This is to certify that the
dissertation entitled

APPLICATION OF MODEL-DRIVEN META-ANALYSIS
AND LATENT VARIABLE FRAMEWORK IN SYNTHESIZING
STUDIES USING DIVERSE MEASURES

presented by

Soyeon Ahn

has been accepted towards fulfillment
of the requirements for the

Ph. D	degree in	Department of Counseling, Educational Psychology and Special Education

Kymberly A. Maier
Major Professor's Signature

18 June 2008
Date

APPLICATION OF MODEL-DRIVEN META-ANALYSIS AND LATENT VARIABLE FRAMEWORK IN SYNTHESIZING STUDIES USING DIVERSE MEASURES

By

Soyeon Ahn

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Counseling, Educational Psychology and Special Education

2008

ABSTRACT

APPLICATION OF MODEL-DRIVEN META-ANALYSIS
AND LATENT VARIABLE FRAMEWORK IN SYNTHESIZING STUDIES
USING DIVERSE MEASURES

By

Soyeon Ahn

In spite of a growing interest in meta-analysis, the application of existing methodology faces numerous difficulties and limitations. In particular, the use of diverse measures in primary studies introduces two methodological concerns in the application of meta-analytic techniques. First, individual study effects can vary significantly depending on differences in measures employed. Second, the existing methodologies are limited in dealing with very sparse data structures, where effect size has its unique measurement characteristics.

In support of resolving these concerns, the current research proposes a method for handling a very sparse data structure of effect sizes that arises from variations in measures used in primary studies. The proposed model is based on model-driven meta-analysis, structural equation modeling with latent variables, and method-of-moments estimation technique. This study presents the model specification in which the true population relationship between two latent variables is estimated. A method to extract unknowns in estimating the relationship between two underlying constructs (Equation 3. 13) is discussed.

First, several Monte Carlo simulations are performed in order to examine the performance of the proposed estimator under different conditions. Results from simulations indicate that the proposed approach correctly estimates the desired population

parameter. MANOVA results show that the factor loadings and reliabilities of indicators have the largest effect on the bias and MSE values of the estimators.

Second, the application of the proposed approach is demonstrated by re-analyzing a sub-set of studies reviewed by Ahn and Choi (2004). The estimated strength of the relationship between teachers' subject matter knowledge and student achievement included in Ahn and Choi using the proposed method was smaller than the weighted mean correlation corrected for artifacts proposed by Hunter and Schmidt (1990, 1994) and the *z*-transformed variance-weighted mean correlation proposed by Shadish and Haddock (1994), but leads to the same inference.

Lastly, four practical considerations of the proposed approach were discussed, followed by a list of potential future research to resolve those limitations. In this section, I demonstrate how well the proposed approach estimates the strength of the relationship between two underlying constructs when it is based on a misspecified population model.

ACKNOWLEDGEMENTS

Though only my name appears on the cover of this dissertation, a number of great people have contributed to its production. I owe my gratitude to all who have made this dissertation possible and because of whom my graduate training experience has been one that I will never forget.

My deepest gratitude goes to my two spiritual mentors, Drs. Betsy J. Becker and Mary M. Kennedy. I have been exceptionally fortunate to have these two great mentors during my six-year graduate training at MSU. Their mentorship was paramount in providing a well-rounded experience consistent with my long-term career goals. There are only a few graduate students who are given the opportunity to develop their own individuality and self-sufficiency by being allowed to work with such independence. For everything you've done for me, Dr. Betsy J. Becker and Dr. Mary M. Kennedy, I thank both of you and I will pay back to my students. I hope that one day I would become as good an advisor to my students as the two have been to me.

I think a decision to have Dr. Becker as my academic advisor was the best choice I have ever made in my entire life. She gave me the freedom to explore on my own and at the same time the guidance to recover when my steps faltered. Her patience and support helped me overcome many crisis situations and finish this dissertation. She always read my terrible drafts from one day to another; she edited my grammar and APA disasters without a complaint. But, I am not yet perfect for APA format. I would also like to convey many special thanks to Dr. King Beach, who welcomed me whenever I came to Florida for a visit, and provided me very warm and practical advice when I was in difficult situations (in particular, one year ago).

Dr. Mary M. Kennedy! We, all TQ-QT girls (you should remember all three girls in TQ-QT), love your sense of humor (i.e., "piece of cake"), your thoughtful insights, practical and generous advice, never-ending support toward us, and values on research and teaching. Dr. Kennedy, everything is "*a piece of cake*", isn't it?

My co-chair, Dr. Kimberly S. Maier, has been always there to encourage me and give me practical advice to go through such a long and lonely journey. Dr. Maier had always been overwhelmingly generous with graduate students—but it was only after she became my advisor that I realized how committed she was to training, supporting and encouraging her students. I will never forget her warm and cheering note whenever I felt miserable.

I would like to thank Dr. Richard T. Houang for his insights and enormous assistance to solve the technical part of this dissertation. Dr. Houang was always open to my call for meetings and welcomed me with his great smile. When I was deadly frustrated, he was the one who shed light on problems and shared his knowledge and insights. Without his help, I doubt I would be able to complete the dissertation and graduate in time.

My thanks also go to Dr. Frederick L. Oswald, the other member of my dissertation committee. Dr. Oswald went well beyond his duty in reading and copiously commenting on my research.

I would like to acknowledge Dr. Ralph Putnam's help on developing the expert judgment assessment. He nicely shared his experience and knowledge to develop the assessment tool for expert judgments. In addition, I want to acknowledge my appreciation to five graduate students at Michigan State University.

I am also grateful to the following former or current TQ-QT staff members. Among them, special thanks go to Dr. Meng-jia Wu (at Loyola University, Chicago), Dr. Jinyoung Choi (at Ewha Womans University, Korea), and Rae-Seon (Sunny) Kim (at Florida State University). I am also thankful to Steve J. Pierce in Community Psychology at Michigan State University for his encouragement and practical advice on research.

Many friends have helped me stay sane through these difficult years. Their support and care helped me overcome setbacks and stay focused on my graduate study. I greatly value their friendship and deeply appreciate their confidence in me. I am also grateful to the Korean Buddhism Organization with which I shared numerous happy moments.

I also want to acknowledge Dr. Andrew C. Shin, who has shared five years of my last twenty. I owe him a lot and I will not forget everything he has done for me during the five years. When I was up till 3 a.m. in the main library, he was always there by my side.

Last, but not the least, I would like to convey my sincere appreciation to my parents and my younger brother in Korea. In particular, I want to say a very special thanks to my dad, Young Gil Ahn, who truly believes in me and supports my success during the entire 30 years of my life. I strongly believe that most of who I am has been built on my dad's very extraordinary support. Also, I really believe that my mom's sincere prayer to Buddha at 5 a.m. every morning truly worked for my success. Thank you very much, my lovely mom Okhee Shim!

PREFACE

Nearly six years of research experience in the Teacher Qualifications and Quality of Teaching (TQ-QT) project[1] under the direction of principal investigators Drs. Betsy J. Becker and Mary M. Kennedy at Michigan State University provided me a solid theoretical and practical background for completion of this dissertation. Approximately 500 studies that examine the relationship between teacher qualifications and quality of teaching vary tremendously and introduce several interesting methodological questions in research synthesis.

This dissertation focuses on how to combine studies when the original studies use diverse measures with different measurement characteristics such as reliability and validity, even though researchers intend these to represent the same underlying constructs. In this research, I have tried to develop an approach whereby we can combine the very sparse data structure that arises from large variations across studies in measures. The proposed method is based on the assumption that all measures are attempting to represent the same underlying construct even though their measurement characteristics are quite different.

The proposed approach is developed based on three existing ideas in statistics and measurement – model-driven meta-analysis, structural equation modeling (SEM) with latent variables, and a method-of-moments estimation technique. Even though the proposed method is built on a simple one-factor model, it is possible to expand this model to solve more complicated issues in meta-analysis. As presented in the section on practical considerations, more attention should be paid to developing a method that can

[1] For more detailed information, please see the website http://www.msu.edu/user/mkennedy/TQQT/

handle missing data in research synthesis. In addition, the robustness of the proposed model should be examined before applying the proposed model in practice.

It is customary to list a long series of acknowledgements somewhere in the preface of a dissertation. I have gained enormous personal and scientific benefits during my time spent on the TQ-QT project at MSU, both from the people with whom I have worked and the environment that they have created. I am only going to personally thank four people, my mentors Drs. Betsy J. Becker and Mary M. Kennedy (we often call them "Spiritual Mentors (SM)"), to whom I owe so much that it would be pointless to try to encapsulate it, Dr. Meng-Jia Wu (at Loyola University at Chicago), and Rae-Seon (Sunny) Kim (at Florida State University), who have played multiple roles as colleague, friend, and big sister. Their academic and emotional support helped me go through a long and sometimes lonely journey toward the completion of this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

From its first appearance, meta-analysis has been widely used in various disciplines including medicine, economics, psychology, epidemiology, and education (Chalmers, Hedges, & Cooper, 2002; Hedges, 1983; Slavin, 2008; Vanhonacker, Lehmann, & Sultan, 1990). In spite of a growing interest in meta-analytic techniques as a means of providing rigorous evidence in many fields (Borman, 2002; Slavin, 2008; Towne, Wise, & Winters, 2005), the application of existing methodology in research synthesis faces numerous difficulties and limitations due to the inherent nature of research in education and social sciences (Berk, 2006; Rubin, 1992; Slavin, 1984; Thum & Ahn, 2007).

1.1. Challenges of Research Synthesis in Education and Social Science

As Kennedy (2007) has pointed out, multiple factors simultaneously influence outcomes within naturally occurring settings in education and social sciences. Many researchers have thus used multiple regressions or hierarchical linear models to eliminate numerous confounding variables in the primary research (Kennedy, Ahn, & Choi, 2008). However, their study findings have been often excluded from meta-analyses (e.g., Ahn & Choi, 2004; Qu & Becker, 2003) because no generally accepted methods exist for integrating results of multiple regressions or hierarchical linear models (Becker & Schram, 1994; Becker & Wu, 2007; Wu, 2006a, 2006b). As discussed in Becker and Schram (1994), regression analyses, path analyses, canonical correlations, and factor analyses are not easily synthesized. This is because partial correlations provided in such

analyses seldom represent the same parameters, which vary depending on other variables included in the model.

For example, Ahn and Choi (2004) found that among 49 studies examining the relationship between teacher subject matter knowledge and student achievement in mathematics, 11 used regression analysis and 4 used more advanced data-analytic techniques such as hierarchical linear modeling (HLM) or structural equation modeling (SEM). Consequently, Ahn and Choi (2004) excluded those 15 studies from their meta-analysis, and synthesized only the remaining 34 studies that provided correlation coefficients between teacher knowledge and student achievement.

In addition, in the social sciences and education, no natural scales of measurements exist (Hedges & Olkin, 1985). Consequently, studies employ a variety of measures. While these may represent "*the same*" underlying construct (e.g., student learning, depression, or other broad constructs), meta-analysts often encounter difficulties in putting effects on a common outcome metric across studies using various measures (Rubin, 1992). As Bollen (1989) demonstrated, study findings (e.g., correlations or regression coefficients) differ if the measurement errors of indicators (with variations in the reliabilities of indicators) are introduced or if the factor loadings (i.e., validity) of indicators are not equal to one.

Choi, Ahn, and Kennedy (under review) discovered that 15 different measures of teacher knowledge in mathematics (e.g., Glennon Test of Mathematical Understanding, Test of Understanding of the Real Number System (TURNS), etc) were used across the 16 studies included in their meta-analysis on teachers' subject matter knowledge in mathematics. Similarly, Becker and Wu (2007) identified 79 unique measures of student

learning used to represent the quality of teaching across 65 studies that investigate the relationship between teacher qualifications and quality of teaching. Such use of diverse measures in the primary studies often introduces the following two methodological concerns in the application of meta-analytic techniques.

First, individual study effects can vary significantly depending on measurement differences in the variables employed (Baugh, 2002; Lipsey & Wilson, 2001; Nugent, 2006; Oswald & Converse, 2005; Oswald & Johnson, 1998; Rubin, 1992; Slavin, 1984). Thus, many researchers (Hunter & Schmidt, 1990; Oswald & Converse, 2005; Oswald & Johnson, 1998; Raju, Anselmi, Goodman, & Thomas, 1998; Raju, Burke, Normand, & Langlois, 1991; Raju, Fralicx, & Steinhaus, 1986) have proposed methods for correcting study effects for differences in measurement. Hunter and Schmidt's (1990) approach, which adjusts correlation coefficients for potential measurement artifacts including sampling error, measurement unreliability, and range restriction, has been widely adopted in social sciences, particularly in applied psychology.

However, some researchers (Lambert & Curlette, 1995; Oswald & Johnson, 1998) have demonstrated that the estimate of the population correlation ($\hat{\rho}$) obtained via the Hunter and Schmidt's approach does not always estimate the true value (ρ) and its associated variance (σ^2_{ρ}) is also somewhat inaccurate. For example, based on Monte Carlo simulations, Oswald and Johnson (1998) demonstrated that discrepancies between $\hat{\rho}$ and ρ get larger with small within-study sample sizes and with smaller numbers of effect sizes included in the meta-analysis.

Recently, Thum and Ahn (2007) have applied a latent variable framework in research synthesis and proposed to adjust for differences in regression coefficients due to

the factor loadings, the measurement errors, and the variances of latent variables before combining the coefficients. On the other hand, a number of limitations stand in the way of practical application of Thum and Ahn's approach. In particular, many components of the model are unreported, including the factor loadings of both criterion and predictor variables, an index of the true relationship between two constructs, and information on measurement errors. Even if reasonable priors on the unknowns can be selected, the estimation process outlined by Thum and Ahn requires information not easily available and thus practical applications may be limited.

The second concern is that the existing univariate or multivariate statistical modeling approaches for meta-analysis (e.g., the Generalized Least Squares (GLS) method presented by Raudenbush, Becker, & Kalaian, 1988) are limited in dealing with very sparse data structures, which occur when each effect size (e.g., correlation or regression coefficient) has its unique measurement characteristics for predictor and outcome variables.

For example, in the meta-analysis by Choi, Ahn, and Kennedy (under review), none of the correlation coefficients from 16 studies uses the same measures of both teacher's knowledge and student achievement in mathematics. In such a case, the GLS method, which is frequently used to combine non-independent effect-sizes in the meta-analysis, is inapplicable due to a singular design matrix for estimating the true population correlation coefficient and its variance.

1.2. Empirical Example

Figure 1.1 in the Appendix C displays studies included in the meta-analysis by Ahn and Choi (2004) that focuses on the effect of how much math teachers know on student learning in mathematics. In Figure 1.1 in the Appendix C, three aforementioned challenges in synthesizing studies are well delineated: 1) Studies provide results from diverse data-analytic techniques (e.g., correlation coefficients in Brown, 1988; regression coefficients in Chaney, 1995; HLM coefficients in Chiang, 1996). 2) Different sets of predictors (i.e., coursework, degree level, major, GPA, and test scores for teacher knowledge in mathematics) and outcome variables (i.e., California Achievement Test (CAT), National Assessment of Educational Progress (NAEP), and Iowa Test of Basic Skills (ITBS) for student achievement in mathematics) are used across studies. 3) Only two studies (i.e., Teddlie, Falk, & Falkowski, 1983 and Hill, Rowan, & Ball, 2005) provide exactly identical links between the same sets of predictors and criterion variables, leading to a very sparse data structure for further analyses.

1.3. Purpose of Research

The current research proposes a new methodology for handling a very sparse data structure of the effect sizes (i.e., correlations or regression coefficients) that mostly arises from the variations in the measures used in the primary studies. To accomplish this, I use a Structural Equation Modeling (SEM) approach with latent variables (Bollen, 1989), the ideas of model-driven meta-analysis (Becker & Schram, 1994), and a method-of-moments estimation technique (Casella & Berger, 1990; Gelman, 1995). This method

quantifies the relationship between two underlying constructs measured by different sets of indicators with unique measurement characteristics such as reliability and validity.

As Messick (1993) indicated, there are several ways of conceptualizing validity (e.g., content validity, criterion validity, predictive validity, etc). I use the term *validity* to refer to the structural relationship (correlation) between the indicator and its underlying construct, which can be understood based on a structural equations approach (Bollen, 1989). As Bollen (1989) pointed out, the validity of a measure is defined as the magnitude of the direct structural relation between the indicator and its associated construct.

In this dissertation, I first present the specification of a population model using model-driven meta-analysis and SEM with latent variables. Based on the specified population model, the true population relationship between two latent variables is quantified by applying the method-of-moments estimation technique. Moreover, three approaches are discussed for obtaining the unknown values needed to compute the method-of-moments estimator of the strength of the relationship between two underlying constructs. Then a series of Monte Carlo simulations is conducted to test the performance of the proposed approach under different conditions. Last, its practical application is demonstrated by synthesizing a set of studies that are reviewed by Ahn and Choi (2004), in which the relationship between teachers' subject matter knowledge and student achievement in mathematics was investigated.

CHAPTER 2

LITERATURE REVIEW

In many disciplines a variety of measures with different measurement characteristics are often used to represent "*the same*" underlying construct in the primary studies (Farley, Lehmann, & Ryan, 1981). For instance, Crowl, Ahn, and Baker (in press) reported that the parent-child relationship quality is measured in several ways across the 19 studies included in their meta-analysis. These measures include standard observational techniques, structured interviews, and several standardized assessments such as the Family Relations Test, the Parenting Stress Index, and the Dyadic Adjustment Scale.

Also, many studies no longer focus on only a few simple bivariate relationships (e.g., zero-order correlations), or differences (main effects) on a few outcomes (Becker, 2001; Becker & Schram, 1994). An example from an ongoing synthesis of the studies examining the relationship between teacher qualifications and the quality of teaching (TQ-QT)[2] indicates that only 55 out of 461 coded studies used a bivariate correlation analysis, and 17 others reported simple *t* tests. Most other studies examined the effect of teacher qualifications on the quality of teaching based on more advanced data analytic techniques such as multiple regression, Multivariate Analysis of Variance (MANOVA), and Analysis of Covariance (ANCOVA). In this section, I first review how these challenges have been handled in research synthesis.

[2] More details about the TQ-QT project can be found in http://www.msu.edu/user/mkennedy/TQQT/.

2.1. Meta-analytic Methods For Synthesizing Studies using Various Indicators

In the literature, three methods are often used to synthesize studies using various measures of the predictor and outcome variables. These are a univariate method, an artifact- correction approach, and a multivariate method.

2.1.1. Univariate Method

The first approach involves creating collections of studies that use the same measures and then performing a series of separate univariate analyses of effect sizes on each relationship. This is accomplished by calculating an average effect for each category (see Hedges & Olkin, 1985; Hunter & Schmidt, 1990; Shadish & Haddock, 1994) based on traditional research-synthesis techniques (e.g., the *z*-transformed variance-weighted average proposed by Hedges and Olkin (1985) or Rosenthal and Rubin (1991)).

For instance, Choi, Ahn, and Kennedy (under review) categorized 51 correlation coefficients extracted from 19 studies into 8 categories in terms of the content domain (i.e., arithmetic, algebra, and geometry) and the cognitive demands of the student mathematics achievement measure (i.e., computation, concepts, and applications). Then they obtained the *z*-transformed variance-weighted average estimates for 8 categories by performing a series of separate univariate analyses, one for each subgroup of studies.

A univariate data-analysis is often used due to its ease of application. However, it is limited when the interest is in an overall picture of interrelationships among all variables included in the model as a whole. Moreover, when individual studies contribute multiple measures of relationships, the univariate method ignores possible dependence in

the data, and thus might lead to inaccurate conclusions (Becker & Schram, 1994; Gleser & Olkin, 1994).

2.1.2. Artifact Correction

Some methodologists (e.g., Bollen, 1989; Nugent, 2006) have argued that effect sizes (i.e., standardized mean differences, correlation coefficients) based on variables with different measurement characteristics are not directly comparable. For instance, Nugent (2006) demonstrated that the distribution of the standardized mean difference, which is the most widely used scale invariant effect-size measure in the current practice of meta-analysis, varies depending on the reliabilities of measures used in the comparison groups. It has been also known that the correlation coefficient varies depending on the reliability of one or both measures (Baugh, 2002; Bollen, 1989; Hancock, 1997; Hunter & Schmidt, 1990, 1994).

Although most discussions have been limited to correlation coefficients, particularly in applied psychology, a number of researchers have suggested using the correction formulas with other effect-size measures such as regression coefficients and standardized mean differences attenuated due to measurement characteristics such as reliability and range restriction (Hunter & Schmidt, 1990; Oswald & Converse, 2005; Oswald & Johnson, 1998; Raju et al., 1986; Raju et al., 1991; Raju et al., 1998).

In fact, corrections for correlation coefficients are heavily used in the meta-analytic procedures proposed by Hunter, Schmidt, and Jackson (1982) and elaborated by Hunter and Schmidt (1990, 2004). Hunter and Schmidt (1990, 1994) have indicated that the study population correlation ρ_0 is always lower than the actual correlation ρ. This is

because we cannot do any study perfectly, and study imperfections produce the artifacts that systematically reduce the actual correlation parameter. Therefore, they have identified 10 possible sources of artifacts, and propose to correct the attenuated sample correlation by multiplying it by appropriate "artifact multipliers" a_i shown in Table 2.1 in the Appendix C.

After disattenuating each sample correlation using appropriate artifact multipliers a_i, the weighted mean correlation $\bar{r}$ is obtained by

$$\bar{r} = \sum w_s r_s / \sum w_s \tag{2.1}$$

where r_s is the s^{th} study correlation; the weight for study s suggested by Hunter and Schmidt is

$$w_s = N_s A_s^2, \tag{2.2}$$

where N_s is the sample size for study s, and A_s is the compound artifact multiplier for study s.

More elaborations of Hunter and Schmidt's method have been developed by a number of researchers (e.g., Le, 2003; Sackett & Yang, 2000 for correcting range restriction; Hancock, 1997; Raju & Brand, 2003; Raju, Burke, Normand, & Langlois, 1991 for correcting reliability and range restriction; Oswald & Converse, 2005 for correcting the unrestricted predictor reliability, the range-restricted criterion reliability, and the restricted validity coefficient). The focus of recent studies (Raju, Burke, Normand, & Langlois, 1991) has been on how to correct correlation coefficients for study artifacts when not all the included studies provide information related to study artifacts. Some researchers (Baugh, 2002; Bollen, 1989; Raju et al., 1986; Raju et al., 1991; Raju

et al., 1998) have also expanded their discussions to include attenuation in either unstandardized or standardized regression coefficients. More details can be found in Table 2.2 in the Appendix C.

However, some research has indicated that some of the correction formulas frequently used in research synthesis fail to fully eliminate the effects of study artifacts. Based on Monte Carlo simulations, Oswald and Johnson (1998) found that the Hunter and Schmidt's method, which corrects study artifacts, yields estimates of the population parameter that do not estimate properly the true value under some conditions, even for bivariate normal data. In addition, Lambert and Curlette (1995) have shown that the variance of the corresponding mean correlation coefficient can be greatly underestimated when some measures have skewed distributions of the predictor and criterion scores.

Such findings suggest that the existing methods for correcting the attenuation of correlation coefficients might not fully eliminate the consequences of study artifacts on effect-size measures. Moreover, no one has suggested how information on some of the artifacts can be obtained from primary studies. In particular, the construct validities of both predictor and outcome variables, which are briefly mentioned in Hunter and Schmidt (1990, 1994), are seldom reported in the primary studies. Considering that these artifact multipliers are not often reported, the application of this correction will be limited in practice unless methods are developed for obtaining the unreported values.

2.1.3. Multivariate Method

The third approach for combining dependent effect sizes from multiple measures is to use multivariate methods. By using multivariate methods, intercorrelations

(dependencies) among several effects can be taken into account. This should lead to a more accurate error rate and ensure that samples with more data do not over-influence the results (Becker & Schram, 1994).

The most frequently used multivariate approach is a Generalized Least Squares (GLS) method suggested by Raudenbush, Becker, and Kalain (1988). The GLS method is a feasible and flexible approach for analyzing multivariate data (Becker & Schram, 1994). Depending on how the covariances between correlations for the variance-covariance matrix $\hat{\Sigma}$ are computed, several variations of the GLS method have been proposed by Becker and Fahrbach (1994), Cheung (2000), Furlow (2003), and Furlow and Beretvas (2005).

