* Power is the percentage of time an accurate identification of lack of equivalence is made.

Appendix L. Statistical Power* of W-index Procedure by Sample Size

p	2			4			6		
r	0.2	0.4	0.6	0.2	0.4	0.6	0.2	0.4	0.6
n									
150	0.12	0.10	0.16	0.24	0.16	0.06	0.28	0.18	0.10
500	0.16	0.16	0.12	0.38	0.28	0.08	0.22	0.20	0.12
1000	0.52	0.26	0.22	0.62	0.38	0.26	0.68	0.30	0.22
2000	1.00	0.92	0.60	0.90	0.92	0.60	1.00	0.90	0.64

* Power is the percentage of time an accurate identification of lack of equivalence is made.

Appendix M. Statistical Power* of W-index Procedure by Intertrait Correlation

Number of items lack equivalence	Sample Size	Intertrait Correlation	Power
2 items (~ 8%) lack equivalence (*p2*)	n = 150	r = 0.2	0.12
		r = 0.4	0.10
		r = 0.6	0.06
	n = 500	r = 0.2	0.16
		r = 0.4	0.16
		r = 0.6	0.12
	n = 1000	r = 0.2	0.52
		r = 0.4	0.26
		r = 0.6	0.22
	n = 2000	r = 0.2	1.00
		r = 0.4	0.92
		r = 0.6	0.60
4 items (~15%) lack equivalence (*p3*)	n = 150	r = 0.2	0.24
		r = 0.4	0.16
		r = 0.6	0.06
	n = 500	r = 0.2	0.38
		r = 0.4	0.28
		r = 0.6	0.08
	n = 1000	r = 0.2	0.62
		r = 0.4	0.38
		r = 0.6	0.26
	n = 2000	r = 0.2	0.90
		r = 0.4	0.92
		r = 0.6	0.60
6 items (~ 23%) lack equivalence (*p4*)	n = 150	r = 0.2	0.28
		r = 0.4	0.18
		r = 0.6	0.10
	n = 500	r = 0.2	0.22
		r = 0.4	0.20
		r = 0.6	0.12
	n = 1000	r = 0.2	0.68
		r = 0.4	0.30
		r = 0.6	0.22
	n = 2000	r = 0.2	1.00
		r = 0.4	0.90
		r = 0.6	0.64

* Power is the proportion of cases for which an accurate identification of lack of equivalence is made.

Appendix N. Factor Loadings for Real Data Survey Instrument

Rotation Method: Promax (power = 3)

	Factor1	Factor2	Factor3	Factor4
item1	13	78 *	26	23
item2	20	71 *	14	-1
item3	33	45	52*	23
item4	35	80 *	28	-5
item5	33	71 *	30	-6
item6	35	42	43 *	4
item7	72 *	27	47	7
item8	79 *	18	40	14
item9	66 *	22	27	30
item10	78 *	25	33	10
item11	71 *	25	17	15
item12	77 *	17	23	11
item13	72 *	33	15	9
item14	8	3	17	42 *
item15	17	11	20	87 *
item16	15	6	20	88 *
item17	22	6	44	69 *
item18	16	7	29	30 *
item19	7	9	15	54 *
item20	16	61 *	43	18
item21	34	22	72 *	15
item22	18	38	61 *	19
item23	26	21	74 *	13
item24	25	56 *	41	0
item25	16	19	67 *	25
item26	24	41	73 *	23

Printed values are multiplied by 100 and rounded to the nearest integer.
Largest values are flagged by an '*'.