In this section, a general overview of the GLS method is presented with a special focus on pooling correlation matrices. I begin by considering that the goal is to estimate the pooled $m \times m$ correlation matrix from the correlation coefficients which are reported in k studies using m variables. To accomplish the GLS analysis, the correlation coefficients should be stacked in a vector $\mathbf{r}$. The fixed-effects model for the correlation r_{sj} ($s = 1$ to k and $j = 1$ to m^*, $m^* = m(m-1)/2$) can be written as

$$r_{sj} = \rho_j + e_{sj}, \text{ for } s = 1 \text{ to } k, \text{ and } j = 1 \text{ to } m^*. \tag{2.3}$$

This model can be re-written as a multiple regression in matrix form, in which the product of a matrix $\mathbf{X}$ and a set of population correlations ρ_j predict a set of sample correlations. Specifically

$$\mathbf{r} = \mathbf{X}\boldsymbol{\rho}_\bullet + \mathbf{e}, \tag{2.4}$$

where the matrix **X** is a stack of $m^* \times m^*$ identity matrices for k studies, and identifies which correlations are estimated in each study and $\rho_\bullet = (\rho_1, ..., \rho_m)'$ contains the population correlations. The pooled correlations and their standard errors are estimated by the following GLS formula shown in Becker (1992)

$$\hat{\boldsymbol{\rho}}_\bullet = (\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}^{-1}\mathbf{r} \tag{2.5}$$

and

$$\hat{\mathbf{V}}(\hat{\boldsymbol{\rho}}_\bullet) = (\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}, \tag{2.6}$$

where **C** is the variance-covariance matrix among the correlations within studies included in the meta-analysis on the diagonal, with blocks of zeros in the upper and lower triangles. See Olkin and Siotani (1967) for formulas for **C**. Also, other ways of estimating **C** can be found in Becker and Fahrbach (1994); S. Cheung (2000); Cheung and Chan (2005); Furlow (2003); and Furlow and Beretvas (2005).

However, the application of the GLS method might be problematic for very sparse datasets, in which few studies use the same measures of variables of interest. This is because the design matrix **X** in equation 2.5 and equation 2.6 may become singular, and GLS analysis would be impossible when estimating the true population correlation coefficient and its variance.

2.2. Model-driven Meta-analysis

Becker (e.g., Becker, 2001; Becker & Schram, 1994; Whiteside & Becker, 2000) described model-driven meta-analysis as an efficient tool to deal with the growing complexity of primary studies in research synthesis. Becker (2001) refers to the model-

driven meta-analysis as a review that incorporates models from the substantive theory and informs us about the strength of relations posited by a population model. In a model-driven meta-analysis, the interrelationships among multiple constructs or measures that are explicit in the model are individually as well as simultaneously examined. Eventually, a model-driven meta-analysis can delineate a more complete system of relationships among constructs or variables than a traditional synthesis and provide a model for making further predictions based on real or hypothetical predictor values.

Becker and Schram (1994) discuss the rationale for employing models in synthesizing studies. First, they emphasize the importance of theory and theoretical models in primary studies, which are useful to verify or refute competing models. Similarly, a model-driven meta-analysis can help the reviewer build a stronger basis of explanation for the mechanisms behind a phenomenon of interest. Second, a model-based research synthesis can provide an overall picture of patterns among variables across the existing studies, by piecing together parts of a process that has been studied by different researchers or studied using different samples. Last, they point out that theoretical models can also guide reviewers in the conduct of the review process, much as they can help the conduct of primary research.

In a model-driven meta-analysis, models can arise empirically or be derived from theory (Becker, 1997). Figure 2.1 in the Appendix C shows one example of a model used in the meta-analysis conducted by Whiteside and Becker (2000), in which multiple factors affecting child outcomes including externalizing symptoms, internalizing symptoms, social skills, and cognitive skills are investigated. As seen in Figure 2.1, models are often illustrated using flowcharts or path diagrams. Such a diagram has two

components – boxes representing a construct or a set of constructs, and arrows representing paths indicating interrelationships among a set of constructs. In Figure 2.1, Whiteside and Becker have used 14 boxes representing variables or constructs (10 for predictors, and 4 for outcomes), and 19 arrows representing paths for interrelationships (including bidirectional relationships) among 14 variables or constructs. Due to the limited number of studies, a slightly reduced model was finally estimated in their meta-analysis. More details can be found in Whiteside and Becker (2000).

Based on Cooper's five stages of the review (1982), Becker (1992, 1997, 2001) drew parallels for incorporating models in conducting a model-driven meta-analysis. At the first stage of *problem formulation*, the models can guide reviewers to conceptualize the problem, define the constructs, and determine study relevance, even though they could also limit the generalization from the review by limiting variables and underlying constructs. At the *data collection* stage, researchers can easily establish explicit inclusion rules. This can occur because researchers who set up their models are fully informed about the research related to their own model and the research on competing models. The next stage is *data evaluation*, in which reviewers judge the procedural adequacy of studies in the review. At this stage, models can be used to identify and code aspects of study features, extract outcomes, and determine the type of data that will be used in data analysis. At the *data analysis* stage, models allow reviewers to test not only individual paths, but also interrelationships among several constructs or variables in the models. Furthermore, researchers can examine the extent to which the relationships posited in the models are observed in the data. At the *public presentation* stage, reviewers are expected

to describe explicitly the use of models in each stage. This helps readers evaluate the generalizability of the findings from the proposed model in a model-driven meta-analysis.

As Becker (2001) mentioned, the major benefit of employing a model-driven meta-analysis is its capacity to provide information about different theoretical and empirical models. Moreover, researchers can obtain the overall picture of a complicated system reflected in the primary studies, by estimating interrelationships among constructs or variables specified in the models. Consequently, the synthesized models can be useful to establish the validity of proposed models against other competing models and to help further formulate stronger explanations for the mechanisms of the phenomenon.

However, several statistical and practical problems in synthesizing models have been identified. One of the most prominent issues is the missing data problem, which can occur as the result of several causes (e.g., researchers may contribute to publication bias by failing to report nonsignificant results (the file-drawer problem), or all the variables of interest for the meta-analysis may not be included in any specific study). Missing data at the synthesis level can make estimation impossible or difficult. Also, a sufficiently large sample size is required for performing a model-driven meta-analysis. Other practical, but less technical issues concern 1) variations in defining the constructs across studies, 2) between-studies and within-study variation in synthetic models, 3) sources of artifactual variation, and 4) model misspecification.

2.3. Structural Equation Modeling with Latent Variables

Bollen (1989) argues that structural equation models with latent variables encompass two general model types. One is *a latent variable model* that summarizes the structural relationship between latent variables as

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}, \tag{2.7}$$

where $\boldsymbol{\eta}$ is the vector of latent endogenous random variables; $\boldsymbol{\xi}$ represents the latent exogenous random variables; $\mathbf{B}$ is the coefficient matrix showing the effect of the latent endogenous variables on each other; and $\boldsymbol{\Gamma}$ is the coefficient matrix for the effects of $\boldsymbol{\xi}$ on $\boldsymbol{\eta}$.

The second component is *a measurement model* that specifies the structural relation of observed to latent variables as

$$\mathbf{x} = \boldsymbol{\Lambda}_{\mathbf{X}}\boldsymbol{\xi} + \boldsymbol{\delta}, \tag{2.8}$$

and

$$\mathbf{y} = \boldsymbol{\Lambda}_{\mathbf{Y}}\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \tag{2.9}$$

where $\mathbf{y}$ and $\mathbf{x}$ are vectors of observed variables; $\boldsymbol{\Lambda}_{\mathbf{X}}$ and $\boldsymbol{\Lambda}_{\mathbf{Y}}$ are the factor-loading matrices that show the relations of $\mathbf{x}$ to $\boldsymbol{\xi}$ and $\mathbf{y}$ to $\boldsymbol{\eta}$, respectively; and $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$ are the errors of measurement for $\mathbf{y}$ and $\mathbf{x}$.

2.3.1. Structural Equation Modeling in Meta-Analysis

Although other statistical methods (e.g., a standardized regression equation from the pooled correlation matrix) can be used to obtain an empirical synthesized model, many researchers (e.g., Becker, 1992; S. Cheung, 2000; Cheung & Chan, 2005; Furlow,

2003) have applied structural equation modeling (SEM) to model-driven meta-analysis. In general, the application of structural equation modeling in the meta-analysis involves two steps. The two-step approach in meta-analytic SEM entails first pooling a correlation matrix across studies included in the meta-analysis, and then performing the SEM by inputting the pooled correlation matrix into standard SEM software such as LISREL or EQS. The meta-analytic SEM has been widely employed in literature (e.g.,Brown & Peterson, 1993; Hom, Caranikas-Walker, Prussia, Griffeth, 1992; Premack & Hunter, 1988; Schmidt, Hunter, & Outerbridge, 1986), focusing on a path analytic method (e.g., Cheung & Chan, 2005; Furlow, 2003).

However, a few researchers (i.e., Cheung & Chan, 2005) have recently applied meta-analytic SEM to estimate a confirmatory factor analysis (CFA) model (Furlow, 2003). Cheung and Chan (2005) have proposed a slightly different technique, which is called the 2-stage structural equation modeling (TTSEM) method. In their TTSEM method, the correlation matrices are first pooled using the technique of multiple-group analysis in SEM, and then the pooled correlation matrices are used to fit the CFA model. Advances in Cheung and Chan's method are 1) to introduce observed variables and their corresponding constructs in the model, and 2) to estimate factor loadings and measurement errors of observed variables for measuring their constructs in the synthesized model.

2.3.2. Latent Variable Framework in Meta-analysis

Recently, Thum and Ahn (2007) have introduced a latent variable framework for synthesizing studies. The latent variable model consists of a *measurement model* that

specifies the relation of observed to latent variables and a *latent variable model* that shows the influence of latent variables on each other. Thum and Ahn (2007) suggested the application of the latent variable model to reach the ultimate goal of research synthesis -- to understand the true relationship among constructs represented by the latent variables, which are measured using various indicators across the included studies.

If the objective in each study i is to reveal the underlying relationship among specific unobserved constructs say, γ, the relationship between γ and each study-specific estimate, say, ($\hat{\beta}_i$) based on the observable indicators employed has a predictable functional relationship that ties the observable indicators to their respective constructs. Furthermore, Thum and Ahn analytically showed that the study-specific estimates (i.e., ordinary least square (OLS) regression coefficients, $\hat{\beta}_i$) can be related to the underlying relationship among unobserved constructs γ based on validity and reliability, the covariance among constructs, sampling factors, and misspecifications of the structural model. Therefore, Thum and Ahn proposed to first adjust study-specific estimates using their respective measurement and structural parameters, and then obtain an average effect. A simulation by Thum and Ahn indicates that the average estimate of the OLS regression coefficients corrected by the reliabilities of predictors and validities of predictors and outcomes is the least unbiased of several estimates.

However, a number of limitations stand in the way of practical application of Thum and Ahn's approach in the real world. In particular, many components for correcting OLS regression coefficients are seldom reported, including factor loadings of both criterion and predictor variables, and information on measurement errors. Even if

reasonable priors on unknowns can be selected, the estimation process outlined by Thum and Ahn is quite complicated and thus practical applications are limited.

2.4. Method of Moments Estimation Technique

The method of moments is the oldest method of finding point estimators (Gelman, 1995), which is to estimate the population parameters such as mean, variance, median and etc. of a probability distribution by matching theoretical moments to specified values (Casella & Berger, 1990). This method is preferable to other approaches because it is simple in that it always provides some sort of estimate.

Let $X_1,...,X_n$ be a sample from a population with probability density function $f(x \mid \theta_1,...,\theta_k)$ with finite moments $E[x^k]$. Methods-of-moments estimators are obtained by equating the first k sample moments to the corresponding k population moments, and solving the resulting system of simultaneous equations. The sample consists of n observations, $x_1,...,x_n$. The k^{th} raw or uncentered moments are

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i^1,\ \mu_1 = Ex^1,$$

$$m_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2,\ \mu_2 = Ex^2,$$

$$\vdots$$

$$m_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k,\ \mu_k = Ex^k. \qquad (2.10)$$

The population moments μ_j will typically be a function of $\theta_1,...,\theta_k$, say $\mu_j(\theta_1,...,\theta_k)$. The method-of-moments estimator $(\tilde{\theta}_1,...,\tilde{\theta}_k)$ of $(\theta_1,...,\theta_k)$ is obtained by solving the following system of equations for $(\tilde{\theta}_1,...,\tilde{\theta}_k)$ in terms of $(m_1,...,m_k)$:

$$m_1 = \mu_1(\theta_1,...,\theta_k),$$

$$m_2 = \mu_2(\theta_1,...,\theta_k),$$

$$\vdots$$

$$m_k = \mu_k(\theta_1,...,\theta_k). \qquad (2.11)$$

The method-of-moments estimation technique is preferable to other estimation techniques such as Fisher's maximum likelihood estimation technique, if the family of probability models is not known or when estimating parameters of a known family of probability distributions (Gelman, 1995). It also provides consistent estimators of parameters (Greene, 1997). However, the method-of moments estimators are not necessarily efficient and sufficient. Therefore, the method-of-moments estimators are often used as the first approximation to the solutions of the likelihood equations or a Bayes prior (Gelman, 1995).

CHAPTER 3

METHODOLOGIES

As discussed in the previous sections, meta-analysts face challenges and difficulties in synthesizing studies when the original studies use diverse measures. Study effects vary considerably depending on the differences in measures employed and thus they are not directly comparable. In addition, data can be too sparse to apply the existing univariate or multivariate meta-analytic methods. Therefore, the existing methods are unable to fully resolve these challenges in research synthesis.

As a result, I propose a methodology in which the strength of the relationship between two latent variables is estimated. In this proposed approach, the underlying population model that is applied to all included studies is first formulated based on two perspectives; one is based on model-driven meta-analysis, and the other is structural equation modeling with latent variables. Then the final estimator, in which the strength of the relationship between two constructs that are measured differently across studies is quantified, is obtained by applying the method-of-moments estimation technique.

3.1 Model Specification

Suppose that the primary goal in the meta-analysis is to understand the strength of relationship between two latent variables, the exogenous (ξ) and endogenous (η) variables, which is represented by γ. All k studies in the meta-analysis provide study-specific effects (i.e., correlations or regression coefficients) estimating γ from a set of predictors $\mathbf{x} = \left[x_1, x_2, ..., x_{p-1}, x_p \right]$ and different outcome variables

from $\mathbf{y} = \left[y_1, y_2, ..., y_{q-1}, y_q \right]$. As shown in Figure 3.1 in the Appendix C, for instance, the first study may provide a zero-order correlation coefficient between x_1 and y_1, and the k^{th} study reports regression coefficients predicting y_2 using x_3 and x_p.

Figure 3.2 in the Appendix C specifies the population model that underlies the k included studies in the hypothetical meta-analysis. Our primary goal in the meta-analysis is to estimate γ from the study-specific effects linking observed predictors $\mathbf{x} = \left[x_1, x_2, ..., x_{p-1}, x_p \right]$ and criterion variables $\mathbf{y} = \left[y_1, y_2, ..., y_{q-1}, y_q \right]$. Each of these represents its corresponding underlying constructs, $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$, with different accuracy.

3.2. Structural Equation Modeling with Latent Variables

The underlying measurement model delineated in Figure 3.2 implies that the indicator variables and their corresponding latent variable are related. Specifically,

$$\mathbf{x} = \boldsymbol{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta}, \tag{3.1}$$

$$\mathbf{y} = \boldsymbol{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\varepsilon}, \tag{3.2}$$

where $\mathbf{x}$ ($p \times 1$) and $\mathbf{y}$ ($q \times 1$) are vectors of observed variables; $\boldsymbol{\Lambda}_\mathbf{x}$ ($p \times c$, c is the total number of ξ) and $\boldsymbol{\Lambda}_\mathbf{y}$ ($q \times d$, , d is the total number of $\boldsymbol{\eta}$) are the factor-loading matrices that show the relations of $\mathbf{y}$ to $\boldsymbol{\eta}$ and $\mathbf{x}$ to ξ, respectively; and $\boldsymbol{\delta}$ ($p \times 1$) and $\boldsymbol{\varepsilon}$ ($q \times 1$) are the errors of measurement for $\mathbf{y}$ and $\mathbf{x}$. The errors in $\boldsymbol{\varepsilon}$ are assumed to be uncorrelated with $\boldsymbol{\eta}$, ξ and $\boldsymbol{\delta}$, and $\boldsymbol{\delta}$ is in turn uncorrelated with $\boldsymbol{\eta}$, ξ and $\boldsymbol{\varepsilon}$.

Let $\mathbf{T}$ be the $p + q$ dimensional column vector of both indicators $\mathbf{x}$ and $\mathbf{y}$ $\mathbf{T} = \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_{p-1} & x_p & y_1 & y_2 & \cdots & y_{q-1} & y_q \end{bmatrix}'$. The corresponding population covariance matrix of $\mathbf{T}$ is schematized as

$$\Sigma = \Sigma(\mathbf{T}) = \begin{bmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{bmatrix}. \tag{3.3}$$

The covariance matrix $\Sigma(\mathbf{T})$ consists of four submatrices: (1) the covariance matrix among the ys, $\Sigma_{\mathbf{yy}}$, (2) the covariance matrix of x with y, $\Sigma_{\mathbf{xy}}$, (3) the transpose of the covariance matrix of x with y, $\Sigma_{\mathbf{yx}}$, and (4) the covariance matrix among the xs, $\Sigma_{\mathbf{xx}}$.

Let us consider the implied covariance matrix of y, $\Sigma_{\mathbf{yy}}(\mathbf{T})$. It is

$$\begin{aligned} \Sigma_{\mathbf{yy}}(\mathbf{T}) &= \mathbf{E}(\mathbf{yy}') = \mathbf{E}[(\Lambda_{\mathbf{y}}\boldsymbol{\eta} + \boldsymbol{\varepsilon})(\Lambda_{\mathbf{y}}\boldsymbol{\eta} + \boldsymbol{\varepsilon})'] \\ &= \Lambda_{\mathbf{y}}\mathbf{E}(\boldsymbol{\eta}\boldsymbol{\eta}')\Lambda'_{\mathbf{y}} + \Theta_{\boldsymbol{\varepsilon}}, \end{aligned} \tag{3.4}$$

where $\Theta_{\boldsymbol{\varepsilon}}$ is the $q \times q$ variance-covariance matrix of $\boldsymbol{\varepsilon}$.

The covariance matrix of x with y, $\Sigma_{\mathbf{xy}}(\mathbf{T})$, and its transpose, $\Sigma_{\mathbf{yx}}(\mathbf{T})$, are equal to

$$\begin{aligned} \Sigma_{\mathbf{xy}}(\mathbf{T}) &= \mathbf{E}(\mathbf{xy}') = \mathbf{E}[(\Lambda_{\mathbf{x}}\boldsymbol{\xi} + \boldsymbol{\delta})(\Lambda_{\mathbf{y}}\boldsymbol{\eta} + \boldsymbol{\varepsilon})'] \\ &= \Lambda_{\mathbf{x}}\mathbf{E}(\boldsymbol{\xi}\boldsymbol{\eta}')\Lambda'_{\mathbf{y}}, \end{aligned} \tag{3.5}$$

and

$$\begin{aligned} \Sigma_{\mathbf{yx}}(\mathbf{T}) &= \mathbf{E}(\mathbf{yx}') = \mathbf{E}[(\Lambda_{\mathbf{y}}\boldsymbol{\eta} + \boldsymbol{\varepsilon})(\Lambda_{\mathbf{x}}\boldsymbol{\xi} + \boldsymbol{\delta})'] \\ &= \Lambda_{\mathbf{y}}\mathbf{E}(\boldsymbol{\eta}\boldsymbol{\xi}')\Lambda'_{\mathbf{x}}. \end{aligned} \tag{3.6}$$

Finally, the covariance matrix of x, $\Sigma_{xx}(T)$, is written as

$$\begin{aligned}\Sigma_{xx}(T) &= E(xx') = E[(\Lambda_x\xi + \delta)(\Lambda_x\xi + \delta)'] \\ &= \Lambda_x E(\xi\xi')\Lambda_x' + \Theta_\delta,\end{aligned} \tag{3.7}$$

where Θ_δ is the $p \times p$ variance-covariance matrix of δ.

If I assemble equations (3.4) - (3.7) into a single matrix $\Sigma(T)$, the population covariance matrix for the sets of indicator variables is

$$\begin{aligned}\Sigma(T) &= \begin{bmatrix}\Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy}\end{bmatrix} \\ &= \begin{bmatrix}\Lambda_x E(\xi\xi')\Lambda_x' + \Theta_\delta & \Lambda_x E(\xi\eta')\Lambda_y' \\ \Lambda_y E(\eta\xi')\Lambda_x' & \Lambda_y E(\eta\eta')\Lambda_y' + \Theta_\varepsilon\end{bmatrix}.\end{aligned} \tag{3.8}$$

3.3. Estimation

From the population covariance matrix shown in Equation 3.8, let us focus on the covariance matrix of x with y, $\Sigma_{xy}(T)$. If I assume all x and y indicators are standardized with mean of 0 and variance of 1, the covariance matrix of y with x, $\Sigma_{xy}(T)$, becomes a matrix of population correlations,

$$\Sigma_{xy}(T) = E(xy') = \begin{bmatrix} \rho_{x_1y_1} & \rho_{x_2y_1} & \cdots & \rho_{x_{p-1}y_1} & \rho_{x_py_1} \\ \rho_{x_1y_2} & \rho_{x_2y_2} & \cdots & \rho_{x_{p-1}y_2} & \rho_{x_py_2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{x_1y_{q-1}} & \rho_{x_2y_{q-1}} & \cdots & \rho_{x_{p-1}y_{q-1}} & \rho_{x_py_{q-1}} \\ \rho_{x_1y_q} & \rho_{x_2y_q} & \cdots & \rho_{x_{p-1}y_{q-1}} & \rho_{x_py_q} \end{bmatrix}. \tag{3.9}$$

Applying Equation 3.5, this correlation matrix can be written as

$$\begin{aligned}
\boldsymbol{\Sigma}_{\mathbf{xy}}(\mathbf{T}) &= \mathbf{E(xy')} \\
&= \boldsymbol{\Lambda}_{\mathbf{x}}\mathbf{E}(\xi\boldsymbol{\eta}')\boldsymbol{\Lambda}'_{\mathbf{y}} \\
&= \begin{bmatrix}
\rho_{x_1 y_1} & \rho_{x_2 y_1} & \cdots & \rho_{x_{p-1} y_1} & \rho_{x_p y_1} \\
\rho_{x_1 y_2} & \rho_{x_2 y_2} & \cdots & \rho_{x_{p-1} y_2} & \rho_{x_p y_2} \\
\vdots & \vdots & & \vdots & \vdots \\
\rho_{x_1 y_{q-1}} & \rho_{x_2 y_{q-1}} & \cdots & \rho_{x_{p-1} y_{q-1}} & \rho_{x_p y_{q-1}} \\
\rho_{x_1 y_q} & \rho_{x_2 y_q} & \cdots & \rho_{x_{p-1} y_{q-1}} & \rho_{x_p y_q}
\end{bmatrix} \\
&= \begin{bmatrix}
\lambda_{x_1}\lambda_{y_1}E(\xi\eta') & \lambda_{x_2}\lambda_{y_1}E(\xi\eta') & \cdots & \lambda_{x_{p-1}}\lambda_{y_1}E(\xi\eta') & \lambda_{x_p}\lambda_{y_1}E(\xi\eta') \\
\lambda_{x_1}\lambda_{y_2}E(\xi\eta') & \lambda_{x_2}\lambda_{y_2}E(\xi\eta') & \cdots & \lambda_{x_{p-1}}\lambda_{y_2}E(\xi\eta') & \lambda_{x_p}\lambda_{y_2}E(\xi\eta') \\
\vdots & \vdots & & \vdots & \vdots \\
\lambda_{x_1}\lambda_{y_{q-1}}E(\xi\eta') & \lambda_{x_2}\lambda_{y_{q-1}}E(\xi\eta') & \cdots & \lambda_{x_{p-1}}\lambda_{y_{q-1}}E(\xi\eta') & \lambda_{x_p}\lambda_{y_{q-1}}E(\xi\eta') \\
\lambda_{x_1}\lambda_{y_q}E(\xi\eta') & \lambda_{x_2}\lambda_{y_q}E(\xi\eta') & \cdots & \lambda_{x_{p-1}}\lambda_{y_q}E(\xi\eta') & \lambda_{x_p}\lambda_{y_q}E(\xi\eta')
\end{bmatrix}.
\end{aligned} \tag{3.10}$$

Equation 3.10 suggests that the correlation between x_i and y_j, $\rho_{x_i y_j}$, can be written as a function of the factor loadings of x_i and y_j, λ_{x_i} and λ_{y_j}, where $i = 1$ to p, and $j = 1$ to q. Applying the method of moments by equating the sample moments to the corresponding population moments, and solving the resulting system of simultaneous equations (Casella & Berger 1990) leads to

$$\begin{aligned}
\mathbf{1'E(xy')1} &= \mathbf{[1'\Lambda_x][\Lambda'_y 1]E(\xi\eta')} \\
&= [\sum_{\mathbf{j=1}}^{\mathbf{q}} \sum_{\mathbf{i=1}}^{\mathbf{p}} \boldsymbol{\lambda}_{\mathbf{x_i}}\boldsymbol{\lambda}_{\mathbf{y_j}}]\mathbf{E}(\xi\boldsymbol{\eta}').
\end{aligned} \tag{3.11}$$

Recall that $\rho_{\xi\eta}$ is equal to $E(\xi\eta')$, where the means and variances of ξ and $\boldsymbol{\eta}$ are 0 and 1, respectively. Then Equation 3.11 becomes

$$\sum_{j=1}^{q}\sum_{i=1}^{p} E(\xi\eta')\lambda_{x_i}\lambda_{y_j}$$

$$= \rho_{\xi\eta}\sum_{j=1}^{q}\sum_{i=1}^{p}\lambda_{x_i}\lambda_{y_j} \qquad (3.12)$$

$$= \rho_{\xi\eta}[\sum_{j=1}^{q}\lambda_{y_j}(\sum_{i=1}^{p}\lambda_{x_i})].$$

From Equation 3.12, the correlation between two constructs ξ and η, $\rho_{\xi\eta}$ is written as

$$\rho_{\xi\eta} = \frac{\sum_{j=1}^{q}\sum_{i=1}^{p}\rho_{x_i y_j}}{\sum_{j=1}^{q}\lambda_{y_j}(\sum_{i=1}^{p}\lambda_{x_i})} = \hat{\gamma}. \qquad (3.13)$$

Therefore, if I know all of the population correlations between x_i and y_j ($\rho_{x_i y_j}$), and the factor loadings of x_i and y_j, λ_{x_i} and λ_{y_j}, the correlation $\rho_{\xi\eta}$ between two latent variables can be estimated. In general, however, I will estimate each correlation using the sample value of x_i and y_j, and I will also need estimates of the factor loadings. I discuss this issue in the next sections.