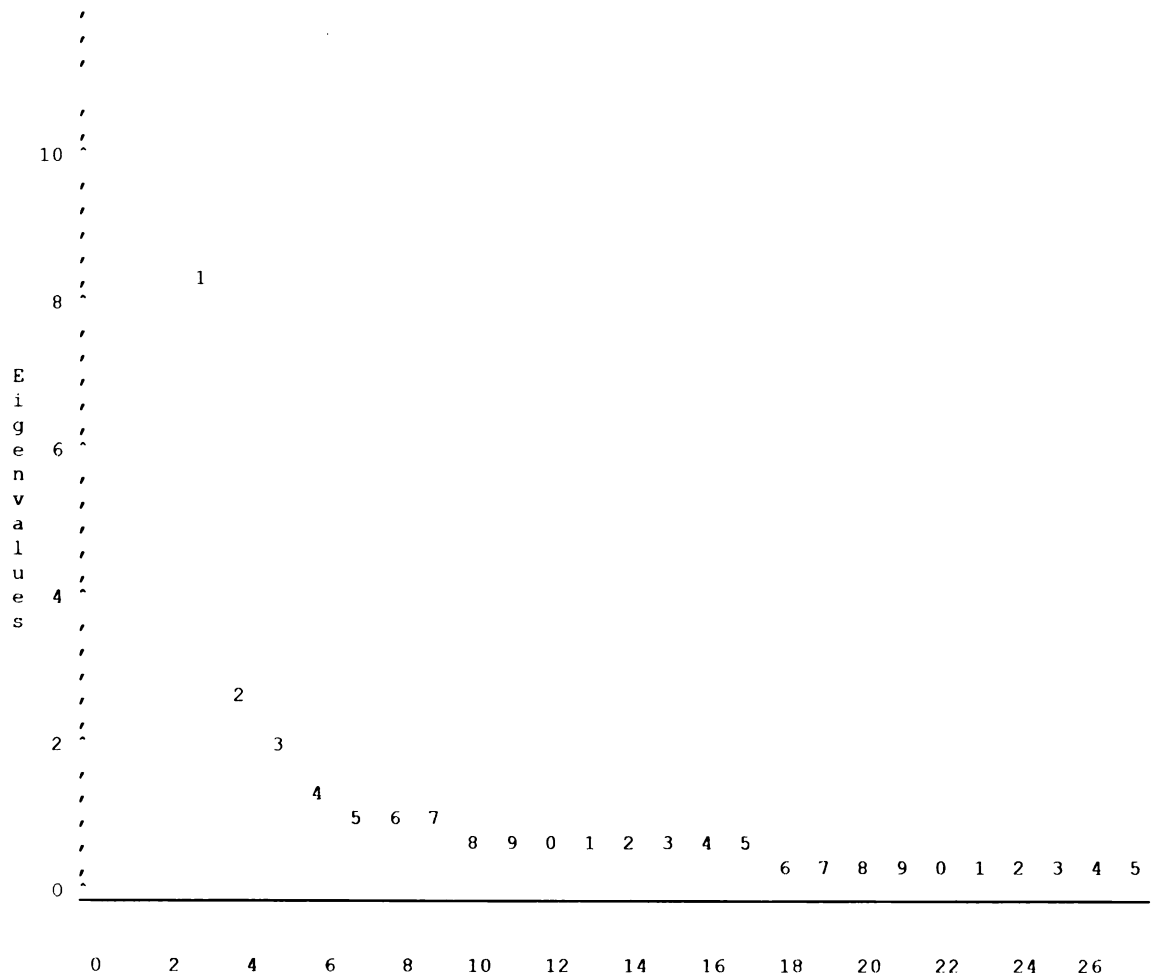
Appendix O. Eigenvalues and Scree Plot for Real Data

Eigenvalues of the Correlation Matrix: Total = 26 Average = 1

	Eigenvalue	Difference	Proportion	Cumulative
1	8.21880021	5.67100975	0.3161	0.3161
2	2.54779046	0.59121528	0.0980	0.4141
3	1.95657518	0.54845975	0.0753	0.4894
4	1.40811543	0.37273563	0.0542	0.5435

4 factors will be retained by the MINEIGEN criterion.

Scree Plot of Eigenvalues



Appendix P. Factor Correlations –Elementary and Secondary Real Data

Elementary Real Data

Inter-Factor Correlations

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
Factor1	100	41	41	19	41	30
Factor2	41	100	25	33	34	37
Factor3	41	25	100	13	30	19
Factor4	19	33	13	100	14	24
Factor5	41	34	30	14	100	27
Factor6	30	37	19	24	27	100

Printed values are multiplied by 100 and rounded to the nearest integer.

Secondary Real Data

Inter-Factor Correlations

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
Factor1	100	34	28	14	12	-6
Factor2	34	100	35	34	12	2
Factor3	28	35	100	12	16	0
Factor4	14	34	12	100	7	14
Factor5	12	12	16	7	100	-4
Factor6	-6	2	0	14	-4	100

Printed values are multiplied by 100 and rounded to the nearest integer.

Appendix Q. Frequency Distribution of *W*-index - Real Data

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.9818271148	1	1.00	1	1.00
0.9929075426	1	1.00	2	2.00
0.994217708	1	1.00	3	3.00
0.9948039607	1	1.00	4	4.00
0.9950878535	1	1.00	5	5.00
0.9969373848	1	1.00	6	6.00
0.9981047364	1	1.00	7	7.00
0.9981400147	1	1.00	8	8.00
0.9988376908	1	1.00	9	9.00
0.9990612701	1	1.00	10	10.00
0.9994145185	1	1.00	11	11.00
1.000490505	1	1.00	12	12.00
1.0007437779	1	1.00	13	13.00
1.0008392412	1	1.00	14	14.00
1.0011448951	1	1.00	15	15.00
1.0014217974	1	1.00	16	16.00
1.0020693426	1	1.00	17	17.00
1.0029324831	1	1.00	18	18.00
1.003041166	1	1.00	19	19.00
1.0031533179	1	1.00	20	20.00
1.0041749996	1	1.00	21	21.00
1.004185419	1	1.00	22	22.00
1.0044031536	1	1.00	23	23.00
1.004695244	1	1.00	24	24.00
1.0052413579	1	1.00	25	25.00
1.0055320335	1	1.00	26	26.00
1.0057100448	1	1.00	27	27.00
1.0062682445	1	1.00	28	28.00
1.0064229166	1	1.00	29	29.00
1.0071436426	1	1.00	30	30.00
1.0074582119	1	1.00	31	31.00
1.0082378441	1	1.00	32	32.00
1.008395885	1	1.00	33	33.00
1.0088362419	1	1.00	34	34.00
1.0095599909	1	1.00	35	35.00
1.010417881	1	1.00	36	36.00
1.0104419935	1	1.00	37	37.00
1.0105333278	1	1.00	38	38.00
1.0105441397	1	1.00	39	39.00
1.0106388743	1	1.00	40	40.00
1.0117580315	1	1.00	41	41.00
1.0120614534	1	1.00	42	42.00
1.0127438694	1	1.00	43	43.00
1.012980207	1	1.00	44	44.00
1.0134237621	1	1.00	45	45.00
1.0135164381	1	1.00	46	46.00
1.0140935602	1	1.00	47	47.00
1.0148141816	1	1.00	48	48.00
1.0149518434	1	1.00	49	49.00
1.0150917903	1	1.00	50	50.00
1.0164749514	1	1.00	51	51.00
1.0165280822	1	1.00	52	52.00

---------- n=2000 r=3 p=1 ----------

f	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1.0167966223	1	1.00	53	53.00
1.0169165685	1	1.00	54	54.00
1.01802137	1	1.00	55	55.00
1.0195277112	1	1.00	56	56.00
1.019594352	1	1.00	57	57.00
1.0197586877	1	1.00	58	58.00
1.01995168	1	1.00	59	59.00
1.0207221649	1	1.00	60	60.00
1.0208308284	1	1.00	61	61.00
1.0215730001	1	1.00	62	62.00
1.0222232355	1	1.00	63	63.00
1.0223687533	1	1.00	64	64.00
1.0226608209	1	1.00	65	65.00
1.0233759987	1	1.00	66	66.00
1.0253016317	1	1.00	67	67.00
1.0255317787	1	1.00	68	68.00
1.0255338267	1	1.00	69	69.00
1.0256768392	1	1.00	70	70.00
1.0265941491	1	1.00	71	71.00
1.0266461055	1	1.00	72	72.00
1.0273727976	1	1.00	73	73.00
1.0274952424	1	1.00	74	74.00
1.027672206	1	1.00	75	75.00
1.0288887232	1	1.00	76	76.00
1.0292020976	1	1.00	77	77.00
1.0296964278	1	1.00	78	78.00
1.0300430773	1	1.00	79	79.00
1.030177346	1	1.00	80	80.00
1.0302856028	1	1.00	81	81.00
1.0303625659	1	1.00	82	82.00
1.0313365074	1	1.00	83	83.00
1.0319094399	1	1.00	84	84.00
1.0328401519	1	1.00	85	85.00
1.0336410911	1	1.00	86	86.00
1.0336818269	1	1.00	87	87.00
1.0340006224	1	1.00	88	88.00
1.0343561362	1	1.00	89	89.00
1.0344726172	1	1.00	90	90.00
1.0354013824	1	1.00	91	91.00
1.0361300233	1	1.00	92	92.00
1.0370143745	1	1.00	93	93.00
1.0375719942	1	1.00	94	94.00
1.0403790147	1	1.00	95	95.00
1.0409811909	1	1.00	96	96.00
1.0434272216	1	1.00	97	97.00
1.0440633346	1	1.00	98	98.00
1.0451529389	1	1.00	99	99.00
1.0506942985	1	1.00	100	100.00