3.4. Information for Estimating $\rho_{\xi\eta}$

Two components are required to estimate $\rho_{\xi\eta}$ using the method-of-moments estimator shown in Equation 3.13. One is the set of estimates of the population correlation coefficients between x_i and y_j (i.e., the estimates of $\rho_{x_i y_j}$), and the other is

the factor loadings of x_i and y_j for measuring the exogenous (ξ) and endogenous (η) variables, λ_{x_i} and λ_{y_j}, respectively.

3.4.1. Population Correlation Coefficients Between *x*s And *y*s, $\rho_{x_i y_j}$

The population correlation coefficients can be estimated if studies provide zero-order correlation coefficients among x_i and y_j. Several methods for estimating mean population correlation coefficients from studies have been widely investigated and discussed (Becker, 1992; Becker & Schram, 1994; Raudenbush, Becker, & Kalaian, 1988; Wu, 2006a, 2006b). In the current research, two methods are used to estimate the population correlation coefficients $\rho_{x_i y_j}$. First, I average sample-size weighted observed correlation coefficients for each *x* and *y* pair (as in Hunter & Schmidt, 1990). A second method is to combine *z*-transformed variance weighted correlations (Shadish & Haddock, 1994).

First, the correlation coefficient estimates of $\rho_{x_i y_j}$ are obtained from the sample-size weighted mean of the observed correlation coefficients between the *x*s and *y*s:

$$\hat{\rho}_{x_i y_j} = \frac{\sum_{s=1}^{k} n_s r(x_i y_j)_s}{\sum_{s=1}^{k} n_s}, \tag{3.14}$$

where $r(x_i y_i)_s$ is the s^{th} reported correlation coefficient between x_i (i = 1 to p) and y_j (j = 1 to q) and k is the number of studies.

Second, the most often used univariate method in meta-analysis, which is to combine *z*-transformed correlations (Shadish & Haddock, 1994) will provide a second estimator. *Z*-transformed variance-weighted correlations are obtained by converting the correlation coefficients $r_{(x_i y_i)_S}$ s by Fisher's variance stabilizing *z* transform:

$$Z_{[r_{(x_i y_j)}]_S} = .5\{\ln[((1 + r_{(x_i y_j)_S}) / (1 - r_{(x_i y_j)_S})]\}, \tag{3.15}$$

where *ln* is the natural logarithm. If the underlying data are bivariate normal, the conditional variance of $Z_{[r_{(x_i y_j)}]_S}$ is

$$v_S = \frac{1}{(n_S - 3)}, \tag{3.16}$$

where n_S is the within-study sample size of the s^{th} study.

The *z*-transformed weighted average correlation coefficient is

$$\bar{Z}_{[r_{(x_i y_j)}]} = \frac{\sum_{s=1}^{k} w_S Z_{[r_{(x_i y_j)}]_S}}{\sum_{s=1}^{k} w_S}, \tag{3.17}$$

where w_S is a weight assigned to the s^{th} study. The weights are calculated by

$$w_S = \frac{1}{v_S}. \tag{3.18}$$

The estimate in the z metric shown in Equation 3.17 is then back-transformed to obtain $\hat{\rho}$ via

$$\hat{\rho} = \frac{\exp(2\bar{z}_{[r_{(x_i y_j)}]}) - 1}{\exp(2\bar{z}_{[r_{(x_i y_j)}]}) + 1}. \tag{3.19}$$

3.4.2. Factor loadings (Validity coefficients)

Factor loadings or validity coefficients of observed variables are rarely reported in primary studies. For instance, only one study included in a meta-analysis by Choi, Ahn, and Kennedy (under review) provided a validity coefficient of the indicators representing how teacher tests measured teachers know math knowledge. Therefore, these values need to be estimated using other information provided in the studies or by other means.

In the case where all studies provide the correlation matrix among all variables used in studies, factor loadings of all variables are easily computed. Consider a simple one-factor, three-indicator model:

$$\begin{bmatrix} \mathbf{x_1} \\ \mathbf{x_2} \\ \mathbf{x_3} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\lambda}_{\mathbf{x_1}} \\ \boldsymbol{\lambda}_{\mathbf{x_2}} \\ \boldsymbol{\lambda}_{\mathbf{x_3}} \end{bmatrix} [\xi] + \begin{bmatrix} \boldsymbol{\delta_1} \\ \boldsymbol{\delta_2} \\ \boldsymbol{\delta_3} \end{bmatrix}, \tag{3.20}$$

where ξ is uncorrelated with δ_i ($i = 1, 2, 3$). This leads to the following relationship:

$$\begin{bmatrix} var(x_1) & & \\ cov(x_1,x_2) & var(x_2) & \\ cov(x_1,x_3) & cov(x_2,x_3) & var(x_3) \end{bmatrix} = \begin{bmatrix} \lambda_{x_1}^2 \phi_1 + var(\delta_{x_1}) & & \\ \lambda_{x_1}\lambda_{x_2}\phi_1 & \lambda_{x_2}^2 \phi_1 + var(\delta_{x_3}) & \\ \lambda_{x_1}\lambda_{x_3}\phi_1 & \lambda_{x_2}\lambda_{x_3}\phi_1 & \lambda_{x_3}^2 \phi_1 + var(\delta_{x_3}) \end{bmatrix}. \tag{3.21}$$

To ensure the model is identified, I set ϕ_1 to 1 (Bollen, 1999). Then, the covariances among the x_i s are computed as

$$\text{cov}(x_1, x_2) = \lambda_{x_1} \lambda_{x_2}, \text{cov}(x_1, x_3) = \lambda_{x_1} \lambda_{x_3}, \text{cov}(x_2, x_3) = \lambda_{x_2} \lambda_{x_3}. \tag{3.22}$$

Likewise, if I have correlations among all variables used in all studies, their factor loadings are easily computed based on Equation 3.22. The same logic can be applied to obtain the factor loadings of y_j. However, if no information is provided (e.g., no study provides correlations among x_1, x_2, and x_3), they must be approximated based on other information provided in each study. More details for obtaining factor loadings of variables used in the studies are discussed below.

3.5. Extracting Unknowns in the Model

If no correlation coefficients among the x_i or y_j exists, the following two methods can be used to estimate factor loadings of variables used in the studies. One is based on the reliabilities of the observed variables, which are fairly frequently reported in primary studies. The other uses expert judgments about the validities of variables.

3.5.1. Use of Reliability Information

Considering that the reliabilities of measures are likely to be reported, it would be reasonable to use them to estimate the factor loadings of the indicators. Bollen (1989) introduced an alternative way to define the reliability based on classical test theory as well as the measurement model. Based on classical test theory (Allen & Yen, 1979; Crocker & Algina, 1986), the observed score (x) can be written as

$$x = \tau + e, \tag{3.23}$$

where τ is the true score, e is the measurement error score or error of measurement, and the expected value of measurement error is assumed to equal to 0. Thus the expected value of x is τ_i.

In addition, the true scores τ depend on the latent variables ξ such that

$$\tau = \Lambda_X \xi + s, \tag{3.24}$$

where Λ_{x_i} is the coefficient that specifies the structural relationship between τ and ξ, and s represents specific variance unrelated to ξ and to e. Substituting equation 3.24 into equation 3.23 leads to

$$x = \Lambda_X \xi + s + e. \tag{3.25}$$

Since the reliability is the ratio of true score variance to the observed score variance, I can write the reliability of x_i as

$$\rho_{x_i x_i} = \frac{\text{var}(\tau_i)}{\text{var}(x_i)} = \frac{\lambda_{x_i}{}^2 \phi_1 + \text{var}(s_i)}{\text{var}(x_i)}. \tag{3.26}$$

From equation 3.26, if the variance of the latent variable (ϕ_i) is set to 1, the specific variance equals 0, and the variance of x_i is known or can be estimated, the x_i factor loading can be written as

$$\hat{\lambda}_{x_i} = \sqrt{\rho_{x_i x_i} * \text{var}(x_i)}. \tag{3.27}$$

The same logic can be applied to estimate the y_j factor loading as

$$\hat{\lambda}_{y_j} = \sqrt{\rho_{y_j y_j} * \text{var}(y_j)}. \tag{3.28}$$

3.5.2. Use of Expert Judgments

The second method is to use expert judgments about the factor loadings of indicator variables x_i and y_j. Each content expert as an independent rater would be asked to provide information regarding how well each of the indicators used in the

studies represents the corresponding underlying constructs. When judging the validity of each indicator, experts are expected to read the individual studies carefully, and then rank order all indicators in terms of each one's relation to its corresponding construct[3]. Experts would also be asked to provide an approximate value for the validity coefficient of each indicator.

According to Thurstone's (1927) discrete utility model, raters rank indicators based on their utilities, in this case their validities, which are unobserved and vary across respondents (Maydeu-Olivares & Bőckenholt, 2005). I shall denote by t_i the latent random variable associated with the validity for an indicator x_i'. If a respondent prefers an indicator x_i over an indicator x_o his or her perception of the validity of an indicator x_i', u_{x_i}, should be larger than that *n* indicator x_o. This can be specified as

$$u_{x_i'} = \begin{cases} 1 \text{ if } t_{x_o} \geq t_{x_o} \\ 0 \text{ if } t_{x_o} < t_{x_o} \end{cases}. \tag{3.29}$$

The response process shown in equation 3.29 can also be written in terms of differences between the latent utilities, $u^d_{x_i} = t_{x_i} - t_{x_k}$, where $u^d_{x_i}$ is the latent comparative response. Then, u_{x_i} can be re-written as a function of $u^d_{x_i}$:

$$u_{x_i} = \begin{cases} 1 \text{ if } u^d_{x_i} \geq 0 \\ 0 \text{ if } u^d_{x_i} < 0 \end{cases}. \tag{3.30}$$

[3] A possible protocol for obtaining expert judgments is shown in Appendix A.

Then, the latent comparative response as a linear function of latent random variable u_{x_i} is

$$\mathbf{U^d} = \mathbf{At}, \tag{3.31}$$

where $\mathbf{A}$ is an design matrix, consisting of p choice alternatives in the columns and the $\tilde{p}$ paired comparisons in the rows. Note that there are p predictors and q outcomes.

For instance, suppose that individual experts rank order 4 measures [A, B, C, D] in terms of their validity. In the design matrix $\mathbf{A}$, each column corresponds to one of the 4 choice alternatives [A, B, C, D]. The corresponding rows give the six paired comparisons [A, B], [A, C], [A, D], [B, C] [B, D], [C, D]. Thus the design matrix $\mathbf{A}$ is

$$\mathbf{A} = \begin{bmatrix} \mathbf{1} & \mathbf{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{1} & \mathbf{0} & \mathbf{-1} & \mathbf{0} \\ \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{-1} \\ \mathbf{0} & \mathbf{1} & \mathbf{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{-1} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{-1} \end{bmatrix}. \tag{3.32}$$

A different ordering, say [A, C, B, D] would lead to a different $\mathbf{A}$ design matrix.

Assuming that the vector of latent utilities t_{x_i} is normally distributed in the population of respondents, the mean and covariance structure of $u^d_{x_i}$ are

$$\boldsymbol{\mu}_{u^d_{x_i}} = \mathbf{A}\boldsymbol{\mu}_{\mathbf{t}_{x_i}}, \tag{3.33}$$

and

$$\boldsymbol{\Sigma}_{u^d} = \mathbf{A}\boldsymbol{\Sigma}_{\mathbf{t}_{x_i}}\mathbf{A}'. \tag{3.34}$$

As Maydeu-Olivares and Bőckenholt (2005) pointed out, the Thurstonian ranking model can be estimated using the SEM framework. Figure 3.3 in the Appendix C depicts the covariance structure shown in 3.34 as a SEM model for a ranking model with four choice alternatives. In figure 3.3, there are six observed variables $u^d_{x_i}$, and four latent validities t_{x_i}. In fact, $u^d_{x_i}$ are not actually observed, but their dichotomizations u_{x_i} are observed.

In SEM, parameters in the structured multivariate normal distributions of $u^d_{x_i}$ that have been dichotomized according to a set of thresholds are estimated in several stages (Muthen, 1978). First, the thresholds and the tetrachoric correlations among the underlying normal variables are estimated. Then, the parameters are estimated from the thresholds and tetrachoric correlations. The thresholds and tetrachoric correlations are obtained first by standardizing the latent responses $u^d_{x_i}$. The standardized latent response $Z(u^d_{x_i})$ is computed as

$$\mathbf{Z(u^d_{x_i}) = D(u^d_{x_i} - \mu_{u^d_{x_i}}),} \tag{3.35}$$

where **D** is a diagonal matrix with the reciprocals of the standard deviations of $u^d_{x_i}$ on the diagonal:

$$\mathbf{D = [Diag(\Sigma_{u^d_{x_i}})]^{-1/2}}. \tag{3.36}$$

Then, the standardized latent responses $Z(u^{d}_{x_i})$ are multivariate normal with mean 0 and correlation matrix $P_{u^{d}_{x_i}}$, where

$$\mathbf{P}_{Z(u^{d}_{x_i})} = \mathbf{D}(\boldsymbol{\Sigma}_{u^{d}_{x_i}})\mathbf{D} = \mathbf{D}(\mathbf{A}\boldsymbol{\Sigma}_{x_t}\mathbf{A}')\mathbf{D}. \tag{3.37}$$

Also, the standardized latent difference responses $Z(u^{d}_{x_i})$ are related to the observed u_{x_i} by a threshold relationship as

$$u_{x_i} = \begin{cases} 1 \text{ if } z(u^{d}_{x_i}) \geq \tau_{x_i} \\ 0 \text{ if } z(u^{d}_{x_i}) < \tau_{x_i} \end{cases}. \tag{3.38}$$

Since there are $\tilde{p}$ paired comparisons, there will be $\tilde{p}$ thresholds τ_{x_i}. The vector of τ_{x_i} values has the following structure, which is proven in Maydeu-Olivares and Bőckenholt (2005)

$$\boldsymbol{\tau} = -\mathbf{D}\mathbf{A}\mu_{t_{x_i}}. \tag{3.39}$$

Thus, the parameters of interest $\mu_{t_{x_i}}$ and $\Sigma_{t_{x_i}}$ are estimated using equations 3.33 – 3.34. The estimation process, which is a SEM with categorical indicators, can be performed using standard SEM software such as LISREL, EQS or MPLUS.

CHAPTER 4

SIMULATION

A number of Monte Carlo simulations are conducted to test the performance of the proposed approach under different conditions. In these simulations, I estimate the relationship between exogenous variables ξ and endogenous variables η, each of which is measured using 3 indicators x_i and y_j, with $\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} = [x_1, x_2, x_3, y_1, y_2, y_3]'$. It is also assumed that these indicators are each standardized with a mean of 0 and a standard deviation of 1.

In each hypothetical meta-analysis, the data to be combined are from a series of k independent studies, in which the s^{th} study reports zero-order correlation coefficients $r_{x_i y_j}$, with population correlation coefficients $\rho_{x_i y_j}$, where $i = 1,2,3$, $j = 1,2,3$, and $s = 1,2,...,k-1,k$. The sample correlation coefficients ($r_{x_i y_j}$) are obtained from a fixed sample size of 30 in each study.

4.1. Data Generation

R (R Development Core Team, 2008) version 2.6.2 is used to generate data and examine the performance of the proposed approach for estimating the correlation between two latent constructs.

The method-of-moments estimator given in Equation 3.13 may be affected by the features of the population model underlying the meta-analysis, including how well x_i and y_j represent the underlying constructs (i.e., the validities of indicators), the total number of sample correlation coefficients included in the hypothetical meta-analysis (i.e., sample

size, the number of studies), the number of sample correlation coefficients $r_{x_i y_j}$ to be included for each study, and how reliable the predictors and outcome variables are. Also, estimates are likely to be affected by the amount of missing data (i.e., the number of *X-Y* correlations that are not reported) and the quality of missing data, so several missing-data conditions are investigated.

For a series of hypothetical meta-analyses, the sample correlation coefficients $r_{x_i y_j}$ are generated from a multivariate normal distribution of $n = 30$ cases per pseudo study, for the vector containing x_i and y_j, with $\begin{bmatrix}\mathbf{X}\\ \mathbf{Y}\end{bmatrix} = [x_1, x_2, x_3, y_1, y_2, y_3]'$, assumed to have mean vector of $\mathbf{0}$ and variance-covariance matrix of $\mathbf{\Sigma}$. $\mathbf{\Sigma}$, shown in Equation 3.10, is determined from the factor loadings of x_i and y_j ($\Lambda_{x_i}, i = 1,2,3$ and Λ_{y_j}, $j = 1,2,3$), the true relationship between ξ and η, and the measurement errors of x_i and y_j (δ_i, $i = 1,2,3$ and ε_j, $j = 1,2,3$).

4.1.1. Choice of Parameters

The parameters to be varied in the simulations are the index of true relationship between ξ and η (γ), the reliabilities of the predictor and outcome variables, the number of studies (*k*) included in the hypothetical meta-analysis, and the quantities of missing $r_{x_i y_j}$ values. The first two simulation parameters are used to create the variance-covariance matrix $\mathbf{\Sigma}$ that generates the zero-order correlations for each study. The next two sets of simulation parameters represent the characteristics of studies included in each hypothetical meta-analysis.

True relationship between ξ and η. Two values are used to characterize the true relationship between the two underlying constructs. These values are selected to investigate how well the proposed model estimates its true relationship 1) when there is no relationship between ξ and η (i.e., $\gamma = 0$), or 2) when there is a medium and positive relationship between ξ and η (i.e., $\gamma = .5$).

Reliabilities of xs and ys. Four sets of the reliabilities of indicators, x_i and y_j, (i.e., $\{\rho_{x_1x_1}, \rho_{x_2x_2}, \rho_{x_3x_3}, \rho_{y_1y_1}, \rho_{y_2y_2}, \rho_{y_3y_3}\} = \{.9,.9,.9,.9,.9,.9\}$, $\{.5,.5,.5,.5,.5,.5\}$, $\{.2,.2,.2,.2,.2,.2\}$, $\{.9,.5,.2,.9,.5,.2\}$) are used to compute the factor loadings of x_i and y_j based on Equations 3.27 – 3.28. Three values of reliabilities - .9, .5, and .2 - are chosen to represent high, medium, and low reliabilities of the indicators, respectively.

Once the factor loadings of x_i and y_j are determined, the variances of the measurement errors are obtained from $\text{var}(x_i) = \lambda^2_{x_i}\phi + \text{var}(\delta_{x_i})$, and $\text{var}(y_j) = \lambda^2_{y_j}\phi + \text{var}(\varepsilon_{y_j})$. Since the variances of the indicators and underlying construct (ϕ) are set to 1 in this simulation, the variances of the measurement errors are obtained by

$$\text{var}(\delta_{x_i}) = 1 - \lambda^2_{x_i}, \tag{4.1}$$

and

$$\text{var}(\varepsilon_{y_j}) = 1 - \lambda^2_{y_j}. \tag{4.2}$$

Total number of studies included. In the published research syntheses in *Review of Educational Research* from 1990 to 2004 and *Psychological Bulletin* from 1995 to 2004, the number of independent studies, k, varied from 12 to 180 (Ahn & Becker, 2005). Ahn and Becker (2005) indicate that approximately 75% of meta-analyses were based on fewer than 40 studies. Therefore, two values (i.e., k = 9 and 36), which are multiples of nine, are used in this simulation, considering that 9 pairs of zero-order correlations using 3 *x*s and 3 *y*s can be generated from the population model. For each of k studies, as many as 9 sample correlation coefficients $r_{x_i y_j}$ are generated.

Number of missing $r_{x_i y_j}$. In practice, the reported correlation coefficients between x_i and y_j vary considerably. Since the population model is established under the assumption that 3 *x*s and 3 *y*s are observed in the primary studies, at least three pairs of zero-order correlation coefficients (i.e., $r_{x_1 y_1}$, $r_{x_2 y_2}$, and $r_{x_3 y_3}$) should be provided. Thus, the quantities of missing $r_{x_i y_j}$ values manipulated in the simulation varied from 0 (i.e., all nine $r_{x_i y_j}$ values are provided) to 6 (i.e., all other $r_{x_i y_j}$ except $r_{x_1 y_1}$, $r_{x_2 y_2}$, and $r_{x_3 y_3}$ are missing). Therefore, 7 variations (i.e., the number of missing *r*s equals 0, 1, 2, 3, 4, 5, and 6) are used in this simulation.

4.1.2. Replications

From 8 population variance-covariance matrices (i.e., 2 values of γ × 4 sets of reliabilities of *x*s and *y*s), a total of 112 meta-analyses (i.e., 2 values of k × 7 variations regarding the qualities of missing variables) are generated in this simulation. These 112

different conditions are replicated 1,000 times, leading to 112,000 hypothetical meta-analyses in the simulation.

4.2. Data Evaluation

The index of the relationship between the two constructs, which is the method-of-moments estimator based on Equation 3.13, is obtained as follows: 1) Nine estimated population correlation coefficients ($\hat{\rho}_{x_i y_j}$ s) are obtained from k $r_{x_i y_j}$ s by pooling the values of the $r_{x_i y_j}$ based on a sample-size weighted average (ES1) and a z-transformed variance-weighted average (ES2), 2) Sums of factor loadings of x_i s and sums of factor loadings of y_j s are computed, and 3) Two values of the index of relationship between the two constructs, which is shown in Equation 3.13, are computed from two sums of nine estimated population correlation coefficients in step 1, and the sums of the factor loadings from step 2.

These estimates ES1 and ES2 are compared to the strength of the true relationship between two latent variables (i.e., $\gamma = \rho_{\xi\eta} = 0$ and $\gamma = \rho_{\xi\eta} = .5$). In particular, the bias and mean-squared error (MSE) of the estimators are evaluated. Denoting each of the two effect-size estimators as $\hat{\delta}$ and the population effect size as δ, I computed

$\text{Bias}(\hat{\delta}) = \text{E}(\hat{\delta}) - \delta$, and

$\text{MSE}(\hat{\delta}) = [\text{Bias}(\hat{\delta})]^2 + \text{Var}(\hat{\delta})$,

where $\text{E}(\hat{\delta})$ is computed as the mean $\hat{\delta}$ value and $\text{Var}(\hat{\delta})$ is the empirical variance of the $\hat{\delta}$ values across the 1,000 replications for each combination.

Then, multivariate analysis of variance (MANOVA) was performed on the bias and MSE values of the estimators in order to examine the relative performance of the proposed methods for estimating the strength of the true relationship between the two underlying constructs. The simulation features used as factors in the MANOVAs are 1) the factor loadings of predictor and outcome variables, 2) the number of studies included in the hypothetical meta-analysis, and 3) the number of missing sample correlation coefficients.

4.3. Simulation Results

4.3.1. Estimators

Figure 4.1 and Figure 4.2 in the Appendix C display the distributions of the estimators (i.e., using a sample-size weighted average (ES1) and a *z*-transformed variance weighted average (ES2)). They are summarized for two γ values of 0 and .5. Figure 4.1 shows that both ES1 and ES2 are normally distributed with the mean of 0 when γ value is set to 0. And, Figure 4.2 displays that the strength of the relationship between two underlying constructs is underestimated with γ value of .5.

4.3.2. Bias and MSE of Estimators

Table 4.1 in the Appendix C presents the average bias and MSE values of ES1 and ES2 that represent the strength of the relationship between the two underlying constructs (γ). They are also summarized for two γ values of 0 and .5.

The biases and MSEs of the estimators when $\gamma = 0$ are .0001 and .008, respectively. This indicates that the strength of the relationship between the two

underlying constructs seems correctly and accurately estimated. No noticeable differences are found between the two estimators based on sample-size weighted average *r*s (ES1) and that computed from *z*-transformed variance-weighted *r*s (ES2) in terms of their bias and MSE values.

When γ is equal to .5, the bias values of ES1 and ES2 are .008 and .023 and the MSE values of ES1 and ES2 are .008 and .009, respectively. This shows that the proposed approach slightly overestimates the true relationship between two underlying constructs for $\gamma = .5$, and does so with less accuracy. Similar to the case with $\gamma = 0$, no noticeable differences between ES1 and ES2 are found in terms of the MSE. However, the bias value of ES2 is bigger than that of ES1.

Table 4.2 and Table 4.3 in the Appendix C show the average bias and MSE values of ES1 and ES2 according to three factors manipulated in this simulation – the reliabilities and factor loadings of indicators, the number of missing *r*s, and the number of studies (*k*). Table 4.2 shows that when $\gamma = 0$ the bias and MSE values of both estimators are not affected by the three factors used in the simulation.

In Table 4.3, however, when $\gamma = .5$, the bias and MSE values of the estimators depend on some factors (i.e., reliabilities and factor loadings of indicators and the number of studies included in the meta-analysis). In particular, as fewer studies are included or indicators with smaller reliabilities and factor loadings are included in the meta-analysis, the bias and MSE values of estimators get bigger. As is true for the estimators when γ is set to 0, no noticeable differences between ES1 and ES2 are found in terms of their bias and MSE.

The quality of this simulation is evaluated by comparing the average biases shown in Table 4.2 and Table 4.3 to those presented in Field (2001). Field (2001) reports mean effect sizes using two well-known methods of synthesizing correlation coefficients. They are 1) Hedges and Olkin (1985) or Rosenthal and Rubin (1991) (i.e., ES2 in this dissertation) and 2) Hunter and Schmidt (1990) (i.e., ES1 in this dissertation). He reports results for different average sample sizes, different numbers of studies in the meta-analysis, and different levels of population effect size for the homogeneous case (i.e., Table 1 *p.* 170 in Field (2001)) and the heterogeneous case (i.e., Table 4 *p.* 174 in Field (2001)). See Field (2001) for the simulation design in more detail.

Table 4.4 in the Appendix C displays the mean bias of *r* obtained from the simulation conducted by Field (2001). As shown in Table 4.4, when the population correlation between two underlying constructs is set to 0, the average mean bias obtained in this simulation is similar to those obtained by Field. For example, the mean biases of nearly 0 in this simulation with γ of 0 for both ES1 and ES2 based on 9 and 36 studies with 30 sample size (see Table 4.2) are equal to 0 in Field's results. No bias is found regardless of the values of the factor loadings and reliability of indicators and the number of missing correlations.

When the population correlation (γ) is set to .5, the mean bias values of both ES1 and ES2 without any missing *r*s (see Table 4.3) are similar to the values reported in Field (2001). However, the mean bias in this simulation with missing *r*s gets larger as the number of missing *r*s increases. With 3 missing *r*s, the mean bias of ES2 is closer to the mean bias for heterogeneous case shown in Field (2001).

4.3.2. Factors Affecting Estimators of the Strength of Relationship Between Two Underlying Constructs

The two indicators of the quality of the estimated strength of the relationship between two true constructs, Bias and MSE, are evaluated in relation to the following factors: 1) the number of studies (*k*), 2) the number of missing *r*s, and 3) the reliabilities (factor loadings) of the indicators. The MANOVAs examine the effects of these characteristics on the bias and MSE of two estimates.

Factors affecting bias and MSE of estimators when $\gamma = 0$. Table 4.5 and Table 4.6 in the Appendix C display results from MANOVAs for the bias of the estimators, when γ is equal to 0. As shown in Table 4.5, statistically significant differences are found in the bias values of the overall estimates across the levels of all factors. In particular, the significant Wilks' Lambdas indicate that the MSEs of estimators differ depending on the number of studies (*k*), the number of missing *r*s, the reliabilities and the factor loadings of the indicators. In addition, the univariate Analysis of Variance (ANOVA) for all factors shown in Table 4.6 in the Appendix C indicates negligible impact of these factors on the bias of the effect-size estimators. Also, the partial Eta-squares for all the study features manipulated in this simulation equal zero.

Table 4.7 in the Appendix C indicates that the three factors significantly affect the MSEs of the both estimators with $\gamma = 0$. In particular, the significant Wilks' Lambdas indicate that the MSEs of estimators differ depending on the number of studies (*k*), the number of missing *r*s, the reliabilities and the factor loadings of the indicators. In addition, the univariate Analysis of Variance (ANOVA) for all factors displayed in Table 4.8 in the Appendix C shows a statistically significant impact of these factors on the MSEs of the

effect-size estimators. The factor loadings of the indicators have the largest effect on the MSEs of the estimators with an Eta-square of .68. Also, the Eta-square of .53 for the number of studies included in the meta-analysis implies that this factor has a medium impact on their MSEs. Finally, the impact of the number of missing *r*s on the MSEs is negligible (e.g., the η^2 of the number of missing *r*s is .07).

Factors affecting bias and MSE of estimators when $\gamma = .5$. Table 4.9 in the Appendix C displays results from MANOVAs when γ is set to .5. Table 4.9 and Table 4.11 in the Appendix C show that the bias and MSEs of estimators differ depending on all factors used in this simulation. In particular, the significant Wilk's Lambdas suggest that the bias and MSEs of both ES1 and ES2 significantly differ depending on the number of studies (*k*), the number of missing *r*s, and the factor loadings and reliabilities of indicators. In addition, the univariate Analysis of Variance (ANOVA) presented in Table 4.10 in the Appendix C indicates that the reliabilities of indicators have the largest effect on the biases of estimators based on an Eta-square of .71.

Also, the univariate Analysis of Variance (ANOVA) for all factors displayed in Table 4.12 in the Appendix C indicates the statistically significant impacts of these factors on the MSEs of the effect-size estimators. While the number of missing variables has the smallest effect on the estimators' MSEs ($\hat{\eta}^2 = .12$), the factor loadings and reliabilities of indicators and the number of studies included in the meta-analysis have medium effects on the estimators' MSEs with Eta-square values of .62 and .40, respectively.

Below, the influence of each study factor on the bias and MSE values of both estimators is discussed.

True relationship between ξ and η. In Figure 4.3 and Figure 4.4 in the Appendix C, the average biases values of two estimates are displayed according to the true population relationship between the two underlying constructs (γ). In particular, the bias values of estimators with $\gamma = 0$ differ from those with $\gamma = .5$, indicating that when $\gamma = 0$ the proposed model correctly estimates the strength of the relationship between two underlying constructs (i.e., Bias is nearly 0 with $\gamma = 0$). However, no differences on the MSE values of estimators are found according to the true population relationship between the two underlying constructs (γ).

Reliabilities of indicators. Figure 4.5 in the Appendix C displays the mean biases of two estimators according to the reliabilities of indicators when γ is set to 0. Regardless of which set of the reliabilities of the indicators is used, the mean biases of the estimators are not far off from 0 and they do not noticeably differ from one another.

As shown in Figure 4.6 in the Appendix C, both ES1 and ES2 have higher bias values when the reliabilities of three indicators vary (i.e., .2, .5, and .9). However, no obvious differences are observed among the mean bias values in terms of the reliabilities of indicators.

Figure 4.7 and Figure 4.8 in the Appendix C compare the MSEs for different magnitudes of the indicators' reliabilities with γ of 0 and .5, respectively. Although no significant differences in the mean MSEs are found, the estimators based on indicators with the reliabilities of .2 have bigger mean square errors. This indicates that the accuracy of estimating the strength of the relationship between two underlying constructs is lower when *r*s arise from indicators with lower reliabilities.

Factor loadings of indicators. As shown in Figure 4.9 in the Appendix C, when γ is set to 0, mean bias values of both ES1 and ES2 almost equal 0.

As displayed in Figure 4.10 in the Appendix C, when *r*s based on the indicators with different factor loadings (i.e., .45, .71, and .95) are combined, the estimators have the highest mean bias values of estimators (i.e., .07).

Figure 4.11 in the Appendix C compares the MSEs of the estimators according to the factor loadings of indicators with $\gamma = 0$. Figure 4.10 shows that the MSE values of the estimators are highest when the factor loadings of indicators are .45. The MSEs are nearly zero when *r*s are based on factor loadings greater than .70, implying that the strength of relationship between two underlying constructs is accurately estimated when indicators with high factor loadings are combined.

Figure 4.12 in the Appendix C displays the MSEs of estimators according to the factor loadings of indicators with γ of .5. With γ of .5, the MSE values of indicators are nearly zero when *r*s from the indicators with the factor loadings of .95 are combined, while they get bigger when indicators with factor loadings of .45 are combined.

Number of studies (k). As shown in Figure 4.13 and Figure 4.14 in the Appendix C, there is no significant relationship between the number of studies included in the meta-analysis and the bias values of estimators.

However, *k* is negatively related to the MSEs of estimators, which is shown in Figure 4.15 and Figure 4.16 in the Appendix C. For example, the mean MSEs are bigger with *k* of 9, while they get smaller with *k* of 36. This makes sense since less information is available with fewer studies.

Number of missing rs. Figure 4.17 and Figure 4.18 in the Appendix C compare the biases of estimators depending on the number of missing zero-order correlations. Figure 4.17 shows the mean biases with γ of 0 do not noticeably differ depending on how many correlation coefficients are missing.

On the other hand, when γ is set to .5, the mean biases slightly increases when the number of missing *r*s increases, but they are close to 0. In particular, the mean bias values of estimators are nearly zero without any missing *r*s, while it is approximately .04 off from zero with only three *r*s (i.e., $r_{x1y1}, r_{x2y2}, r_{x3y3}$) included. However, regardless of the number of missing *r*s, the mean bias values are not far off from 0.

As shown in Figure 4.19 in the Appendix C, no relationship between the MSEs and the number of missing *r*s is found when γ is set to 0. However, when γ is set to .5, there is a slightly positive relationship between MSEs and the number of missing *r*s increases (see Figure 4.20 in the Appendix C).

Quality of missing rs. Lastly, the effect of which *r*s are included on the performance of the proposed approach is examined. This is accomplished by looking at the bias and MSE values of estimators when correlation coefficients from indicators with different reliabilities are included.

Table 4.13 in the Appendix C shows the correlation matrix of the six indicators in terms of their reliabilities. At least three zero-order correlations (i.e., $r_{x1y1}, r_{x2y2}, r_{x3y3}$) on the diagonal (i.e., shaded in Table 4.12 in the Appendix C) should be always included. The quality of *r*s is determined by the reliability of each indicator. For instance, the quality of $r_{x1,y2}$ equals to that of $r_{x2,y1}$ because they are based on two indicators with high reliability (i.e., .9) and medium reliability (i.e., .5).

For a simple demonstration here, I only present how the strength of the relationship between two underlying constructs is estimated by not including each one of six correlations, which are not shaded in Table 4.13 in the Appendix C.

Table 4.14 in the Appendix C shows the results from the analysis of variance (ANOVA) on the effect of which *r* is not included in the meta-analysis on the bias value of ES1. Since there is no significant difference between ES1 and ES2 in terms of their bias and MSE, I here present an ANOVA result for ES1.

As shown in Table 4.14, which *r* is not included in meta-analysis does not have an significant impact on the bias of the ES1 with γ of 0 ($F_{(5, 24000)} = 2.070, p = .07$). However, ANOVA result indicates that the bias of estimators depend on the quality of missing *r* when γ is equal to .5 ($F_{(5, 24000)} = 12.60, p < .05, \hat{\eta}^2 = .003$).

Figure 4.20 in the Appendix C compares the bias values of ES1 depending on which correlation coefficient is included in a meta-analysis, in addition to three zero-order correlations (i.e., $r_{x_1y_1}, r_{x_2y_2}, r_{x_3y_3}$) on the diagonal (i.e., three shaded areas in Table 4.12). As shown in Figure 4.20, ES1 includes correlation coefficients from indicators with high (i.e., .9) and medium (i.e. .5) reliabilities (i.e., $r_{x_1y_2}$ and $r_{x_2y_1}$) and it has smallest bias (i.e., |-.25|). However, ES1 including correlations from indicators with medium (i.e., .5) and low (i.e. .2) reliabilities (i.e., $r_{x_2y_3}$ and $r_{x_3y_1}$) has the biggest bias (i.e., |-.27|).

In addition, statistically significant results from pairwise comparisons shown in Table 4.15 in the Appendix C indicate that the bias of ES1 differ according to the quality of *r*. For example, the bias of ES1 including correlations from indicators with high

(i.e., .9) and medium (i.e. .5) reliabilities (i.e., $r_{x_1 y_2}$ and $r_{x_2 y_1}$) is significantly different from that based on correlations from indicators with medium (i.e., .5) and low (i.e. .2) reliabilities (i.e., $r_{x_2 y_3}$ and $r_{x_3 y_1}$).

4.4. Conclusions

Findings from this simulation in which the performance of the proposed approach shown in Equation 3.13 for estimating the strength of the relationship between the two underlying constructs is evaluated in relation to the different factors in this simulation are as follows:

First, the average bias and MSE values of the estimators are approximately zero when the true relationship between the two underlying constructs is set to 0 and .5,. Overall, no distinguishable difference between the bias and MSE values of ES1 and ES2 is found. The mean bias values of ES1 and ES2 from this simulation are comparable to those presented by Field (2001) with γ of 0 and .5.

Second, the bias values of the estimators differ according to the number of studies, the number of the missing *r*s, the factor loadings and the reliabilities of indicators with γ of both 0 and .5. Among them, the factor loadings and the reliabilities of indicators have the biggest effect on the bias of the estimators, while the number of the missing *r*s has the smallest effect on the bias of the estimators

Third, the MSE values of the estimators differ according to the number of studies, the number of missing *r*s, the factor loadings and the reliabilities of indicators. This is found in the MSE of the estimator with γ of both 0 and .5. The factor loadings and reliabilities of indicators have the largest effect on the MSE values of the estimators. The

number of studies included in the meta-analysis has the next largest effect on the MSEs of estimators.

Fifth, when no *r*s are missing, the mean bias and MSE values of the estimators become nearly zero with γ of 0 and .5, indicating that the strength of the relationship between the two underlying constructs is correctly estimated using the proposed approach. When all 6 *r*s are not included and $\gamma = .5$, the mean bias values of estimators are about .03. In addition, under these same conditions, the mean MSE values of estimators are about .01.

Lastly, a statistically significant effect on the bias of which correlation is included is found for γ of .5. This indicates that the reliabilities of the missing *r*s have a significant influence on the accuracy of the estimators.

CHAPTER 5

APPLICATION

The application of the proposed method is demonstrated by re-analyzing a subset of studies reviewed by Ahn and Choi (2004). In their meta-analysis, Ahn and Choi investigated the relationship between teachers' subject matter knowledge (SMK) and student learning (SL) in mathematics. In order to deal with considerable variation in measures, Ahn and Choi first categorized the included studies in terms of how teachers' SMK was measured (e.g., teachers' test scores and teachers' coursework in mathematics). Then they conducted a series of univariate analyses, one for each subgroup of studies.

However, their meta-analysis, based on a univariate method, is limited in its ability to portray the overall picture of the relationship between what teachers know and how much students learn in mathematics. For instance, one of the conclusions that Ahn and Choi drew in their meta-analysis was that the relationship between teachers' subject matter knowledge and student learning is stronger when *teachers' SMK is measured by teacher test scores*. Such a finding can be limited if I want to make an inference about the overall picture of the relationship between two underlying constructs - teachers' knowledge and student achievement in mathematics.

Therefore, I demonstrate here how to use my proposed method to assess the strength of the relationship between teachers' knowledge in math and students' achievement, each of which is measured differently across studies. I expect that the proposed method will deal with the sparse data structure that mostly comes from variation in measures and further provide an assessment of the strength of the relationship

between the two underlying constructs (i.e., teachers' knowledge and student achievement in mathematics).

5.1. Study Description

The main purpose of this meta-analysis is to understand the relationship between teachers' subject matter knowledge (SMK) and student learning (SL) in mathematics. Studies included in this research synthesis were drawn from a larger literature database gathered as part of the *Teacher Qualifications and the Quality of Teaching* (TQ-QT) study. The TQ-QT project aims to synthesize studies investigating the relationship between indicator(s) of teacher qualifications (TQ) and the quality of teaching (QT), which have been conducted in the United States since 1960 (Wu, Becker, & Kennedy, 2002). As of winter 2005, the data base included about 480 studies. More details about search criteria and selected studies in the TQ-QT study can be found at http://www.msu.edu/~mkennedy/TQQT.

From the 480 studies in the TQ-QT database, 27 studies including 18 dissertations, 4 journal articles, 1 conference paper, and 4 reports were included in Ahn and Choi's meta-analysis. Details about inclusion rules can be found in Ahn and Choi (2004). Among 27 studies, only 8 studies based on 6^{th} grade-level students are used in the demonstration of the proposed method. These 8 studies are a relatively homogenous group in terms of statistical analysis, grade level, whether reliability is reported, and content domain (i.e., arithmetic) of students' mathematical knowledge. For instance, all 8 studies provide correlation coefficients between teachers' subject matter knowledge as measured by tests and student learning.

However, these 8 studies vary in terms of the measures used to represent teachers' and students' knowledge; by test type – researcher-made local, researcher-made large-scale, or commercial; by whether the gain score metric is utilized in analyzing the student achievement test and in terms of the unit of analysis and the time interval between pre- and post-test in year. For instance, 7 studies used commercial student measures (i.e., CAT, SAT, SRA, CTBS, and ITBS), while 1 study used a researcher-made large-scale assessment of student learning. In addition, all except two studies (i.e., Turgoose, 1996 and Lampela, 1966) provide correlation coefficients based on gain scores. However, some (such as Caezza, 1969) are based on 2 year gains, while others (e.g., Cox, 1970) are based on 1 year gain scores. There are also differences in the unit of analysis across the 8 studies (i.e., student level is used in Caezza, 1969 and Cox, 1970 vs. classroom level in Bassham, 1962; Koch, 1972; Lampela, 1966; Moore, 1964; Prekeges, 1973, and Turgoose, 1996).

Let us closely look at the various measures that are used to represent both teachers and students' knowledge in mathematics in these 8 studies displayed in Table 5.1 in the Appendix C. Two studies (i.e., Bassham, 1962; Moore, 1964) used the same test (i.e., Glennon test of basic mathematical understanding) as a measure of teacher's subject matter knowledge. Bassham (1962) and Lampela (1966) used California Achievement Test (CAT) and Cox (1970) and Moore (1964) used the SRA Achievement Series as a measure of student's knowledge. However, no pair of studies provide correlation coefficients based on the same measures of teachers' and students' knowledge. Figure 5.1 in the Appendix C shows the empirically driven population model.

5.2. Method

Since the proposed approach in this research makes use of correlation coefficients among the predictor and outcome variables, correlation coefficients that estimate the strength of the relationship between teachers' subject matter knowledge and student achievement in mathematics are extracted from the 8 studies. Table 5.2 in the Appendix C displays the study characteristics of the 8 studies in more detail.

Four studies (i.e., Caezza, 1969; Cox, 1970; Moore, 1964; Prekeges, 1970) provide more than one correlation coefficient, which are from subtests of their student achievement tests. For example, Caezza provided three correlation coefficients from three subtests of the Stanford Achievement Test (SAT), which are SAT: Concept, SAT: Computation, and SAT: Application. When several correlation coefficients are provided for the same sample, the following two rules are applied to extract one independent study effect.

First, if available, a correlation coefficient from the total score is used. For instance, a correlation coefficient of .16 based on a *total* score on the Achievement Series (SRA) is obtained from Cox (1970). Second, if a study provides several correlation coefficients from subtests of the student achievement test, the average correlation coefficient is obtained. This rule was applied to three studies (Caezza, 1969; Moore, 1964, and Prekeges, 1970).

Among the 8 studies, only one study (i.e., Turgoose, 1996) provides the validity coefficients for the variables. Turgoose (1996) reports the concurrent validity for the Tests of Achievement and Proficiency (TAP) ranged from .69 to .79. Seven studies (i.e., Caezza, 1969; Cox, 1970; Lampela, 1966; Moore, 1964; Prekeges, 1973; Turgoose,

1996) present reliability information related to teachers' subject matter knowledge measures. And four studies (i.e., Caezza, 1969; Moore, 1964; Prekeges, 1973; Turgoose, 1996) provide the reliability of student learning measures in mathematics. Different types of reliability information are reported for the indicators, including Cronbach's alpha, test-retest, and KR-20 reliabilities.

For studies that do not report the reliability of indicators (e.g., Bassham, 1962), whenever possible reliability is obtained from other studies that use the same measure. For example, the reliability for the Glennon test of mathematical understanding (an indicator of teachers' math knowledge) is obtained from Moore (1964). Similarly, the reliability of the Achievement Series (SRA) test used in Cox (1970) is obtained from Darakjian and Michele (1982). Therefore, the reported reliabilities of .51 to .67 provided by Darakjian and Michele (1982) are used for the reliability of the SRA used in Cox (1970).

The specific procedure to estimate the relationship between teachers' SMK and student learning is as follows. First, the average population correlation coefficients between the two sets of observed indicators are computed by averaging sample-size weighted correlation coefficients and averaging z-transformed variance-weighted correlation coefficients. Second, the validity information for the indicators is extracted from three sources of available information. 1) The validity of indicators that are provided in studies or borrowed from other studies (i.e., Cox, 1980; Moore, 1964; Turgoose, 1996) is directly used, 2) If the reliability information is available, the validity is extracted using Equation 3.27 and Equation 3.28, and 3) When the reliability or the validity of indicators is not available from the individual study (e.g., the California

Achievement Test (CAT) used in Bassham (1962) and Lampela (1966)), the validity of the indicator is obtained from expert judgments. The detailed procedure for gathering the expert judgments about the validity of indicators is later discussed in more detail. Lastly, the strength of the relationship between SMK and students' learning is computed using Equation 3.13.

5.3. Expert judgments

When no information about the reliabilities and the validities of indicators is available, validity information of indicators is obtained from content experts in the domain. Here, I define content expert as a person who is fairly familiar with research on teachers' subject matter knowledge with mathematical teaching experience. Based on this definition of *content expert*, five graduate students focusing on mathematics education, from the Department of Counseling, Educational Psychology and Special Education and the Department of Teacher Education at Michigan State University, provided their judgments on the validity information of indicators used in 8 studies.