Appendix R. Exploratory Factor Analysis – Elementary Real Data

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
item1	7	-10	5	-12	77*	18
item2	4	-1	-1	7	87*	-5
item3	16	26	0	13	54*	1
item4	67*	24	-6	-9	25	-1
item5	81*	6	-9	-4	12	4
item6	-28	35	0	-18	11	52*
item7	-21	81*	12	-5	9	2
item8	19	79*	4	-2	-12	-9
item9	24	40*	6	27	9	-9
item10	2	75*	2	7	5	-10
item11	6	72*	0	1	-4	14
item12	19	80*	-16	10	-19	4
item13	1	60*	9	-6	14	-2
item14	11	-1	-17	16	12	65*
item15	-15	8	-2	79*	-4	15
item16	-5	-4	8	87*	6	-1
item17	-2	6	19	58*	-2	19
item18	14	13	37	-3	-13	36
item19	12	-17	1	22	-2	73*
item20	0	19	42*	-5	9	29
item21	4	-7	62*	-2	-4	36
item22	-7	4	82*	0	14	-27
item23	-9	-7	73*	8	15	10
item24	72*	-1	33	-13	-10	11
item25	1	9	68*	11	-15	-7
item26	20	-2	75*	4	-2	-5

Note: Printed values are multiplied by 100 and rounded to the nearest integer. Values greater than 0.4 are flagged by an '*'.

Appendix S. Exploratory Factor Analysis – Secondary Real Data

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
item1	82*	-14	-3	-23	-9	-4
item2	70*	3	-11	-5	-7	17
item3	30	10	38	4	-5	-1
item4	70*	12	2	-13	21	-7
item5	46*	13	10	-12	53*	-12
item6	14	16	32	-6	35	1
item7	4	67*	30	-13	-26	12
item8	-9	75*	19	-3	-1	2
item9	-6	61*	-5	27	23	-12
item10	-2	76*	8	-1	-3	-21
item11	3	73*	-15	11	7	4
item12	-8	78*	-3	1	22	10
item13	21	72*	-17	1	-6	6
item14	3	2	17	23	6	66*
item15	3	6	-9	88*	14	-4
item16	-1	5	-6	88*	4	7
item17	3	9	30	56*	-37	10
item18	-16	0	20	30	60*	7
item19	4	-7	-1	52*	39	36
item20	44*	-8	26	15	-2	-3
item21	-9	12	72*	-5	-4	-12
item22	10	-11	61*	2	29	26
item23	-10	-1	79*	-11	28	5
item24	36	1	24	-3	18	-41*
item25	-10	-9	66*	12	5	-42*
item26	20	-2	70*	0	4	-8

Note: Printed values are multiplied by 100 and rounded to the nearest integer. Values greater than 0.4 are flagged by an '*'.

REFERENCES

References

Ackerman, T.A. (1992). A didactic explanation of item bias item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement, 29*, 67-91.

Adams, R. J., & Wilson, M. (1996). Formulating the Rasch model as a mixed coefficients multinomial logit model (pp. 143-166). In *Objective Measurement: Theory into Practice Volume 3.* G. Engelhard and M. Wilson (Eds.) Norwood, NJ: Ablex Publishing.

Adams, R.J., Wilson, M., & Wang, W.C. (1997, March). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21*(1), 1-23.

Amemiya, T. (1985). *Advanced econometrics.* Cambridge, Mass.: Harvard University Press.

Bejar, I. (1980). A procedure for investigating the unidimensionality of achievement tests based on item parameter estimates. *Journal of Educational Measurement* 17(4), 283-96.

Bock, R.D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46,* 443-459.

Bock, R.D., Gibbons, R., & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement, 12*, 261-280.

Boles, J.S., Dean, D.H., Ricks, J.M., Short, J. C., & Want, G. (2000). The dimensionality of the Maslach Burnout Inventory across small business owners and educators. *Journal of Vocational Behavior, 56*, 12-34.

Bollen, K.A., & Long, J.S. (1993). *Testing structural equation models.* Newbury Park, CA: Sage.

Bress, P. (2000, December). Gender differences in teaching styles. *English Teaching Forum, 38* (4), 17-27.

Briggs, D.C., & Wilson, M. (2003). Understanding Rasch measurement: An introduction to multidimensional measurement using Rasch models. *Journal of Applied Measurement, 4*(1), 87-100.

Bryk, A.S. & Driscoll, M.E. (1988). *The high school as community: Contextual influences and consequences for students and teachers.* Madison: National Center on Effective Secondary Schools, University of Wisconsin.

Buss, A. R. & Royce, J.R. (1975). Detecting cross-cultural commonalties and differences: Intergroup factor analysis. *Psychological Bulletin*, 82(1),128-136.

Byne, B.M. (1994). Testing for the factorial validity, replication, and invariance of a measurement instrument: A paradigmatic application based on the Maslach Burnout Inventory. *Multivariate Behavioral Research, 29*, 289-311.