Those five content experts approximated the validities of indicators based on the protocol shown in Appendix A. The protocol was developed to help content experts make better judgments about the indicators' validities. Based on the concept of concurrent validity, experts were first asked to compare all measures that are present in the population model (see Figure 5.1) in terms of how closely each indicator represents what constitutes teachers' subject matter knowledge in mathematics. In the process, either the actual test items (for the Glennon test of basic mathematical understanding) or descriptions of the assessments (i.e., CAT, SAT, and ITBS) were provided.

Second, content experts are asked to rate how well each indicator measures the conceptual dimension of teachers' subject matter knowledge in mathematics. The conceptual dimension of teachers' subject matter knowledge is derived by integrating various researchers' definition of teachers' knowledge in mathematics. Many researchers (e.g., Ball, 1990a; Ball, 1990b; Hill et al., 2004; Shulman, 1986) have different but similar conceptual definitions of teachers' knowledge in mathematics. Therefore, I have adopted a working definition based on the number and operations standards for grade 6-8 that are suggested by the National Council of Teachers of Mathematics (NCTM)[4].

After rating how closely each indicator represents three dimensions of teachers' subject matter knowledge, experts are asked to rank order the indicators from the one most likely to measure teachers' subject matter knowledge to the least likely measure. Then, they are finally asked to provide the approximate value of a factor loading, which is essentially a correlation coefficient between what is measured in each indicator and what constitutes teachers' subject matter knowledge in arithmetic. It took an average of 1 hour for each content expert to produce their judgments on the validity of indicators.

Since I need the validity of CAT used by Bassham (1962) and Lampela (1966), I took the average of the validity values provided by five content experts and used it as an approximate factor loading of CAT. The validity of CAT obtained from the ratings of the five content experts was .329.

[4] Number and operations standards for grade 6-8 that are suggested by NCTM are available at http://standards.nctm.org/document/chapter6/numb.htm.

5.4. Results

The estimated strength of the relationship between teachers' subject matter knowledge and student achievement using Equation 3.13 was .0005 computed using the sample-size weighted average correlation coefficient (ES1), and .0006 using *z*-transformed weighted correlation coefficients (ES2). These nearly zero estimates based on the proposed method-of-moments estimator given in Equation 3.13 indicate that there is no relationship between how much teachers know and 6th grade student learning in mathematics.

ES1 = .0005 and ES2 = .0006 were also compared to the weighted mean correlation corrected for artifacts (i.e., reliabilities and construct validity of indicators) proposed by Hunter and Schmidt (1990, 1994) and to the *z*-transformed variance-weighted mean correlation proposed by Shadish and Haddock (1994). The weighted mean correlation corrected for artifacts is .007 and the *z*-transformed variance-weighted mean correlation is .008, all indicating that there is no relationship between how much teachers know and 6th grade student learning in mathematics. Although the same inference is drawn, the estimated strength of the relationship between teachers' subject matter knowledge ands student achievement using Equation 3.13 is much smaller than the estimators based on the methods proposed by Hunter and Schmidt (1990, 1994) and Shadish and Haddock (1994).

However, a nearly zero correlation between teachers' subject matter knowledge and student achievement should not be over-interpreted. As shown in the previous section, the method-of-moments estimators based on Equation 3.13 might be affected by several factors in the population model. Above all, none of the 8 studies provided all possible

pairs of correlation coefficients that are present in Figure 5.1 in the Appendix C. Since the effect of missing *r*s on the bias and MSEs of estimators is large, the estimators computed in the presence of missing correlation coefficients might be biased.

In addition, the estimated validity coefficients of measures might be incorrect. First of all, 8 studies reported the reliabilities of measures based on different methods such as test-retest and Cronbach alpha. No acceptable methods exist to put the different types of reliabilities of measures in a common scale (though all range between 0 and 1). Therefore, the obtained validities could be either over-estimates or under-estimates, because each type of coefficient may tap different sources of error.

In fact, if the true population correlation coefficient is 0 and thus study factors do not affect the bias values of estimators as found in the simulation, the estimators based on Equation 3.13 might not be far off. However, there is no certainty that the true correlation between teachers' subject matter knowledge and student achievement is 0, thus it would be unwise to ignore the effect of study factors on computing the strength of the relationship between two constructs.

CHAPTER 6

PRACTICAL CONSIDERATIONS

The proposed method provides an innovative way to deal with one of the challenges in research synthesis that comes from variations in measures. By using the proposed approach, a reviewer can combine the sparse data that arise from the large variations in measures. Thus, the strength of the relationship between the two underlying constructs can be estimated. However, the method's application in practice raises several methodological questions. In this section, some practical considerations of the proposed approach are discussed, followed by an outline of potential future research to examine those limitations.

First, the most critical issue in the proposed approach is that some components needed to compute the index of the relationship between the two underlying constructs given in Equation 3.13 are not often reported in primary studies. In particular, researchers virtually never report the factor loadings (a.k.a. the validity coefficients) of variables employed in the studies. Even though three alternative methods (i.e., use of correlation matrices, use of expert judgments, and use of reliability information) are suggested for obtaining validity information, these all might introduce errors in estimating the true relationship between two underlying constructs. One potential solution is to use other available sources to obtain information for estimating the factor loadings of variables. For instance, reliability information used for obtaining the factor loadings of variables can be acquired from test manuals, other similar studies using the same variables, or personal contact with study authors.

Second, the other issue in using the proposed approach is related to the missing data. As described above, the studies included in the meta-analysis do not always report all of the relationships or paths included in the population model. As discussed in the simulation, studies often also do not use exactly parallel models, instead using different predictors and outcome variables. Results from the simulation show that the number of missing *r*s and the quality of the *r*s have statistically significant impacts on the bias and MSE of estimators. Therefore, more attention should be paid to developing an approach that can deal with missing data in research synthesis.

Third, the proposed approach presumes that the included studies are based on the same population, which indicates the fixed-effect model. Without this assumption, study-specific effects (i.e., correlation coefficients) should not be combined. For instance, if there are significant differences in the correlations among teachers' subject matter knowledge and student achievement in terms of grade level (See Ahn & Choi, 2004), these correlations should not be pooled. In fact, the test for the homogeneity of the correlation matrices (Becker & Schram, 1994) can be used to confirm this assumption.

Fourth, the proposed method assumes that the theoretically or empirically driven model is correctly specified. In fact, the development of a population model that is as comprehensive as possible is required. However, it is highly probable that the empirically or theoretically driven model may be misspecified. For example, a meta-analyst may derive a one-factor model even though the underlying population model is really a two-factor model. Therefore, understanding the robustness of the proposed model can be helpful for the application of this model, so I here investigate how well the proposed

model estimates the strength of the relationship between two underlying constructs when it is based on a misspecified population model.

Let us suppose that the observed scores of indicators include one additional component, which is the specific variance (s_i) introduced in Equation 3.23. As discussed in section 3.5.1, the specific variance is unrelated to the underlying constructs or to the measurement errors. For instance, if teachers' test scores that represent how much they know in a subject depend on teachers' age, teachers' age introduces variation in their test scores that is not related to the underlying construct of teachers' subject matter knowledge or to measurement error. When the specific variance (s_i) of an indicator is nonzero, its factor loading can be smaller than its reliability (see the relationship between factor loadings and reliabilities described in Equation 3.25).

In order to examine the robustness of the proposed approach when specific variances of indicators are introduced, the bias and MSE values of the estimators are compared for different values of the specific variances of the indicators (i.e., SV = 0, .15, .45, or .85). As shown in Figure 6.1 through Figure 6.4 in the Appendix C, the bias and MSE values of the estimators are compared under the following four conditions depending on the values of the specific variances: 1) All six indicators have zero specific variances (i.e., SV(x_1 or y_1) = SV2(x_2 or y_2) = SV3(x_3 or y_3) = 0), 2) All six indicators have specific variances of .15 (i.e., SV(x_1 or y_1) = SV2(x_2 or y_2) = SV3(x_3 or y_3) = .15), 3) All six indicators have specific variances of .45 (i.e., SV(x_1 or y_1) = SV2(x_2 or y_2) = SV3(x_3 or y_3) = .45), and 4) All six indicators have specific variances of .85 (i.e., SV(x_1 or y_1) = SV2(x_2 or y_2) = SV3(x_3 or y_3) = .85), and 5) The three xs and three ys have

different specific variances (i.e., SV(x_1 or y_1) = .85, SV (x_2 or y_2) = .45, SV (x_3 or y_3) = .15).

Figure 6.1 and Figure 6.2 in the Appendix C show that when the specific variances of the six indicators are zero, the bias values of both estimators are the smallest regardless of the values of the index of the true relationship between two underlying constructs (γ). In fact, regardless of γ, the biases of estimators are nearly zero.

However, the effect on the MSEs of estimators of having nonzero specific variances in the indicators seems to be greater and more obvious. As shown in Figure 6.3 and Figure 6.4 in the Appendix C, the MSEs of estimators based on indicators with nonzero specific variances are bigger by .2 or more, compared to MSEs of estimator using indicators with zero specific variance. This pattern is shown regardless of the population values of the true relationship between two underlying constructs (γ). This indicates that the strength of the relationship is accurately estimated when combining correlations that are generated from indicators without specific variance introduced. Therefore, it should be fully understood that the true relationship between two constructs might be underestimated when combining correlations using indicators with specific variances.

CHAPTER 7

DISCUSSION

With the recent movement toward evidence-based policy and practice in education (Whitehurst, 2002), a growing interest has been devoted to meta-analytic techniques as a means of providing rigorous educational evidence (Slavin, 2008). As attractive and useful as these may seem in providing critical determinants for policy decisions, the application of existing methodology in research synthesis faces numerous difficulties and limitations due to the inherent nature of research in education and social science.

A particular problem in the social sciences and education is that studies employ diverse measures, even though researchers intend these to represent *the same* underlying constructs. Due to the use of diverse measures in primary studies, meta-analysts encounter three important problems. One is the variety of measures that are used to measure the same underlying construct. Another is the variety of statistical methods that are used to estimate the relationship between the two constructs. Third, there are not many pairs of the studies that use the same combinations of measures or statistical approaches, which leads to call as a *sparse* data structure. There may be many studies but no easy way to form a collection of studies that use the same measures or the same statistical methods to generate their estimates of effect sizes.

Therefore, I have proposed a new method for quantifying the strength of the relationship between two constructs that are measured in many different ways across studies. In this research, I have developed an approach that can handle the very sparse data structure that arises from variations in measures and statistical techniques. I also

want an approach that recognizes the fact that, even though a set of measurements can be quite different in their characteristics, they are all attempting to measure the same underlying construct.

One advantage of using the proposed approach given in Equation 3.13 is that it can estimate the strength of the relationship between two underlying constructs that are measured in different ways. By using the proposed method, variation in measures, which may lead to considerable heterogeneity across studies, is taken into account. Thus, the proposed method can provide more precise estimates by combining the corrected study effects. Contrary to Hunter and Schmidt's approach, the proposed method also suggests practical ways to adjust measure differences (i.e., how to obtain the validities of the indicators) that are derived from the population model. In addition,

In spite of the advantages of using the proposed method mentioned above, further research is required to resolve some practical issues. First of all, although the approach focuses on a simple bivariate relationship between two underlying constructs, each of which is measured using various indicators, it is certainly plausible to expand the proposed method for application to more complicated models. For instance, the partial correlation between two constructs controlling for a 3rd construct can be obtained from pooled correlations of control variables with both outcome and predictor variables.

As demonstrated in the simulation, the proposed approach correctly estimates the desired population parameters (i.e., the true relationship between two underlying constructs) if no missing *r*s exist. However, the bias and MSE values of the estimators get bigger as the number of missing *r*s is slightly increased. Since estimators differ depending on the number of missing *r*s, further investigations are needed to deal with the

missing *r*s. As discussed in the previous sections, one potential solution would be to impute missing information based on the available information. However, more attention should be paid to developing an approach that can deal with the missing data, which might be missing by "design".

One potential study would be to look into the robustness of the proposed model when the population model is not correctly specified. Knowing how robust the proposed approach is would definitely offer useful insights for applying the proposed approach in practice. Therefore, the generalizability of the proposed approach should be investigated under different scenarios. For example, the performance of the proposed method based on a one-factor model can be examined, although the true population model is a two-factor model.

In addition, the proposed approach has been examined for the case in which all indicators are assumed to be continuous variables. Future research could also examine how to combine relations involving both continuous and categorical variables.

APPENDIX A:

Protocol for Obtaining Expert Judgments

Expert Judgments on Measure Validity

: How teacher knowledge and student learning in math have been measured?

Many researchers have long been interested in the effect of how much teachers know in a subject they taught on improving student learning. However, studies and reviews have shown mixed findings for the relationship between teachers' subject matter knowledge and student learning (Ahn & Choi, 2005; Darling-Hammond, 2000; Wilson, Floden, & Ferrini-Mundy, 2001). This research posits that variation in measures to represent both teachers' subject matter knowledge and student learning might lead inconsistent findings. For instance, in literature, different indicators have been used to represent both teachers' subject matter knowledge and student learning in mathematics; some researchers use the indicator, number of teachers' courses in math as the best representation of teachers' knowledge (the construct) and others use teachers' test scores (as indicator) for measuring teachers' subject matter knowledge (the construct). Therefore, this assessment aims to obtain your judgments about how the indicators used in 8 primary studies represent their corresponding constructs (i.e., teachers' subject matter knowledge and student learning in arithmetic). These 8 studies focus on the relationship between teachers' knowledge and 6th grade students' test scores in arithmetic. Please read the following instruction and provide your judgments on the attached sheet.

[Instruction]

You as an independent rater are expected to work individually. First, you should be familiar with the instrument(s) used in each study. You want to look at the provided information regarding to instrument (e.g., sample item(s) of the instrument, the instrument, and method section in the study). If needed, you can also use any accessible resources (e.g., internet, test manual, other study using the same instrument). For more resources, please contact Soyeon Ahn (ahnso@msu.edu or 517-256-1891).

Second, you should evaluate how well each indicator represents each dimension of math knowledge in arithmetic.

Finally, based on your evaluation, you want to rank order measures used in studies for which you think would be most likely to measure each dimension of teachers' subject matter knowledge and students' understanding of arithmetic knowledge. Also, you want to provide the approximate value of validities of each indicator in terms of correlation value.

[Teachers' Subject Matter Knowledge Measures in Mathematics]

Definitions of Mathematical Knowledge in Arithmetic[5]

A. Understand numbers, ways of representing numbers, relationships among numbers, and number systems
B. Understand meanings of operations and how they relate to one another
C. Compute fluently and make reasonable estimates
D. Other: _____

ID	Measures	A					B					C					D				
		Well				*Poorly*	*Well*				*Poorly*	*Well*				*Poorly*	*Well*				*Poorly*
1	Glennon test score of basic mathematical understanding	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
2	Callahan Test of Mathematical Knowledge	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
3	Dr. Leroy's Test of Mathematical Understanding	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
4	Test of Teacher understanding	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5

1. Rate each measure in terms of the degree to which three components of mathematical knowledge in a domain are measured:

2. Please rank order the IDs of measures according to which you think would be most likely to measure teachers' subject matter knowledge:

_____ _____ _____ _____ _____ _____

3. Please provide the approximate correlation between what is measured in each assessment and how much teachers know in mathematics:

_____ _____ _____ _____ _____ _____

[5] Standards are obtained from NCTM numbers and operations standard for grades 6-8 available in http://standards.nctm.org/document/chapter6/numb.htm.

[Students' Knowledge Measures in Mathematics]

Definitions of Knowledge in Arithmetic[6]

A. Understand numbers, ways of representing numbers, relationships among numbers, and number systems (Conceptual Knowledge)
B. Understand meanings of operations and how they relate to one another
C. Compute fluently and make reasonable estimates
D. Other: _____

ID	Measures	A					B					C					D				
		Well				*poorly*	*Well*				*Poorly*	*Well*				*Poorly*	*Well*				*Poorly*
1	California Achievement Test (CAT)	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
2	Stanford Achievement Test (SAT)	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
3	Edwards	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
4	Iowa Test Basic Skill (ITBS)	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
5	Growth in mathematics	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
6	Achievement Series (SRA)	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5

1. Rate each measure in terms of the degree to which three components of mathematical knowledge in a domain are measured:

2. Please rank order the IDs of measures according to which you think would be most likely to measure teachers' subject matter knowledge: _____ _____ _____ _____ _____ _____

3. Please provide the approximate correlation between what is measured in each assessment and how much teachers know in mathematics:

_____ _____ _____ _____ _____ _____

[6] Standards are obtained from NCTM numbers and operations standard for grades 6-8 available in http://standards.nctm.org/document/chapter6/numb.htm.

APPEDIX B:

R Code for the Proposed Model

APPEDIX B:

R Code for the Proposed Model

```
# ------------------------------------------------------------------------------------------------------------.
# Author: Soyeon Ahn
# Date: 2008-02-01/ 2008-02-03: Fifth revision
# Simulation in detail: This is a simulation set-up for the dissertation. In this simulation, the k
independent studies with sample size of 30 are generated from the population model. In the
population model, the relationship between two underlying factors (ksi & eta), each of which is
measured using three indicators (three xs p=3 & three ys q=3), is of our interest. In the population,
data with 30 sample size are generated from the multivariate distribution with mean vector of 0 and
variance-covariance matrix, which is obtained from the population parameters. The zero-order
correlation coefficients between xs and ys are computed for meta-analysis.
# 1000 replications per conditions.
# 0 missing case - all the included studies provide all 9 possible zero-order correlation coefficients
between 3xs and 3ys.
# ------------------------------------------------------------------------------------------------------------.

# Set-up the directory.
setwd ("C:/Documents and Settings/Soyeon Ahn/Desktop/Dissertation/Simulation/sim_data/0
missing")
getwd()

library(MASS)

# ------------------------------------------------------------------------------------------------------------.
# Population parameters: Reliability of x, reliability of y, and gamma. For all simulation, it is
assumed that ksi and eta are standardized with mean of 0 and variance of 1. This indicates that
phi and psi are set to 1.
# Gamma is set to .5 & 0.
# Reliabilities of xs: [.9, .9, .9]; [.5, .5, .5]; [.9,.5,.2]; [.2,.2,.2].
# Reliabilities of ys: [.9, .9, .9]; [.5, .5, .5]; [.9,.5,.2]; [.2,.2,.2].
# Specific variances: 0 or (Reliability - .05).
# Factor loading: sqrt(Reliability-specific variance)
# Delta=specific variance+(1-factor loadidng^2).; Epsilon=specific variance+(1-factor loading^2).
# Assuming that we have correlation coefficients for diagonal, # of missingness will be manipulated
(0/6, 1/6, 2/6, 3/6, 4/6, 5/6, 6/6). Choice of what should be unavailable will be based on random
selection.
# ------------------------------------------------------------------------------------------------------------.

# Detailed procedures.
# 1. make var-cov; 2. generate multivariate normal distribution using mean vector & var-cov
produced in 1; 3. generate correlation coefficients; 4. generate random number for choosing
missingness; 5. get the estimator.
```

```
MV<-matrix(c(0), nrow=1, ncol=6, byrow=TRUE) #MV is mean-vector of variables (3xs +3ys)

# Create hypothetical meta-analysis for each condition.
# Function called "hypothetical meta" creates 1. variance-covariance matrix for generating zero-order correlations for each study in a hypothetical meta-analysis.

hypothetical.meta<-function(GAMMA, V, Rel1,Rel2, Rel3, study){
# For creating variance-covariance matrix, we need the following information - Gamma, Phi, Psi, Lambda_X (from reliability of Xs), Lambda_Y (from reliability of Ys), Theta_delta, and Theta_epsilon.
        GA<-matrix(c(GAMMA), c(1,1))
                phi<-matrix(1,c(1,1))
                psi<-matrix(1,c(1,1))
                rel<-matrix(c(Rel1, Rel2, Rel3), c(3,1))
# Here, one additional condition is added for this simulation. If V=1, no specific variances exist on any side of exogenous and endogenous variables.
                        if (V==1) {sv <-matrix(c(0,0,0), c(3,1))} else {sv<-matrix(c(Rel1-.05,Rel2-.05,Rel3-.05), c(3,1))}
                LX<-sqrt(rel-sv)
                LY<-sqrt(rel-sv)
                TD<-matrix(c((1-LX[1,1]^2),0,0,0,(1-LX[2,1]^2),0,0,0, (1-LX[3,1]^2)),nrow=3, ncol=3,byrow=TRUE)
                TE<-matrix(c((1-LY[1,1]^2),0,0,0,(1-LY[2,1]^2),0,0,0, (1-LY[3,1]^2)),nrow=3, ncol=3,byrow=TRUE)
# Now, using seven parameters above, variance-covariance matrix (called sigma_XX, sigma_YY, sigma_XY) will be established.
# Sigma_XX= LX*phi*LX'+TD; Sigma_YY=LY*psi*LY'+TE; Sigma_XY=LX*GA*LY' (Use these equastions)
                Sigma_XX<-LX %*% phi %*% t(LX)+TD
                Sigma_YY<-LY %*% psi %*% t(LY)+TE
                Sigma_XY<-LX%*%GA%*%t(LY)
                Sigma_YX<-LY%*%GA%*%t(LX)
                Sigma<-rbind(cbind(Sigma_XX,Sigma_XY),cbind(Sigma_YX, Sigma_YY))

# Second, I'm generating multivariate normal distribution and creating correlation coefficients among all indicators.

#       x1   x2 x3
# y1 [1,1] [2,1] [3,1]
# y2 [1,2] [2,2] [3,2]
# y3 [1,3] [2,3] [3,3]

# For missing 4 cases.
es.meta<-matrix(0,study*9,4)
ES<-matrix(0,1,17)

 # change the 1: (1,6,15, 10, 6, 3, 1)
```

```
for (j in 1:1){
for (i in 1:study){
cor.data<-mvrnorm(30, MV, Sigma, empirical=FALSE)
a<-cor(cor.data)
cmatrix<-cor(cor.data)[1:3,4:6]
attributes(cmatrix)

# ----------------------------------------------------------------------------------------------------------------.
# For missing information, indicate elements in the correlation matrix that is generated from
multivariate normal distribution called "cor.data". It is named as C1-C15.
C1<-cbind(1,2,cmatrix[1,2])
C2<-cbind(1,3,cmatrix[1,3])
C3<-cbind(2,3,cmatrix[2,3])
C4<-cbind(2,1,cmatrix[2,1])
C5<-cbind(3,1,cmatrix[3,1])
C6<-cbind(3,2,cmatrix[3,2])
C7<-cbind(1,1,cmatrix[1,1]) #C7 is on diagonal.
C8<-cbind(2,2,cmatrix[2,2])  #C8 is on diagonal.
C9<-cbind(3,3,cmatrix[2,2])  #C9 is on diagonal.

# By having random #, we can assgin the same # of Rs across all three elements on the diagonal.
#random<-runif(1,0,1)
# ind<-ifelse(random<1/3,1,ifelse(random>=1/3 & random<2/3,2,3))

# 0 missing
es.meta[9*i-8,]<-cbind(i, C7)  #C1, C2, C3, C4, C5, C6, C7, C8, C9
es.meta[9*i-7,]<-cbind(i, C8)
es.meta[9*i-6,]<-cbind(i, C9)
es.meta[9*i-5,]<-cbind(i, C1)
es.meta[9*i-4,]<-cbind(i, C2)
es.meta[9*i-3,]<-cbind(i, C3)
es.meta[9*i-2,]<-cbind(i, C4)
es.meta[9*i-1,]<-cbind(i, C5)
es.meta[9*i,]<-cbind(i, C6)}
# ----------------------------------------------------------------------------------------------------------------.

# ----------------------------------------------------------------------------------------------------------------.
# 1 missing case
# For missing information, indicate elements in the correlation matrix that is generated from
multivariate normal distribution called "cor.data". It is named as C1-C15.
C1<-cbind(1,2,cmatrix[1,2])
C2<-cbind(1,3,cmatrix[1,3])
C3<-cbind(2,3,cmatrix[2,3])
C4<-cbind(2,1,cmatrix[2,1])
C5<-cbind(3,1,cmatrix[3,1])
C6<-cbind(3,2,cmatrix[3,2])
C7<-cbind(1,1,cmatrix[1,1]) #C7 is on diagonal.
```

```
C8<-cbind(2,2,cmatrix[2,2])  #C8 is on diagonal.
C9<-cbind(3,3,cmatrix[2,2])  #C9 is on diagonal.
all.element1<-rbind(C2, C1,C1,C1, C1, C1)
all.element2<-rbind(C3, C3,C2,C2, C2, C2)
all.element3<-rbind(C4, C4,C4,C3, C3, C3)
all.element4<-rbind(C5, C5,C5,C5, C4, C4)
all.element5<-rbind(C6, C6,C6,C6, C6, C5)


#(2,3,4,5,6), (1,3,4,5,6), (1,2,4,5,6), (1,2,3,5,6), (1,2,3,4,6), (1,2,3,4,5).
# By having random #, we can assgin the same # of Rs across all three elements on the diagonal.
#random<-runif(1,0,1)
# ind<-ifelse(random<1/3,1,ifelse(random>=1/3 & random<2/3,2,3))

es.meta[8*i-7,]<-cbind(i, C7)  #C1, C2, C3, C4, C5, C6, C7, C8, C9
es.meta[8*i-6,]<-cbind(i, C8)
es.meta[8*i-5,]<-cbind(i, C9)
 es.meta[8*i-4,]<-cbind(i, matrix(all.element1[j,],c(1:3)))
 es.meta[8*i-3,]<-cbind(i, matrix(all.element2[j,],c(1:3)))
 es.meta[8*i-2,]<-cbind(i, matrix(all.element3[j,],c(1:3)))
 es.meta[8*i-1,]<-cbind(i, matrix(all.element4[j,],c(1:3)))
 es.meta[8*i,]<-cbind(i, matrix(all.element5[j,],c(1:3)))}
# ----------------------------------------------------------------------------------------------------------.

# ----------------------------------------------------------------------------------------------------------.
# 2 missing case
# For missing information, indicate elements in the correlation matrix that is generated from
multivariate normal distribution called "cor.data". It is named as C1-C15.
C1<-cbind(1,2,cmatrix[1,2])
C2<-cbind(1,3,cmatrix[1,3])
C3<-cbind(2,3,cmatrix[2,3])
C4<-cbind(2,1,cmatrix[2,1])
C5<-cbind(3,1,cmatrix[3,1])
C6<-cbind(3,2,cmatrix[3,2])
C7<-cbind(1,1,cmatrix[1,1]) #C7 is on diagonal.
C8<-cbind(2,2,cmatrix[2,2])  #C8 is on diagonal.
C9<-cbind(3,3,cmatrix[2,2])  #C9 is on diagonal.
all.element1<-rbind(C3,C2,C2,C2,C2,C1,C1,C1,C1,C1,C1,C1,C1,C1,C1)
all.element2<-rbind(C4,C4,C3,C3,C3,C4,C3,C3,C3,C2,C2,C2,C2,C2,C2)
all.element3<-rbind(C5,C5,C5,C4,C4,C5,C5,C4,C4,C5,C4,C4,C3,C3,C3)
all.element4<-rbind(C6,C6,C6,C6,C5,C6,C6,C6,C5,C6,C6,C5,C6,C5,C4)
# (3,4,5,6), (2,4,5,6), (2,3,5,6), (2,3,4,6),
# (2,3,4,5), (1,4,5,6), (1,3,5,6), (1,3,4,6), (1,3,4,5),
#(1,2,5,6), (1,2,4,6), (1,2,4,5), (1,2,3,6),
# (1,2,3,5), (1,2,3,4).

# 2 missing
```

```
es.meta[7*i-6,]<-cbind(i, C7)  #C1, C2, C3, C4, C5, C6, C7, C8, C9
es.meta[7*i-5,]<-cbind(i, C8)
es.meta[7*i-4,]<-cbind(i, C9)
 es.meta[7*i-3,]<-cbind(i, matrix(all.element1[j,],c(1:3)))
 es.meta[7*i-2,]<-cbind(i, matrix(all.element2[j,],c(1:3)))
 es.meta[7*i-1,]<-cbind(i, matrix(all.element3[j,],c(1:3)))
  es.meta[7*i,]<-cbind(i, matrix(all.element4[j,],c(1:3)))}

# (1,2,3,4), (1,2,3,5), (1,2,3,6), (2,3,4,5), (2,3,4,6), (3,4,5,6)
# ------------------------------------------------------------------------------------------------------------.

# ------------------------------------------------------------------------------------------------------------.
# 3 missing case
# For missing information, indicate elements in the correlation matrix that is generated from
multivariate normal distribution called "cor.data". It is named as C1-C15.
C1<-cbind(1,2,cmatrix[1,2])
C2<-cbind(1,3,cmatrix[1,3])
C3<-cbind(2,3,cmatrix[2,3])
C4<-cbind(2,1,cmatrix[2,1])
C5<-cbind(3,1,cmatrix[3,1])
C6<-cbind(3,2,cmatrix[3,2])
C7<-cbind(1,1,cmatrix[1,1]) #C7 is on diagonal.
C8<-cbind(2,2,cmatrix[2,2])  #C8 is on diagonal.
C9<-cbind(3,3,cmatrix[2,2])  #C9 is on diagonal.

all.element1<-rbind(C1,C1,C1,C1,C2,C2,C2,C3,C3,C4)
all.element2<-rbind(C2,C2,C2,C2,C3,C3,C3,C4,C4,C5)
all.element3<-rbind(C3,C4,C5,C6,C4,C5,C6,C5,C6,C6)
# (1,2,3), (1, 2,4), (1, 2,5), (1, 2, 6), (2,3,4), (2,3,5), (2,3,6), (3, 4, 5), (3,4,6), (4, 5, 6)
# ------------------------------------------------------------------------------------------------------------.

# ------------------------------------------------------------------------------------------------------------.
# 4 missing case

# For missing information, indicate elements in the correlation matrix that is generated from
multivariate normal distribution called "cor.data". It is named as C1-C15.
C1<-cbind(1,2,cmatrix[1,2])
C2<-cbind(1,3,cmatrix[1,3])
C3<-cbind(2,3,cmatrix[2,3])
C4<-cbind(2,1,cmatrix[2,1])
C5<-cbind(3,1,cmatrix[3,1])
C6<-cbind(3,2,cmatrix[3,2])
C7<-cbind(1,1,cmatrix[1,1]) #C7 is on diagonal.
C8<-cbind(2,2,cmatrix[2,2])  #C8 is on diagonal.
C9<-cbind(3,3,cmatrix[2,2])  #C9 is on diagonal.
all.element1<-rbind(C1,C1,C1,C1,C1,C1,C2,C2,C2,C2, C3,C3,C3,C4,C4,C5)
all.element2<-rbind(C2,C3,C4,C5,C6,C3,C4,C5,C6,C4, C5,C6,C5,C6,C6,C6)
```

```
# By having random #, we can assgin the same # of Rs across all three elements on the diagonal.
#random<-runif(1,0,1)
 #ind<-ifelse(random<1/3,1,ifelse(random>=1/3 & random<2/3,2,3))

 # 4 missing
 es.meta[5*i-4,]<-cbind(i, C7)  #C1, C2, C3, C4, C5, C6, C7, C8, C9
es.meta[5*i-3,]<-cbind(i, C8)
es.meta[5*i-2,]<-cbind(i, C9)
 es.meta[5*i-1,]<-cbind(i, matrix(all.element1[j,],c(1:3)))
 es.meta[5*i,]<-cbind(i, matrix(all.element2[j,],c(1:3)))}

# (C1, C2), (C1, C3), (C1, C4), (C1, C5), (C1, C6), (C2, C3), (C2, C4), (C2, C5), (C2, C6), (C3,
C4),(C3, C5), (C3, C6), (C4, C5), (C4, C6), (C5, C6).
# ----------------------------------------------------------------------------------------------------.

# ----------------------------------------------------------------------------------------------------.
# 5 missing case
# For missing information, indicate elements in the correlation matrix that is generated from
multivariate normal distribution called "cor.data". It is named as C1-C6.
C1<-cbind(1,2,cmatrix[1,2])
C2<-cbind(1,3,cmatrix[1,3])
C3<-cbind(2,3,cmatrix[2,3])
C4<-cbind(2,1,cmatrix[2,1])
C5<-cbind(3,1,cmatrix[3,1])
C6<-cbind(3,2,cmatrix[3,2])
C7<-cbind(1,1,cmatrix[1,1]) #C7 is on diagonal.
C8<-cbind(2,2,cmatrix[2,2])  #C8 is on diagonal.
C9<-cbind(3,3,cmatrix[2,2])  #C9 is on diagonal.
all.element<-rbind(C1,C2,C3,C4,C5,C6)

# By having random #, we can assgin the same # of Rs across all three elements on the diagonal.
#random<-runif(1,0,1)
 #ind<-ifelse(random<1/3,1,ifelse(random>=1/3 & random<2/3,2,3))

# 5 missing
 es.meta[4*i-3,]<-cbind(i, C7)  #C1, C2, C3, C4, C5, C6, C7, C8, C9
es.meta[4*i-2,]<-cbind(i, C8)
es.meta[4*i-1,]<-cbind(i, C9)
es.meta[4*i,]<-cbind(i, matrix(all.element[j,],c(1:3)))}
# ----------------------------------------------------------------------------------------------------.

# ----------------------------------------------------------------------------------------------------.
# 6 missing
C7<-cbind(1,1,cmatrix[1,1]) #C7 is on diagonal.
C8<-cbind(2,2,cmatrix[2,2])  #C8 is on diagonal.
C9<-cbind(3,3,cmatrix[2,2])  #C9 is on diagonal.
```

```
# By having random #, we can assgin the same # of Rs across all three elements on the diagonal.
#random<-runif(1,0,1)
# ind<-ifelse(random<1/3,1,ifelse(random>=1/3 & random<2/3,2,3))

es.meta[3*i-2,]<-cbind(i, C7)  #C1, C2, C3, C4, C5, C6, C7, C8, C9
es.meta[3*i-1,]<-cbind(i, C8)
es.meta[3*i,]<-cbind(i, C9)}
attributes(es.meta)
# After creating a hypothetical studies with # of studies in it, next step is to compute the final
estimates based on sample-size weighted average Rs & z-transformed variance weighted Rs.

# ----------.

# ES1 is the final ES based on sample-size weighted Rs & ES2 is the final ES based on z-
transformed weighted Rs.
meta<-cbind(es.meta, es.meta[,4]*30, 27*(.5*log((1+es.meta[,4])/(1-es.meta[,4]))))
ES1.sum<-data.matrix(aggregate(meta[,5], list(x=meta[,2], y=meta[,3]),sum))
ES1<-cbind(ES1.sum[,1], ES1.sum[,2],ES1.sum[,3]/(30*study))
ES2.sum<-data.matrix(aggregate(meta[,6], list(x=meta[,2], y=meta[,3]), sum))
ES2<-cbind(ES2.sum[,1], ES2.sum[,2],ES2.sum[,3]/(27*study))
ES2<-cbind(ES2.sum[,1], ES2.sum[,2],(exp(2*ES2[,3])-1)/(exp(2*ES2[,3])+1))

meta_ES1<-as.matrix((apply(ES1,2,sum))/((apply(LX,2,sum))*(apply(LY,2,sum))))
meta_ES2<-as.matrix((apply(ES2,2,sum))/((apply(LX,2,sum))*(apply(LY,2,sum))))

# ES includes both the final ES based on sample-size weighted Rs(ES1) & the final ES based on
z-transforme weighted Rs.
ES[j,]<-cbind(j, LX[1,1], LX[2,1], LX[3,1], LY[1,1], LY[2,1], LY[3,1], sv[1,1], sv[2,1], sv[3,1], study,
GA, Rel1, Rel2, Rel3, meta_ES1[3,1], meta_ES2[3,1])}
result<-return(ES)}

# Make hypothetical meta-analyses depending on different conditions in accordance with different
parameters.

# ----------.
# Definine function depending for getting final ES.
# ----------

matrix.ga<-matrix(c(.5,0), c(1,2))
matrix.specific.variance<-matrix(c(1,0), nrow=1, ncol=2,byrow=TRUE)
matrix.reliability<-matrix(c(.9,.9,.9,.5,.5,.5,.2,.2,.2,.9,.5,.2), nrow=4, ncol=3, byrow=TRUE)
n.study<-matrix(c(9, 36), ncol=1, nrow=2, byrow=FALSE) # k is # of studies included in meta-
analysis;

# 32 conditions by 2 (Gamma) * 4(Reliability sets) * 2(# of study) * 2(# of specific variance) =32
#(0.5,0,.9,.9,.9,9) (0.5,1,.9,.9,.9,9) (0,0,.9,.9,.9,9) (0,1,.9,.9,.9,9)
```

```
#(0.5,0,.9,.9,.9,36) (0.5,1,.9,.9,.9,36) (0,0,.9,.9,.9,36) (0,1,.9,.9,.9,36)
#(0.5,0,.5,.5,.5,9) (0.5,1,.5,.5,.5,9) (0,0,.5,.5,.5,9) (0,1,.5,.5,.5,9)
#(0.5,0,.5,.5,.5,36) (0.5,1,.5,.5,.5,36) (0,0,.5,.5,.5,36) (0,1,.5,.5,.5,36)
#(0.5,0,.2,.2,.2,9) (0.5,1,.2,.2,.2,9) (0,0,.2,.2,.2,9) (0,1,.2,.2,.2,9)
#(0.5,0,.2,.2,.2,36) (0.5,1,.2,.2,.2,36) (0,0,.2,.2,.2,36) (0,1,.2,.2,.2,36)
#(0.5,0,.9,.5,.2,9) (0.5,1,.9,.5,.2,9) (0,0,.9,.5,.2,9) (0,1,.9,.5,.2,9)
#(0.5,0,.9,.5,.2,36) (0.5,1,.9,.5,.2,36) (0,0,.9,.5,.2,36) (0,1,.9,.5,.2,36)

replication<-1000 #write # of replications here
M1<-lapply(1:replication, function(x) hypothetical.meta(0.5,0,.9,.9,.9,9))
M2<-lapply(1:replication, function(x) hypothetical.meta(0.5,1,.9,.9,.9,9))
M3<-lapply(1:replication, function(x) hypothetical.meta(0,0,.9,.9,.9,9) )
M4<-lapply(1:replication, function(x) hypothetical.meta(0,1,.9,.9,.9,9))
M5<-lapply(1:replication, function(x) hypothetical.meta(0.5,0,.9,.9,.9,36))
M6<-lapply(1:replication, function(x) hypothetical.meta(0.5,1,.9,.9,.9,36))
M7<-lapply(1:replication, function(x) hypothetical.meta(0,0,.9,.9,.9,36))
M8<-lapply(1:replication, function(x) hypothetical.meta(0,1,.9,.9,.9,36))
M9<-lapply(1:replication, function(x) hypothetical.meta(0.5,0,.5,.5,.5,9))
M10<-lapply(1:replication, function(x) hypothetical.meta(0.5,1,.5,.5,.5,9))
M11<-lapply(1:replication, function(x) hypothetical.meta(0,0,.5,.5,.5,9))
M12<-lapply(1:replication, function(x) hypothetical.meta(0,1,.5,.5,.5,9))
M13<-lapply(1:replication, function(x) hypothetical.meta(0.5,0,.5,.5,.5,36))
M14<-lapply(1:replication, function(x) hypothetical.meta(0.5,1,.5,.5,.5,36))
M15<-lapply(1:replication, function(x) hypothetical.meta(0,0,.5,.5,.5,36))
M16<-lapply(1:replication, function(x) hypothetical.meta(0,1,.5,.5,.5,36))
M17<-lapply(1:replication, function(x) hypothetical.meta(0.5,0,.2,.2,.2,9))
M18<-lapply(1:replication, function(x) hypothetical.meta(0.5,1,.2,.2,.2,9))
M19<-lapply(1:replication, function(x) hypothetical.meta(0,0,.2,.2,.2,9))
M20<-lapply(1:replication, function(x) hypothetical.meta(0,1,.2,.2,.2,9))
M21<-lapply(1:replication, function(x) hypothetical.meta(0.5,0,.2,.2,.2,36))
M22<-lapply(1:replication, function(x) hypothetical.meta(0.5,1,.2,.2,.2,36))
M23<-lapply(1:replication, function(x) hypothetical.meta(0,0,.2,.2,.2,36))
M24<-lapply(1:replication, function(x) hypothetical.meta(0,1,.2,.2,.2,36))
M25<-lapply(1:replication, function(x) hypothetical.meta(0.5,0,.9,.5,.2,9))
M26<-lapply(1:replication, function(x) hypothetical.meta(0.5,1,.9,.5,.2,9))
M27<-lapply(1:replication, function(x) hypothetical.meta(0,0,.9,.5,.2,9))
M28<-lapply(1:replication, function(x) hypothetical.meta(0,1,.9,.5,.2,9))
M29<-lapply(1:replication, function(x) hypothetical.meta(0.5,0,.9,.5,.2,36))
M30<-lapply(1:replication, function(x) hypothetical.meta(0.5,1,.9,.5,.2,36))
M31<-lapply(1:replication, function(x) hypothetical.meta(0,0,.9,.5,.2,36))
M32<-lapply(1:replication, function(x) hypothetical.meta(0,1,.9,.5,.2,36))
# ------------------------------------------------------------------------------------------------------------------------.
# Creating dataset by combining
# ------------------------------------------------------------------------------------------------------------------------.

#class(META1[[1]])
#replication
```

```
#length(META1)
# replication == length(META1)

library(gdata)

create.data<-function(META1,META2){
META2<-matrix(0,1000,17)
for (i in 1:replication){
META2[i,]<-META1[[i]]}
colnames(META2)<-c( "Ind", "LX1", "LX2", "LX3", "LY1", "LY2", "LY3", "sv1", "sv2", "sv3", "study",
"GA", "Rel1", "Rel2", "Rel3", "ES1", "ES2")
result<-return(META2)}

R1<-create.data(M1,D1)
R2<-create.data(M2,D2)
R3<-create.data(M3,D3)
R4<-create.data(M4,D4)
R5<-create.data(M5,D5)
R6<-create.data(M6,D6)
R7<-create.data(M7,D7)
R8<-create.data(M8,D8)
R9<-create.data(M9,D9)
R10<-create.data(M10,D10)
R11<-create.data(M11,D11)
R12<-create.data(M12,D12)
R13<-create.data(M13,D13)
R14<-create.data(M14,D14)
R15<-create.data(M15,D15)
R16<-create.data(M16,D16)
R17<-create.data(M17,D17)
R18<-create.data(M18,D18)
R19<-create.data(M19,D19)
R20<-create.data(M20,D20)
R21<-create.data(M21,D21)
R22<-create.data(M22,D22)
R23<-create.data(M23,D23)
R24<-create.data(M24,D24)
R25<-create.data(M25,D25)
R26<-create.data(M26,D26)
R27<-create.data(M27,D27)
R28<-create.data(M28,D28)
R29<-create.data(M29,D29)
R30<-create.data(M30,D30)
R31<-create.data(M31,D31)
R32<-create.data(M32,D32)
```

```
big.data<-combine(R1, R2, R3, R4, R5, R6, R7, R8, R9, R10, R11, R12, R13, R14, R15, R16, R17,
R18, R19, R20, R21, R22, R23, R24, R25, R26, R27, R28, R29, R30, R31, R32)
write.matrix(big.data, file="D_0 missing.xls", sep=" ")
```

APPENDIX C:

Tables and Figures

Table 2.1

Attenuated Artifacts and the Corresponding Multiplier

Attenuation Artifacts	**The corresponding multiplier**
Random error of measurement in dependent variable Y	$a_1 = \sqrt{r_{yy}}$, r_{yy} is the reliability of the Y measure
Random error of measurement in independent variable X	$a_2 = \sqrt{r_{xx}}$, r_{xx} is the reliability of the X measure
Artificial dichotomization of continuous dependent variable split into proportions p and q	a_3 = biserial constant = $\phi(c) / \sqrt{pq}$ where $\phi(x) = e^{-x^2} / \sqrt{2\pi}$ is the unit normal density function and c is unit normal distribution cut point corresponding to a split of p
Artificial dichotomization of continuous independent variable split into proportions p and q	a_4 = biserial constant = $\phi(c) / \sqrt{pq}$ where $\phi(x) = e^{-x^2} / \sqrt{2\pi}$ is the unit normal density function and c is unit normal distribution cut point corresponding to a split of p
Imperfect construct validity of the dependent variable Y	a_5 = the construct validity of Y
Imperfect construct variable of the independent variable X	a_6 = the construct validity of X
Range restriction on the dependent variable Y	$a_7 = \sqrt{(u_y{}^2 + \rho^2 - u_y{}^2\rho^2)}$, where u = (SD_y study population)/ (SD_y reference population)
Range restriction on the independent variable X	$a_8 = \sqrt{(u_x{}^2 + \rho^2 - u_x{}^2\rho^2)}$, where u = (SD_x study population)/ (SD_x reference population)
Bias in the correlation coefficients due to small sample sizes	$a_9 = 1 - (1 - \rho^2) / (2N - 2)$
Study-caused variation	Partial correlation to remove the effects of unwanted variation in experience

Table 2.2.

Comparisons of Correction formulas

ES / Study / Correcting factors	*d*	Correlation				Regression
	Nugent (2006)	Le (2003); Sackett & Yang (2000)	Hunter & Schmidt (1990)	Hancock (1997)	Oswald & Converse (2005)	Raju & Brand (2003); Raju et al.(1991)
Reliability (X)	x		x	x	x	x
Reliability (Y)			x	x	x	
Validity (X)			x			
Validity (Y)			x			
Range restriction		x	x		x	
Sampling error			x	x		

Table 4.1.