Byrne, B.M., & Campbell, T. L. (1999, September). Cross-cultural comparisons and the presumption of equivalent measurement and theoretical structure: A look beneath the surface. *Journal of Cross-Cultural Psychology, 30*(5), 555-574.

Byrne, B.M., Shavelson, R.J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*, 456-466.

Camilli, G. (1992). A conceptual analysis of differential item functioning in terms of a multidimensional item response model. *Applied Psychological Measurement, 16*, 129-147.

Chan, D. (2000). Detection of differential item functioning on the Kirton adaptation-innovation iventory using multiple-group mean and covariance structure analyses. *Multivariate Behavior Research, 35*,169-200.

Cheung. G.W., & Rensvold, R.B. (1999). Testing factorial invariance across groups, A reconceptualization and proposed new method. *Journal of Management, 25*(1), 1-28.

Cheung. G.W., & Rensvold, R.B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233 – 255.

Cohen, A.S., & Kim, S.H. (1992). *UW system foreign language placement test validity study.* UW Eau Claire. Madison, WI: University of Wisconsin, Center for Placement Testing.

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

Cole, D.A., & Maxwell, S.E. (1985). Multitrait-multimethod comparisons across populations: A confirmatory factor analytic approach. *Multivariate Behavioral Research, 2*, 389-417.

Collins, W.C., Raju, N.S., & Edwards, J.E. (2000). Assessing differential functioning in a satisfaction scale. *Journal of Applied Psychology, 85*, 451-461.

Cunningham, I.C.A.M, Cunningham, W.H., & Green, R.T. (1973, November). A cross cultural study of subjective product attributes. *Proceedings of the Association of Consumer Research*, 82-98.

Davison, M.L., Chen, T. (1991). *Parameter invariance in the Rasch model.* Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 3-7, 1991).

Deal, T. E., & Peterson, K. D. (1994). *The leadership paradox: Balancing logic and artistry in schools.* San Francisco: Jossey-Bass.

De Champlain, A.F., & Gessaroli, M.E. (1991). *Assessing test dimensionality using an index based on nonlinear factor analysis.* Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 3-7, 1991).

De Champlain, A.F., & Gessaroli, M.E. (1996, April). *Assessing the dimensionality of item response matrices with small sample sizes and short test lengths.* Paper presented at the Annual Meeting of the National Council on Measurement in Education (New York, NY, April 9-11, 1996)

De Champlain, A.F., Gessaroli, M.E., Tang, K.L., & De Champlain, J.E. (1998). *Assessing the dimensionality of polytomous item responses with small sample sizes and short test lengths: A comparison of procedures.* Paper presented at the Annual Meeting of the American Educational Research Association (San Diego, CA, April 13-17, 1998).

Donovan, M.A., Drasgow, F., & Probst, T.M. (2000). Does computerizing paper-and-pencil job attitude scales make a difference? New IRT analyses offer insight. *Journal of Applied Psychology, 85*, 305-313.

Drasgow, F., & Kanfer, R. (1985). Equivalence of psychological measurement in heterogeneous populations. *Journal of Applied Psychology, 70*, 662-680.

Drasgow, F., Levine, M.V., & McLaughlin, M.E. (1991). Appropriateness measurement for some multidimensional test batteries. *Applied Psychological Measurement, 15*, 171-191.

DuFour, R. (1997, Spring). Functional as learning communities enables schools to focus on student achievement. *Journal of Staff Development, 18,* 56-7.

DuFour, R., & Eaker, R. (1998). *Professional learning communities at work: Best practices for enhancing student achievement.* Bloomington, IN: National Educational Service.

Embretson, S.E. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika, 56*, 495-515.

England, G.W., & Harpaz, I. (1983). Some methodological and analytic considerations in cross-national comparative research. *Journal of International Business Studies, 14*(3), 597-622.

Facteau, J.D., & Craig, S.B. (2001, April). Are performance appraisal ratings from different rating sources comparable? *Journal of Applied Psychology, 86*(2), 215.

Flowers, C.P. (1996). *A description and demonstration of the polytomous-DFIT framework.* Paper presented at the Annual Meeting of the American Educational Research Association (New York, NY, April 8-12, 1996).

Flowers, C.P., Oshima, T.C., & Raju, N.S. (1999). A description and demonstration of the polytomous-DFIT framework. *Applied Psychological Measurement, 23*, 309–326.

Flowers, C.P., Raju, N.S., & Oshima, T.C. (2002, April). *A comparison of measurement equivalence methods based on confirmatory factor analysis and item response theory.* Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA, April 2–4, 2002.

Furlow, C. F., & Fouladi, R.T. (2005). *The impact of missing data and differential item functioning for survey researchers using IRT models.* Paper presented at annual meeting of the American Educational Research Association (Monteal, Quebec, Canada, April 11-15, 2005).

Ghorpade, J., Hattrup, K., & Lackritz, J.R. (1999). The use of personality measures in cross-cultural research: A test of three personality scales across two countries. *Journal of Applied Psychology, 84*, 640-679.

Glas, C.A.W. (1992). A Rasch model with a multivariate distribution of ability. In M. Wilson (Ed.). *Objective measurement: Theory into practice, vol. 1*, pp. 236-258. Norwood NJ: Ablex.

Golembiewski, R.T., Billingsley, R., & Yeager, S. (1976). Measuring change persistency in human affairs: Types of change generated by OD designs. *Journal of Applied Behavioral Science, 12*, 133-157.

Gosz, J. K. & Walker, C. M. (2002). *An empirical comparison of multidimensional item response data using TESTFACT/NOHARM.* Paper presented at the Annual

Meeting of the National Council for Measurement in Education. New Orleans, Louisiana, April 2–4, 2002.

Hambleton, R.K., & Rovinelli, R.J. (1986). Assessing the dimensionality of a set of test items. *Applied Psychological Measurement, 10*, 287-302.

Hambleton, R.K., & Swaminathan, H. (1985). *Item response theory: Principles and applications.* Boston, MA: Kluwer-Nijhoff.

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

Hattie, J.A. (1985, June). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement, 9*(2), 139-164.

Hays, W.L. (1988). *Statistics, 4th ed.* New York: Holt, Rinehart and Winston, Inc.

Hidalgo-Montesinos, M. D., & Lopez-Pina, J. A. (2002, February). Two-state equating in differential item functioning detection under the graded response model with the Raju area measures and the Lord statistic. *Educational and Psychological Measurement, 62*(1), 32-44.

Holland, P.W., & Rosenbaum, P.R. (1986). Conditional association and unidimensionality in monotone latent trait models. *Annals of Statistics, 14*, 1523-1543.

Horn, J.L., & McArdle, J.J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18*, 117-144.

Hui, C.H., & Triandis, H.C. (1985). Measurement in cross-cultural psychology. *Journal of Cross-Cultural Psychology, 16*(2), 131-152.

Idaszak, J.R., Bottom, W.P., & Drasgow, F. (1988, November). A test of the measurement equivalence of the revised Job Diagnostic Survey: Past problems and current solutions. *Journal of Applied Psychology, 73*(4), 647-656.

Jackson, P., Wall, T., Martin, R., & Davids, K. (1993). New measures of job control, cognitive demand, and production responsibility. *Journal of Applied Psychology, 78*, 753-762.

Jansens, M., Brett, J.M., & Smith, F.J. (1995). Confirmatory cross-cultural research: Testing the viability of a corporation-wide safety policy. *Academy of Management Journal, 38*, 364-382.

Kelderman, H., & Rijkes, C.P.M. (1994). Loglinear multidimensional IRT models for polytomously scored items. *Psychometrika, 59*, 149-176.

Knight, G., & Hill, N. (1998). Measurement invariance in research involving minority adolescents. In V. McLoyd & L. Steinberg (Eds.), *Research on minority adolescents: Conceptual, methodological and theoretical issues* (pp. 183-210). Hilldale, NJ: Erlbaum.

Knoke, D., & Burke, P.J. (1980). *Log-linear models.* Beverly Hills, CA: Sage.

Knol, D.L., & Berger, M.P.F. (1991). Empirical comparison between factor analysis and multidimensional item response models. *Multivariate BehavioralResearch, 26*(3), 457-77.

Lee, V.E., Dedrick, R.F., & Smith, J.B. (1991, July). The effect of the social organization of schools on teachers' efficacy and satisfaction. *Sociology of Education,* 64(3), 190-208.

Lee, V.E., & Smith, J.B. (1996, February). Collective responsibility for learning and its effects on gains in achievement for early secondary school students. *American Journal of Education, 104*(2), 103-147.

Lim, R.G., & Drasgow, F. (1990). Evaluation of two methods for estimating item response theory parameters when assessing differential item functioning. *Journal of Applied Psychology, 75*, 164-174.

Linacre, J.M. (1994). *Many-facet Rasch measurement.* Chicago. MERSA Press (original work published 1989).

Linden, W.J.v.d.., & Hambelton, R. (Eds.). (1997)*Handbook of modern item response theory.* New York : Springer.

Linn, R.L., & Werts, C.E. (1979). Considerations in studies of test bias. *Journal of Educational Measurement, 8*, 1-4.