Bias and MSE of Estimators

γ	0				.5			
Quality of	Bias		MSE		Bias		MSE	
Estimators	ES1	ES2	ES1	ES2	ES1	ES2	ES1	ES2
Min	-.01	-.01	0	0	-.02	0	0	0
Max	.01	.01	.06	.06	.15	.17	.05	.06
M	.0001	.0001	.0082	.0082	.008	.023	.008	.009
S.D	.003	.003	.01	.01	.03	.03	.01	.01

Table 4.2

Bias and MSE of Sample-size weighted and Z-transformed Variance Weighted Estimators Under Different Conditions ($\gamma = 0$)

Reliability (X1,Y1)	Reliability (X2,Y2)	Reliability (X3,Y3)	Factor Loading (X1,Y1)	Factor Loading (X2,Y2)	Factor Loading (X3,Y3)	k	# of Missing	ES1		ES2	
								Bias	MSE	Bias	MSE
0.9	0.9	0.9	0.95	0.95	0.95	9	0	.003	.004	.003	.004
							1	.000	.004	.000	.004
							2	.001	.004	.001	.004
							3	.000	.004	.000	.004
							4	.000	.004	.000	.004
							5	-.003	.004	-.003	.004
							6	-.002	.004	-.002	.005
						36	0	.000	.001	.000	.001
							1	.001	.001	.001	.001
							2	.000	.001	.000	.001
							3	.000	.001	.000	.001
							4	.000	.001	.000	.001
							5	.000	.001	.000	.001
							6	-.002	.001	-.002	.001
.5	.5	.5	.71	.71	.71	9	0	-.003	.007	-.003	.007
							1	.000	.007	.000	.008
							2	.000	.007	.000	.008
							3	-.001	.008	-.001	.008
							4	.000	.008	.000	.009
							5	.003	.009	.003	.010
							6	-.006	.010	-.006	.011
						36	0	.000	.002	.000	.002

Table 4.2 Continued

							1	.000	.002	.000	.002
							2	.000	.002	.000	.002
							3	.000	.002	.000	.002
							4	.000	.002	.000	.002
							5	.000	.002	.000	.002
							6	.002	.003	.002	.003
.2	.2	.2	.45	.45	.45	9	0	.003	.022	.004	.024
							1	.002	.025	.002	.026
							2	-.001	.027	-.001	.029
							3	.003	.030	.004	.032
							4	.000	.035	.000	.037
							5	.001	.043	.001	.046
							6	.006	.055	.007	.059
						36	0	-.002	.005	-.002	.006
							1	.000	.006	.000	.006
							2	.000	.007	.000	.007
							3	.000	.008	.000	.008
							4	.001	.009	.002	.009
							5	.000	.011	.000	.011
							6	.000	.014	-.001	.015
.9	.5	.2	.95	.71	.45	9	0	.004	.007	.004	.007
							1	.000	.007	.000	.008
							2	.000	.008	.000	.008
							3	-.001	.008	-.001	.009
							4	.000	.009	.000	.010
							5	.000	.010	.000	.011
							6	.005	.012	.005	.013
						36	0	.000	.002	.000	.002
							1	.000	.002	.000	.002

Table 4.2 Continued

							2	.000	.002	.001	.002
							3	.000	.002	.000	.002
							4	.000	.002	.000	.002
							5	-.001	.003	-.001	.003
							6	.001	.003	.001	.003

Table 4.3

Bias and MSE of Sample-size weighted and Z-transformed Variance Weighted Estimators Under Different Conditions ($\gamma = .5$)

Reliability (X1,Y1)	Reliability (X2,Y2)	Reliability (X3,Y3)	Factor Loading (X1,Y1)	Factor Loading (X2,Y2)	Factor Loading (X3,Y3)	*k*	# of Missing	ES1		ES2	
								Bias	MSE	Bias	MSE
.9	.9	.9	.95	.95	.95	9	0	-.006	.003	.006	.003
							1	-.007	.003	.005	.003
							2	-.007	.003	.006	.003
							3	-.007	.003	.006	.003
							4	-.008	.003	.005	.003
							5	-.007	.003	.006	.003
							6	-.009	.003	.004	.003
						36	0	-.006	.001	.008	.001
							1	-.007	.001	.006	.001
							2	-.007	.001	.007	.001
							3	-.007	.001	.006	.001
							4	-.007	.001	.007	.001
							5	-.007	.001	.007	.001
							6	-.008	.001	.006	.001
.5	.5	.5	.71	.71	.71	9	0	-.006	.006	.008	.007
							1	-.008	.006	.006	.007
							2	-.008	.006	.007	.007
							3	-.009	.007	.005	.007
							4	-.008	.007	.007	.008
							5	-.007	.008	.007	.008
							6	-.007	.009	.008	.010
						36	0	-.008	.002	.008	.002
							1	-.008	.002	.008	.002

Table 4.3 Continued

							2	-.008	.002	.008	.002
							3	-.008	.002	.008	.002
							4	-.008	.002	.008	.002
							5	-.008	.002	.008	.002
							6	-.007	.002	.009	.002
.2	.2	.2	.45	.45	.45	9	0	-.007	.022	.008	.023
							1	-.010	.024	.005	.025
							2	-.009	.026	.006	.028
							3	-.009	.030	.007	.032
							4	-.009	.034	.006	.036
							5	-.013	.041	.002	.043
							6	-.006	.053	.009	.057
						36	0	-.011	.006	.006	.006
							1	-.010	.006	.007	.006
							2	-.008	.007	.009	.007
							3	-.008	.008	.009	.008
							4	-.008	.009	.009	.009
							5	-.008	.011	.009	.011
							6	-.013	.014	.004	.015
.9	.5	.2	.95	.71	.45	9	0	.025	.007	.040	.008
							1	.032	.007	.047	.009
							2	.041	.009	.056	.011
							3	.048	.010	.064	.012
							4	.073	.014	.089	.017
							5	.093	.018	.109	.022
							6	.140	.029	.158	.035
						36	0	.027	.002	.043	.004
							1	.033	.003	.050	.005
							2	.042	.004	.059	.006

Table 4.3 Continued

							3	.049	.005	.067	.007
							4	.073	.008	.090	.011
							5	.094	.012	.112	.016
							6	.137	.021	.156	.027

Table 4.4

Mean bias of r from Field (2001)

	Number of studies	Homogeneous case		Heterogeneous case	
		H/O	H/S	H/O	H/S
$\rho = 0$	10	-.0001	-.0001	0	0
	30	-.00005	-.00005	0	0
$\rho = .5$	10	.006	-.007	.1785	-.024
	30	.007	.007	.205	-.024

Note.

H/O is the method suggested by Hedges and Olkin (1985) or Rosenthal and Rubin (1991). This is equivalent to ES2 in this dissertation; H/S is the method suggested by Hunter and Schmidt (1990). This is equivalent to ES1 in this dissertation

Table 4.5

Results from Multivariate Analysis of Variance (MANOVA) on the Bias of Estimators for $\gamma = 0$

Effect		Value	F	df_1	df_2	p	η^2
Number of Missing *r*s	Pillai's Trace	.12	442.06	12	863978	<.01	.058
	Wilks' Lambda	.89	4459.44	12	863976	<.01	.058
	Hotelling's Trace	.12	4498.84	12	863974	<.01	.059
	Roy's Largest Root	.10	6954.08	6	431989	<.01	.088
Reliability (Factor loading) of 6 variables	Pillai's Trace	1.69	779286.84	6	863978	<.01	.844
	Wilks' Lambda	.00	1894142.01	6	863976	<.01	.929
	Hotelling's Trace	6.49	4355172.78	6	863974	<.01	.968
	Roy's Largest Root	58.10	8366262.43	3	431989	<.01	.983
Number of study included (*k*)	Pillai's Trace	.96	4984185.66	2	431988	<.01	.958
	Wilks' Lambda	.04	4984185.66	2	431988	<.01	.958
	Hotelling's Trace	23.08	4984185.66	2	431988	<.01	.958
	Roy's Largest Root	23.08	4984185.66	2	431988	<.01	.958

Table 4.6

Tests of Between-factors Effects for Bias of Overall Effect-size Estimate ($\gamma = 0$)

Source	Estimators	*SS*	*df*	*MS*	*F*	*p*	η^2
Number of Missing *r*s	Bias_ES1	.01	6	.001	123.8	<.01	.002
	Bias_ES2	.01	6	.001	122.4	<.01	.002
Reliability (Factor loading)	Bias_ES1	.04	3	.012	1277.7	<.01	.009
of 6 variables	Bias_ES2	.04	3	.013	1293.2	<.01	.009
Number of study included (*k*)	Bias_ES1	.00	1	.001	69.5	<.01	0
	Bias_ES2	.00	1	.001	75.3	<.01	0
	Bias_ES1	4.02	431989	0			
Error	Bias_ES2	4.30	431989	0			
	Bias_ES1	4.07	432000				
Total	Bias_ES2	4.35	432000				

Table 4.7

Results from Multivariate Analysis of Variance (MANOVA) on the MSEs of Estimators for $\gamma = 0$

Effect		Value	F	df_1	df_2	p	η^2
Number of Missing *r*s	Pillai's Trace	.13	4874.36	12	863978	<.01	.063
	Wilks' Lambda	.87	5045.30	12	863976	<.01	.065
	Hotelling's Trace	.14	5216.62	12	863974	<.01	.068
	Roy's Largest Root	.14	10357.59	6	431989	<.01	.126
Reliability (Factor loading) of 6 variables	Pillai's Trace	.77	89819.13	6	863978	<.01	.384
	Wilks' Lambda	.23	154959.09	6	863976	<.01	.518
	Hotelling's Trace	3.31	238246.58	6	863974	<.01	.623
	Roy's Largest Root	3.31	47644.73	3	431989	<.01	.768
Number of study included (*k*)	Pillai's Trace	.54	250327.34	2	431988	<.01	.537
	Wilks' Lambda	.46	250327.34	2	431988	<.01	.537
	Hotelling's Trace	1.16	250327.34	2	431988	<.01	.537
	Roy's Largest Root	1.16	250327.34	2	431988	<.01	.537

Table 4.8

Tests of Between-Subject Effects for MSEs of Overall Effect-size Estimate ($\gamma = 0$)

Source	Estimators	*SS*	*df*	*MS*	*F*	*p*	η^2
Number of Missing *r*s	MSE_ES1	.68	6	.11	535.30	<.01	.069
	MSE_ES2	.77	6	.13	5387.71	<.01	.070
Reliability (Factor loading) of	MSE_ES1	19.80	3	6.60	310937.39	<.01	.683
6 variables	MSE_ES2	22.48	3	7.49	312586.73	<.01	.685
Number of study included (*k*)	MSE_ES1	1.41	1	1.41	490571.08	<.01	.532
	MSE_ES2	11.75	1	11.75	490052.68	<.01	.531
Error	MSE_ES1	9.17	431989	.00			
	MSE_ES2	1.36	431989	.00			
Total	MSE_ES1	68.87	432000				
	MSE_ES2	78.07	432000				

Table 4.9

Results from Multivariate Analysis of Variance (MANOVA) on the Bias of Estimators for $\gamma = .5$

Effect		Value	F	df_1	df_2	p	η^2
Number of Missing *r*s	Pillai's Trace	.12	442.06	12	863978	<.01	.058
	Wilks' Lambda	.89	4459.44	12	863976	<.01	.058
	Hotelling's Trace	.12	4498.84	12	863974	<.01	.059
	Roy's Largest Root	.10	6954.08	6	431989	<.01	.088
Reliability (Factor loading) of 6 variables	Pillai's Trace	1.69	779286.84	6	863978	<.01	.844
	Wilks' Lambda	.00	1894142.01	6	863976	<.01	.929
	Hotelling's Trace	6.49	4355172.78	6	863974	<.01	.968
	Roy's Largest Root	58.10	8366262.43	3	431989	<.01	.983
Number of study included (*k*)	Pillai's Trace	.96	4984185.66	2	431988	<.01	.958
	Wilks' Lambda	.04	4984185.66	2	431988	<.01	.958
	Hotelling's Trace	23.08	4984185.66	2	431988	<.01	.958
	Roy's Largest Root	23.08	4984185.66	2	431988	<.01	.958

Table 4.10

Tests of Between-Subject Effects for Bias of Overall Effect-size Estimate (γ = .5)

Source	Estimators	*SS*	*df*	*MS*	*F*	*p*	η^2
Number of Missing *r*s	Bias_ES1	13.54	6	2.26	6674.26	<.01	.085
	Bias_ES2	14.21	6	2.37	6656.07	<.01	.085
Reliability (Factor loading)	Bias_ES1	349.10	3	116.37	344171.06	<.01	.705
of 6 variables	Bias_ES2	367.47	3	122.49	34436.33	<.01	.705
Number of study included	Bias_ES1	.02	1	.02	7.52	<.01	.000
(*k*)	Bias_ES2	.39	1	.39	1105.14	<.01	.003
	Bias_ES1	146.06	431989	.00			
Error	Bias_ES2	153.66	431989	.00			
	Bias_ES1	539.54	432000				
Total	Bias_ES2	778.07	432000				

Table 4.11

Results from Multivariate Analysis of Variance (MANOVA) on the MSEs of Estimators for $\gamma = .5$

Effect		Value	F	df_1	df_2	p	η^2
Number of Missing *r*s	Pillai's Trace	.14	5564.54	12	863978	<.01	.072
	Wilks' Lambda	.86	5784.70	12	863976	<.01	.074
	Hotelling's Trace	.17	6005.49	12	863974	<.01	.077
	Roy's Largest Root	.16	1187.50	6	431989	<.01	.142
Reliability (Factor loading) of 6 variables	Pillai's Trace	1.37	310998.25	6	863978	<.01	.684
	Wilks' Lambda	.09	333413.49	6	863976	<.01	.698
	Hotelling's Trace	4.96	356932.97	6	863974	<.01	.713
	Roy's Largest Root	3.53	508628.23	3	431989	<.01	.779
Number of study included (*k*)	Pillai's Trace	.52	233526.95	2	431988	<.01	.520
	Wilks' Lambda	.48	233526.95	2	431988	<.01	.520
	Hotelling's Trace	1.08	233526.95	2	431988	<.01	.520
	Roy's Largest Root	1.08	233526.95	2	431988	<.01	.520

Table 4.12

Tests of Between-Subject Effects for MSEs of Overall Effect-size Estimate (γ = .5)

Source	Estimators	*SS*	*df*	*MS*	*F*	*p*	η^2
Number of Missing *r*s	MSE_ES1	1.73	6	.288446	1047.93	<.01	.127
	MSE_ES2	2.26	6	.376086	11032.71	<.01	.133
Reliability (Factor loading) of	MSE_ES1	19.72	3	6.572996	238607.52	<.01	.624
6 variables	MSE_ES2	22.98	3	7.658813	224676.00	<.01	.609
Number of study included (*k*)	MSE_ES1	8.48	1	8.481522	307889.25	<.01	.416
	MSE_ES2	9.23	1	9.227018	27068.27	<.01	.385
Error	MSE_ES1	11.90	431989	2.75E-05			
	MSE_ES2	14.73	431989	3.41E-05			
Total	MSE_ES1	74.32	432000				
	MSE_ES2	89.33	432000				

Table 4.13

Correlation Matrix of Six Indicators

	x_1 (reliability = .9)	x_2 (reliability =.5)	x_3 (reliability = .2)
y_1 (reliability = .9)		$r_{x2,y1}$ (medium, high)	$r_{x3,y1}$ (low, high)
y_2 (reliability = .5)	$r_{x1,y2}$ (high, medium)		$r_{x3,y2}$ (low, medium)
y_3 (reliability = .2)	$r_{x1,y3}$ (high, low)	$r_{x2,y3}$ (medium, low)	

Table 4.14

ANOVAs Comparing Bias of ES1 Across Which r is Included

γ	Source	*SS*	*df*	*MS*	*F*	*p*	η^2
0	Intercept	.02	1	.02	.65	.42	0
	which *r* is included	.38	5	.08	2.07	.07	0
	Error	88.31	23,994	.04			
	Total	88.71	24,000				
.5	Intercept	1,589.6	1	1,589.60	41,261.5	<.01	.632
	which *r* is included	2.4	5	.49	12.69	<.01	.003
	Error	924.4	23,994	.04			
	Total	2,516.42	24,000				

Table 4.15

Pairwise Comparison Comparing Bias of ES1 Depending On Which r is Included

Ind	Ind	Mean Difference	*SE*	*p*	95% CI	
					Upper	Lower
1	2	.012*	.004	.008	.003	.020
	3	.023*	.004	.000	.014	.032
	4	-.006	.004	.206	-.014	.003
	5	.014*	.004	.002	.005	.023
	6	.019*	.004	.000	.011	.028
2	1	-.012*	.004	.008	-.020	-.003
	3	.011*	.004	.010	.003	.020
	4	-.017*	.004	.000	-.026	-.009
	5	.002	.004	.614	-.006	.011
	6	.007	.004	.090	-.001	.016
3	1	-.023*	.004	.000	-.032	-.014
	2	-.011*	.004	.010	-.020	-.003
	4	-.029*	.004	.000	-.037	-.020
	5	-.009*	.004	.040	-.018	.000
	6	-.004	.004	.387	-.012	.005
4	1	.006	.004	.206	-.003	.014
	2	.017*	.004	.000	.009	.026
	3	.029*	.004	.000	.020	.037
	5	.019*	.004	.000	.011	.028
	6	.025*	.004	.000	.016	.033
5	1	-.014*	.004	.002	-.023	-.005
	2	-.002	.004	.614	-.011	.006
	3	.009*	.004	.040	.000	.018
	4	-.019*	.004	.000	-.028	-.011
	6	.005	.004	.233	-.003	.014
6	1	-.019*	.004	.000	-.028	-.011
	2	-.007	.004	.090	-.016	.001
	3	.004	.004	.387	-.005	.012
	4	-.025*	.004	.000	-.033	-.016
	5	-.005	.004	.233	-.014	.003

Note. 1 = $r_{(x1,y2)}$; 2 = $r_{(x1,y3)}$; 3 = $r_{(x2,y3)}$; 4 = $r_{(x2,y1)}$; 5 = $r_{(x3,y1)}$; 6 = $r_{(x3,y2)}$

Table 5.1

Measures used to represent teachers' and students' knowledge in 8 studies

Studies	Measures of Teachers' Knowledge	Measures of Students' Knowledge
Bassham (1962)	Glennon test score of basic mathematical understanding	California Achievement Test (CAT)
Caezza (1969)	Callahan Test of Mathematical Knowledge	Stanford Achievement Test (SAT)
Cox (1970)	Dr. Leroy's Test of Mathematical Understanding	Achievement Series (SRA)
Koch (1972)	Test of Understandings of the Real Number System (TURNS)	Grade Equivalent Scores from CTBS
Lampela (1966)	Stoneking Test of Basic Arithmetical Principles and Generalizations	California Achievement Test (CAT)
Moore (1964)	Glennon test score of basic mathematical understanding	Achievement Series (SRA)
Turgoose (1996)	Tests of Achievement and Proficiency (TAP)	Iowa Test Basic Skill (ITBS)
Prekeges (1973)	Test of Teacher understanding	Growth in mathematics

Table 5.2.

Description of 8 Studies

Study	Description			Teacher Subject Matter Knowledge Test						Student Mathematical Achievement Test						Analysis		
	Source	*N* of Tch	*N* of Std	Name	Type	Format	*N* of Item	Reliability		Name	Type	Format	*N* of Item	Reliability		Unit	*N* for Link	*R*
								Type	Value					Type	Value			
Bassham (1962)	1	28	NI	Glennon Mathematical Understanding	1	Multi	NI	Alpha	.89	California Achievement Test (CAT)	3	NI	NI	NI	NI	CL	28	.27
Caezza (1969)	2	18	465	Callahan Test of Mathematical Knowledge	1	Multi	40	KR - 20	.86	Stanford Achievement Test (SAT) : Concept	3	NI	NI	Split - Half	.77 ~ .95	Std	465	0
										Stanford Achievement Test (SAT) : Computation	3	NI	NI			Std	465	.10
										Stanford Achievement Test (SAT): Application	3	NI	NI			Std	465	.02

Table 5.2 Continued

Cox (1970)	.	12	216	Dr. Leroy's Test of Mathematical Understanding	1	Multi	45	NI	.59	Achievement Series(SRA) : Total	3	NI	NI	NI	NI	Std	216	.16
										Achievement Series(SRA) : Concept						Std	216	.14
										Achievement Series(SRA): Computation						Std	216	.09
										Achievement Series(SRA): Reasoning						Std	216	.02
Koch (1972)	2	26	NI	Test of Understandings of the Real Number System (TURNS)	1	Multi	50	NI	.85	Total Arithmetic (Grade Equivalent Scores) from CTBS	3	NI	98	NI	.80	CL	26	.07
Lampela (1966)	2	24	NI	Stoneking Test of Basic Arithmetical Principles and Generalizations	1	Multi	67	Froe-lich	.92	Edwards	3	NI	NI	NI	NI	CL	24	.16
Moore (1964)	2	11	NI	Glennon Mathematical Understanding	1	Multi	NI	Alpha	.89	Achievement Series(SRA)-Reasoning	3	NI	49	NI	.87	CL	11	-.02
										Achievement Series(SRA)-Computation		NI	44	NI	.78	CL	11	-.26
										Achievement Series(SRA)-Concept		NI	20	NI	.81	CL	11	-.30

Table 5.2 Continued

Prekeges (1973)	3	61	1611	Test of Teacher Understanding	1	Multi	60	Test/ Retest	.79	Growth in Understanding	1	Multi	42 ~ 44	Test/ Retest	.74 ~ .87	Std	1611	-.01
										Growth in Problem Solving		Multi	13 ~ 16	Test/ Retest	.45 ~ .80	Std	1611	.07
										Growth in Computation		Multi	23 ~ 30	Test/ Retest	.33 ~ .88	Std	1611	-.08
Turgoose (1996)	2	80	NI	Tests of Achievement and Proficiency (TAP)	3	Multi	48	Test/ Retest	.75 ~ .92	Iowa Test Basic Skill (ITBS)	3	NI	NI	Alpha	.89	CL	80	.28

Note.
Sources: 1 = Journal article, 2 = Dissertation, 3 = Reports, and 4 = Conference paper.
Test Type (1 = Researcher-made: Local; 2 = Researcher-made: Large scale; 3 = Commercial), Test Format (Multi = Multiple-choice item; Open-ended = Open-ended item), Unit (Std = Student; CL = Classroom)
N (Number of samples in the correlation). R: Correlation values

Data Analysis	StudyID	Teacher's Content Knowledge						Student Achievement					
		Educational Background						Standardized Test					
		Course work	Major	Degree	GPA	Other	Test	NELS	MAT	NAEP	CAT	ITBS	Researcher-made Test
Correlation	Brown88	*											*
	Reed86											*	
	Hurst67	*											
	Rouse67	*									*		
	Soeteber69				*						*		
	⋮												
Regression	BrewGold96			*				*					
	MonkKing					*				*			
	Chaney95	*			*					*			
	GoldBrew			*				*					
	Turgoose96						*						
	Teddlie83						*						*
	⋮												
HLM	Devaney97												
	Kim92	*								*			
	HillRB05						*						*
	Chiang96	*	*				*	*					
	⋮												
Other	Hurst67D	*							*			*	
	Godgart64	*											
	⋮												

Figure 1.1.

An empirical example from Ahn and Choi (2004)

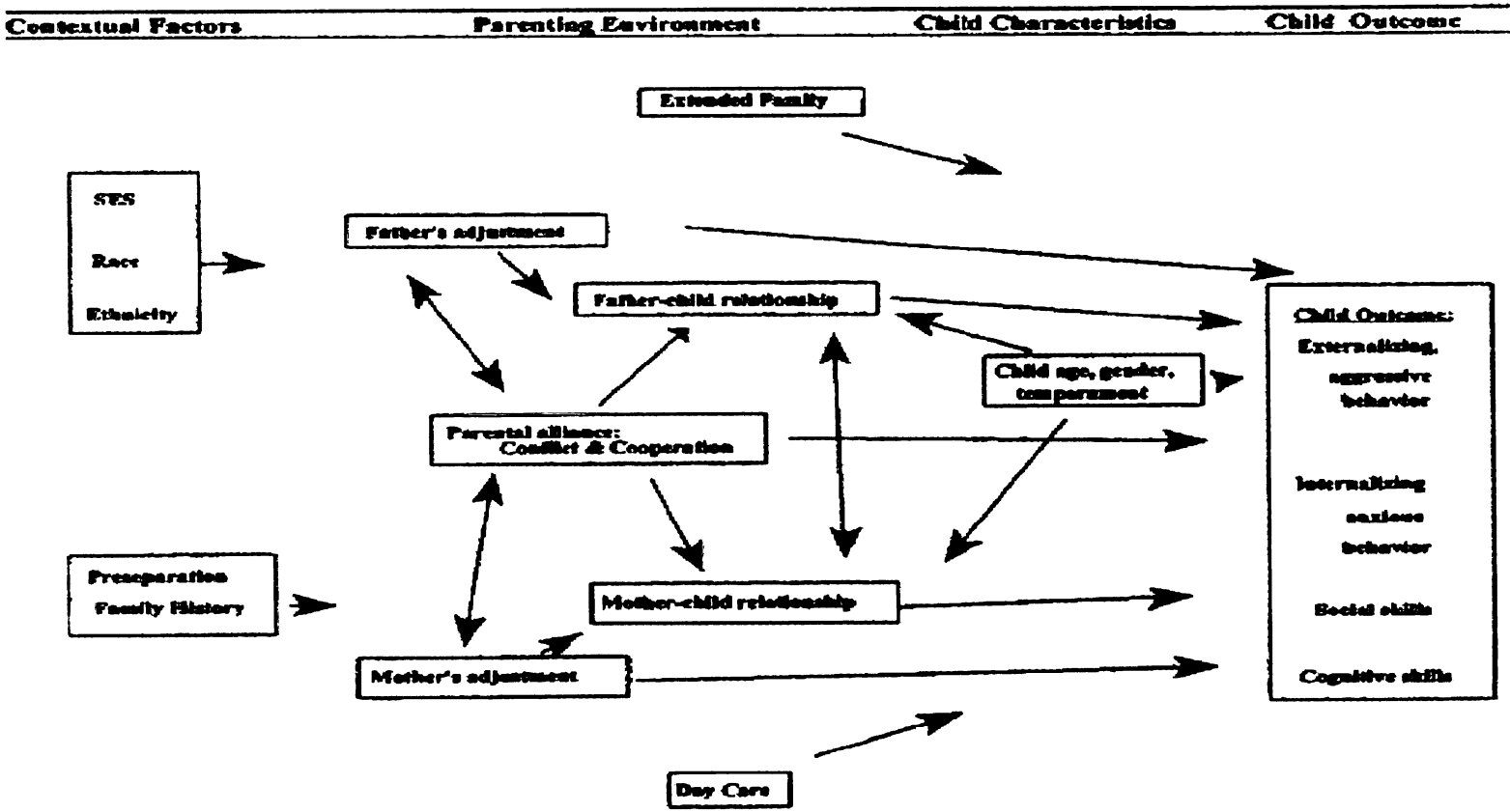


Figure 2.1

The underlying model used in a meta-analysis by Whiteside and Becker (2000)[7].