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

Luczak, S.E., Raine, A., & Venables, P.H. (2001). Invariance in the MAST across religious groups. *Journal of Studies on Alcohol, 62*, 834-837.

Luecht, R., & Miller, R. (1992). Unidimensional calibrations and interpretations of composite traits for multidimensional tests. *Applied Psychological Measurement,* 16, 279-293.

Marsh, H.W. 1(993). The multidimensional structure of academic self-concept: Invariance over gender and age. *American Educational Research Journal, 30*(4), 841-860.

Marsh, H.W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First-and higher order factor models and their invariance across groups. *Psychological Bulletin, 97*, 562-582.

Martin, L.R., & Friedman, H.S. (2000). Comparing personality scales across time: An illustrative study of validity and consistency in life-span archival data. *Journal of Personality, 68,* 85-110.

Maurer, T. J., Raju, N.S., & Collins, W.C. (1998). Peer and subordinate performance appraisal measurement equivalence. *Journal of Applied Psychology, 83*, 693-702.

McDonald, R.P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology 34*, 100-117.

McDonald, R.P. (1993, September). A scale-invariant treatment for recursive path models. *Psychometrika, 58*(3), 431-443.

McKinley, R., & Mills, C. (1985). A comparison of several goodness-of-fit statistics. *Applied Psychological Measurement, 19,* 49-57.

Meade, A.W., Ellington, J.K., & Craig, S.B. (2004, April). *Exploratory measurement invariance: A new method based on item response theory.* Symposium presented at the 19th Annual Conference of the Society for Industrial and Organizational Psychology, Chicago, IL.

Meade, A.W., & Lautenschlager, G.J. Michels, LC, & Gentry, W. (2004, October). A comparison of item response theory and confirmatory factor analytic methodologies for establishing measurement equivalence/invariance. *Organizational Research Methods, 7*(4), 361-387.

Meier, D. (1995). *The power of their ideas: Lessons for America from a small school in Harlem.* Boston: Beacon.

Millsap, R.E., & Everson, H. (1991). Confirmatory measurement model comparisons using latent means. *Multivariate Behavioral Research, 26*, 479-497.

Millsap, R.E., & Meredith, W. (2004). *Factorial invariance: Historical trends and new developments.* Paper presented at the "Factor Analysis at 100" Conference, May 13-15, 2004, L.L. Thurstone Psychometric Laboratory, University of North Carolina.

Mullen, M.R. (1995). Diagnosing measurement invariance in cross-national research. *Journal of International Business Studies, 26*, 573-596.

Nandakumar, R. (1994, Spring). Assessing dimensionality of a set of item responses: Comparison of different approaches. *Journal of Educational Measurement, 31*(1), 17-35.

Oshima, T.C., & Miller, M.D. (1992). Multidimensionality and item bias in item response theory. *Applied Psychological Measurement, 16*, 237-248.

Peterson, K.D., & Deal, T.E. (1998, September). How leaders influence the culture of schools. *Educational Leadership, 56*(1). 28-30.

Pentz, M.A., & Chou, C.P. (1994). Change from development and intervention. *Journal of Consulting and Clinical Psychology, 62*, 450-462.

Ployhart, R.E., Wiechmann, D., Schmitt, N., Saccor, J. M., & Rogg, K. (2002). The cross-cultural equivalence of job performance ratings. *Human Performance, 16*, 49-79.

Raju, N.S., van der Linden, W., & Fleer, P. (1995). An IRT-based internal measure of test bias with applications for differential item functioning. *Applied Psychological Measurement, 19*, 353-368.

Raju, N.S., Laffitte, L.J., & Byrne, B.M. (2002, June). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology, 87*(3), 517-529.

Reckase, M.D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement, 9*, 401-412.

Reckase, M.D. (1997, March). The past and future of multidimensional item response theory. *Applied Psychological Measurement, 21*(1), 25-36.

Reise, S.P., Widaman, K.F., & Pugh, R.H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin, 114*, 552-566.

Rensvold, R.B., & Cheung, G.W. (2001). Testing for metric invariance using structural equation models: Solving the standardization problem. In C.A. Schriesheim & L.L. Neider (Eds.), *Research in management: Equivalence in measurement, vol. 1.* (pp. 21-50). Greenwich, CT: Information Age.

Riordan, C.M., Richardson, H.A., Schaffer, B.S., & Vandenberg, R.J. (2001). Alpha, beta, and gamma change: A review of past research with recommendations for new directions. In C.A. Schriesheim & L.L. Neider, (Eds), *Research in management: Equivalence in measurement, vol. 1.* (pp. 51–97). Greenwich, CT.: Information Age Publishing.

Riordan, C.M., & Vandenberg, R.J. (1994). A central question in cross-cultural research, Do employees of different cultures interpret work-related measures in an equivalent manner? *Journal of Management, 20*, 643-671.

SAS (2004). *SAS/STAT software 8e* [Software manual]. Cary, NC: SAS Institute, Inc.

Schaubroeck, J., & Green, S.G. (1989). Confirmatory factor analytic procedures for assessing change during organizational entry. *Journal of Applied Psychology, 74*, 892-900.

Schmitt, N. (1982). The use of analysis of covariance structures to assess beta and gamma change. *Multivariate Behavioral Research, 17*, 343-358.

Seraphine, A.E. (2000, March). The performance of DIMTEST when latent trait and item difficulty distributions differ. *Applied Psychological Measurement, 24*(1), 82-94.

Singh, J. (1995). Measurement issues in cross-national research. *Journal of International Business Studies, 26*, 597-619.

Steenkamp, L.E. M., & Baumgartner, H. (1998). Assessing measurement invariance in crossnational consumer research. *Journal of Consumer Research, 25*, 78-90.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Stout, W. (1987, December). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika, 52*(4), 589-617.

Taris, T.W., Bok, I.A., & Meijer, Z.Y. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*, 78-90.

Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines. In H. Wainer & H. I. Braun (Eds.), *Test validity*. (pp. 147-169). Hillsdale, NJ: Lawrence Erlbaum.

Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P.W. Holland & H. Wainer (Eds.), *Differential item functioning.* (pp. 67-113). Hillsdale, NJ: Lawrence Erlbaum.

Triandis, H.C. (1994). Cross-cultural industrial and organizational psychology. In H.C. Triandis, M.D. Dunnette, & L.M. Hough (Eds.), *Handbook of industrial and organizational psychology, vol. 4, (2nd ed.)*. Palo Alto, CA: Consulting Psychologists Press, Inc.

van Abswoude, A.A.H., van der Ark, L.A., & Sijtsma, K. (2004, January). A comparative study of test data dimensionality assessment procedures under nonparametric IRT models. *Applied Psychological Measurement, 28*(1), 3-24.

Vanderberg, R. J., (2002). Toward a further understanding of and improvement in measurement invariance methods and procedures. *Organizational Research Methods, 5*(2), 139-59.

Vandenberg, R.J., & Lance, C.E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(4), 69.

Vandenberg, R.J., & Self, R.M. (1993). Assessing newcomers' changing commitments to the organization during the first 6 months of work. *Journal of Applied Psychology, 78*, 557-568.

Volodin, N.A., & Adams, R.J. (1995, April). *Identifying and estimating a D-dimensional item response model.* Paper presented at the International Objective Measurement Workshop, University of California, Berkeley, California.

Wherry, R.J., Sr., Naylor, J.C., Wherry, R.J., Jr., & Fallis, R.F.(1965). Generating multiple samples of multivariate data with arbitrary population parameters. *Psychometrika, 30*, 303-313.

Wilson, D.T., Wood, R., & Gibbons, R.D. (1991). *TESTFACT.* Chicago: Scientific Software.

Windle, M., Iwawaki, S., & Lerner, R.M. (1988). Cross-cultural comparability of temperament among Japanese and American preschool children. *International Journal of Psychology, 23*, 547-567.

Winter, R., & Prohaska. J. (1983) Quantitative analysis of qualitative data. *Psychometrika*, 48(4), 417-448.

WINSTEPS. (1999). *Rasch-model computer program.* Chicago: MESA press. MESA Press.

Wright, B.D., & Masters, G.N. (1982). *Rating scale analysis.* Chicago: MESA Press.

Wu, M. L. (1997). *The development and application of a fit test for use with marginal maximum likelihood estimation and generalized item response models.* Unpublished masters thesis, University of Melbourne.

Wu, M.L., Adams, R.J., & Wilson, M.R. (1998). *ACER ConQuest: Generalized item response modeling software (Version 1.0)* [computer program]. Melbourne, Victoria, Australia: Australian Council for Educational Research.

Xie, Yuyu. (2001). *Dimensionality, dependence, or both? An application of the item bundle model to multidimensional data.* University of California at Berkley.

Yoo, B. (2002). Cross-group comparisons: A cautionary note. *Psychology and Marketing, 19*, 357-368.

Yuen, A., & Ma, W. (2002). Gender differences in teacher computer acceptance. *Journal of Technology and Teacher Education 10*(3), 365-382.