[7] This path diagram used in Whiteside and Becker (2000) is extracted from Becker (2001).

Study	ξ						η					
	x_1	x_2	x_3	⋯	x_{p-1}	x_p	y_1	y_2	y_3	⋯	y_{q-1}	y_q
1	√						√					
2		√	√			√		√				
3			√								√	
:												
k-2				√						√		
k-1	√					√					√	
k		√				√		√				

Figure 3.1.

A hypothetical meta-analysis with *k* studies

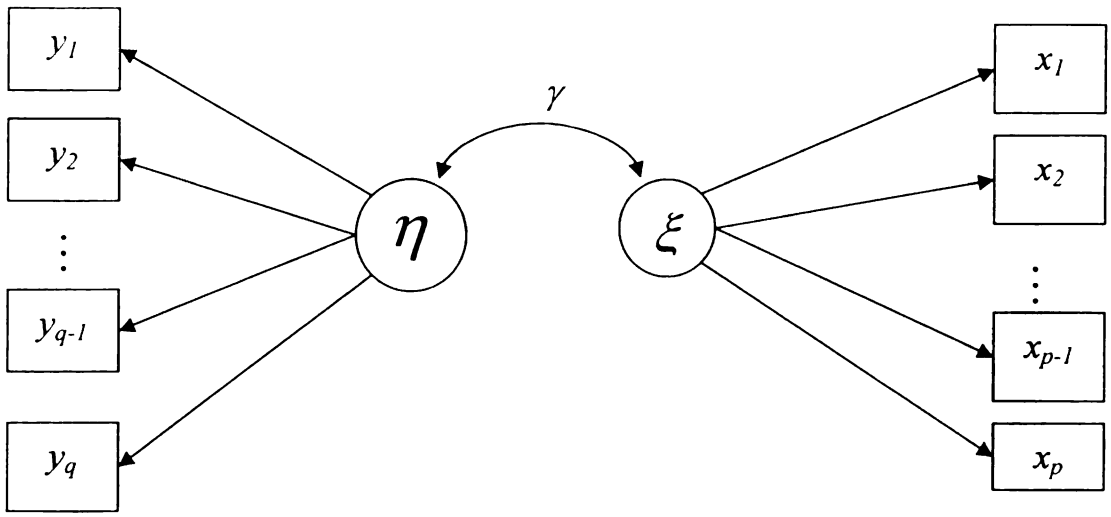


Figure 3.2.

A population model for a hypothetical meta-analysis

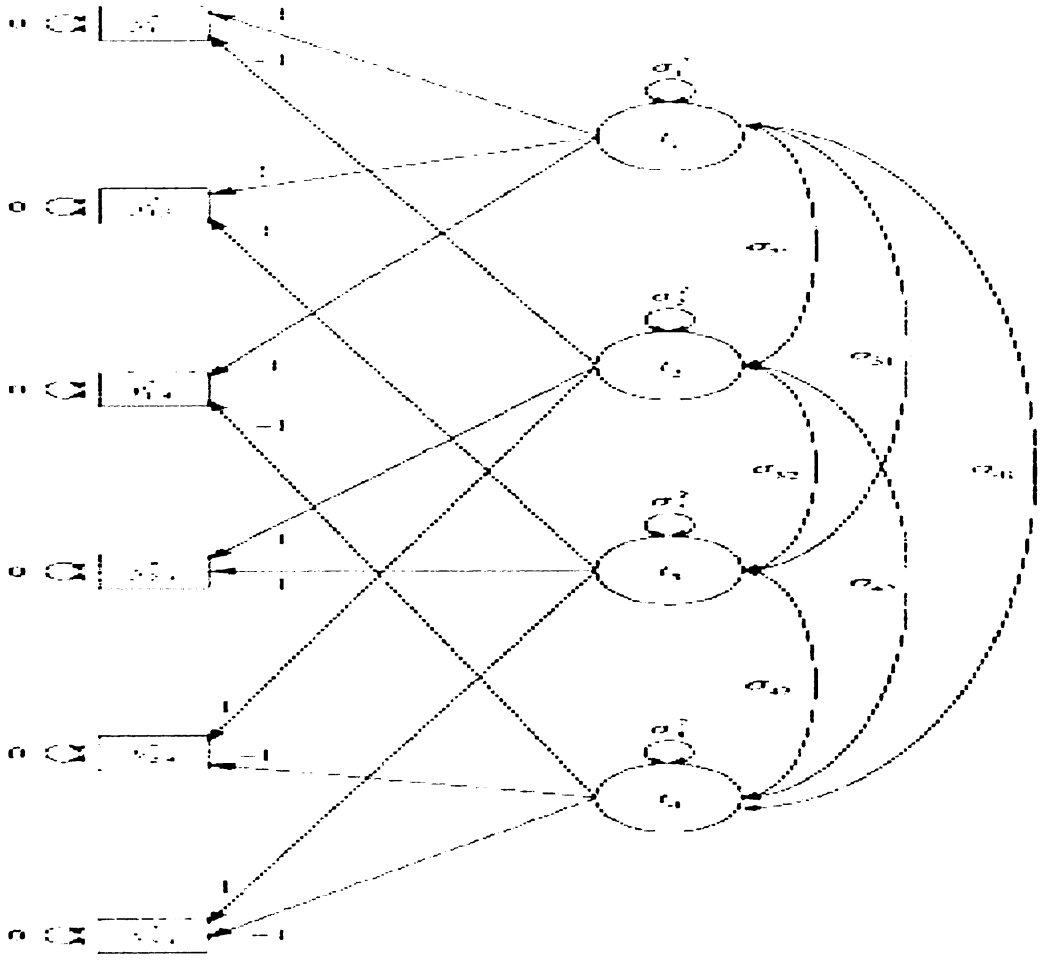
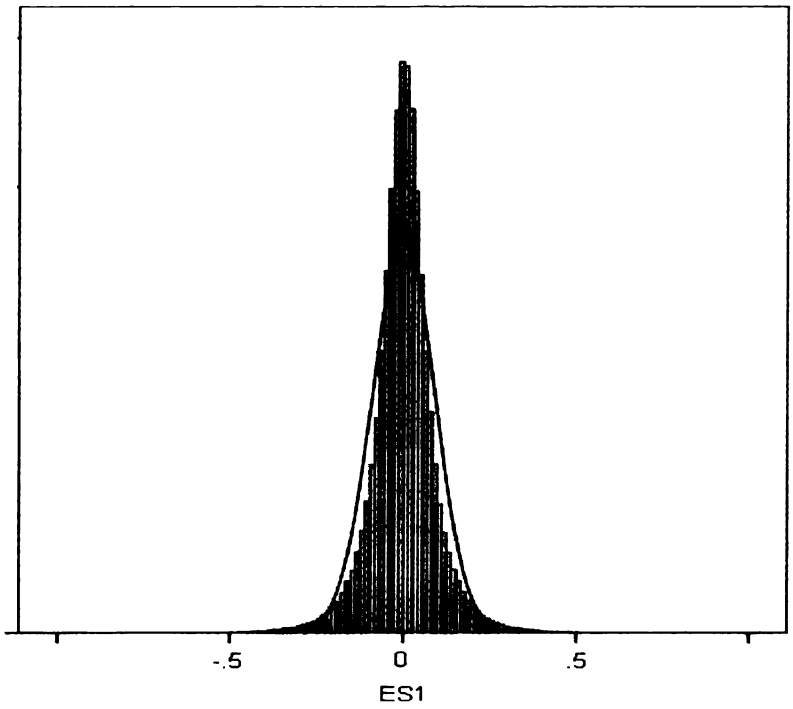


Figure 3.3

Covariance structure model for ranking data with $p = 4$ alternatives[8]

[8] Example is from Maydeu-Olivares & Böckenholt (2005).



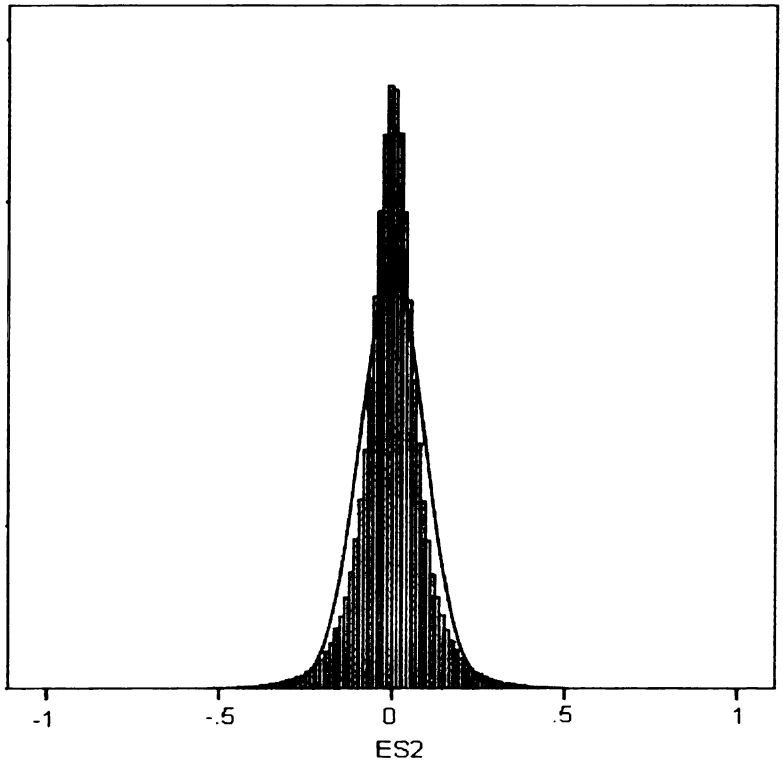
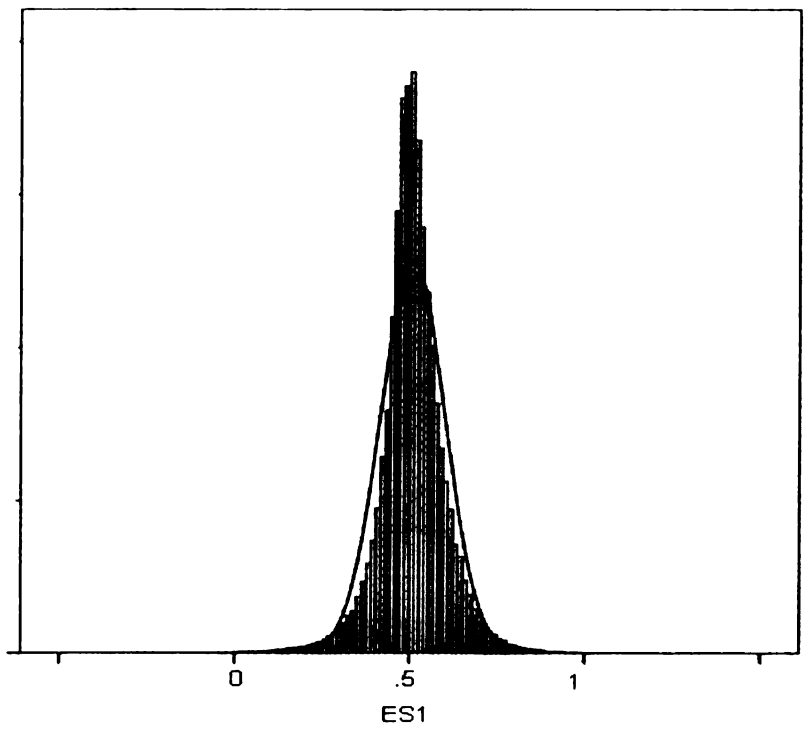


Figure 4.1

Histograms of estimators when γ is set to 0



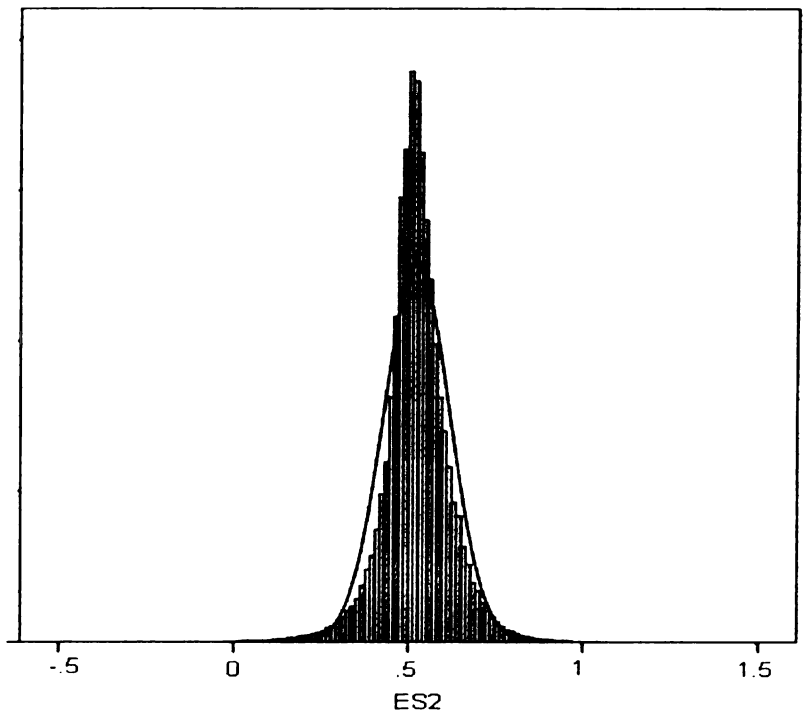


Figure 4.2

Histograms of estimators when γ is set to .5

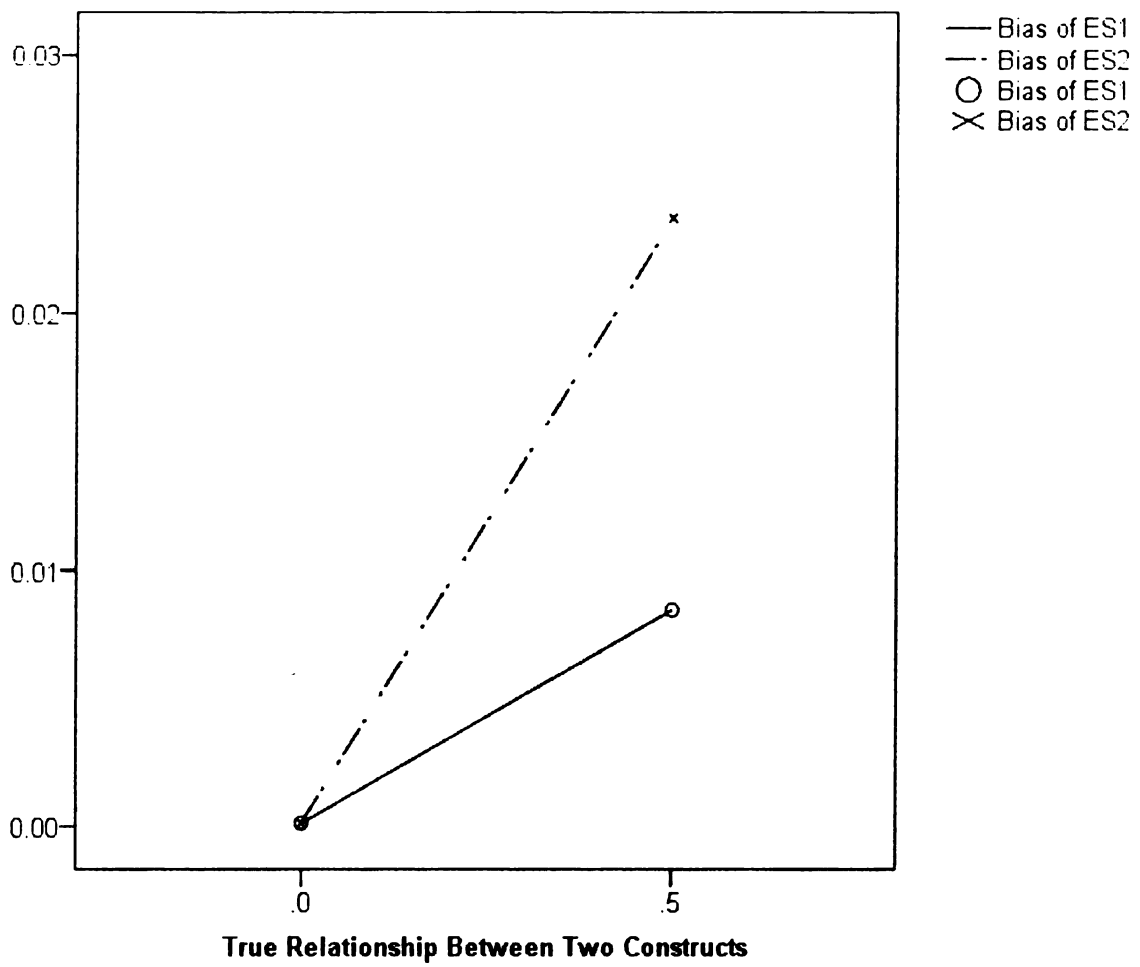


Figure 4.3

Bias of two estimators depending on the true population relationship between two underlying constructs (γ)

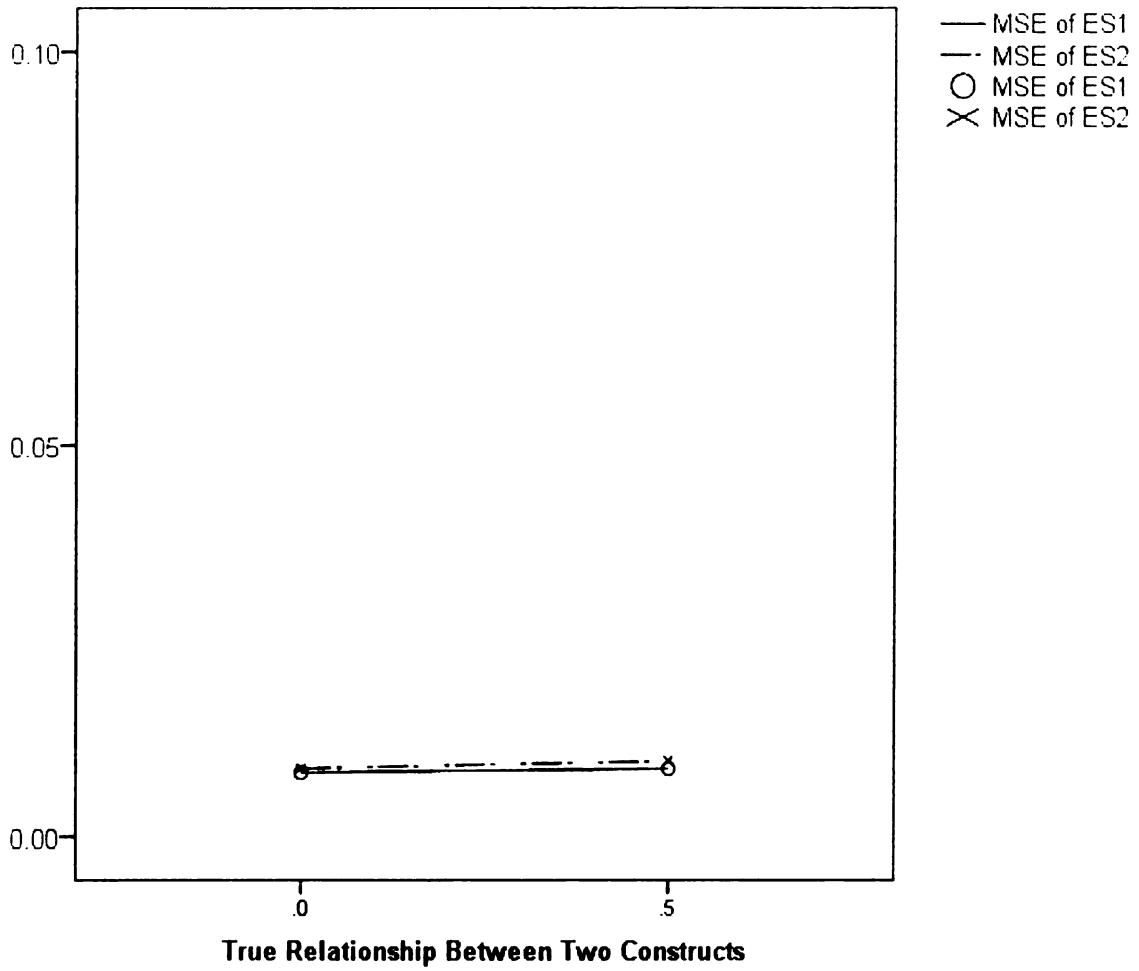


Figure 4.4

MSEs of two estimators depending on the true population relationship between two underlying constructs (γ)

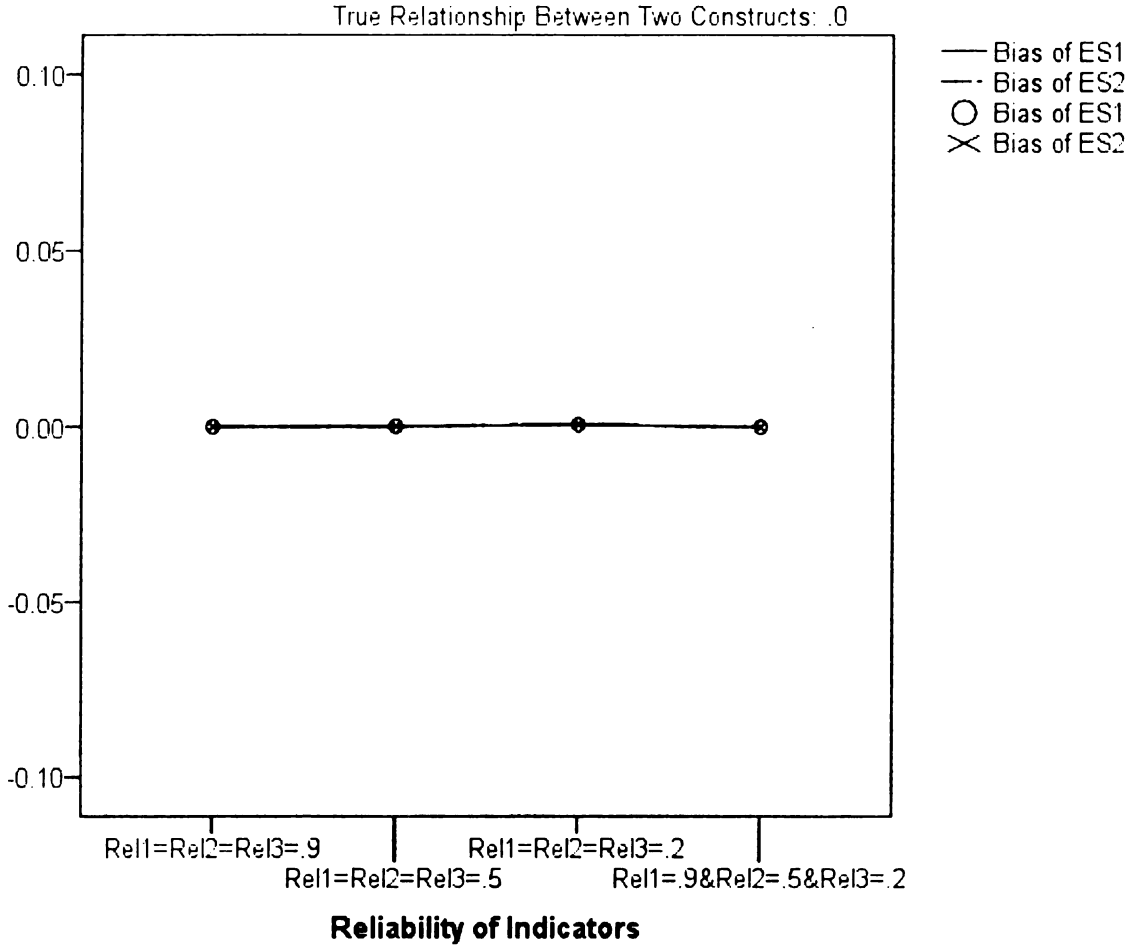


Figure 4.5

Bias of two estimators depending on the reliabilities of indicators when γ is set to 0

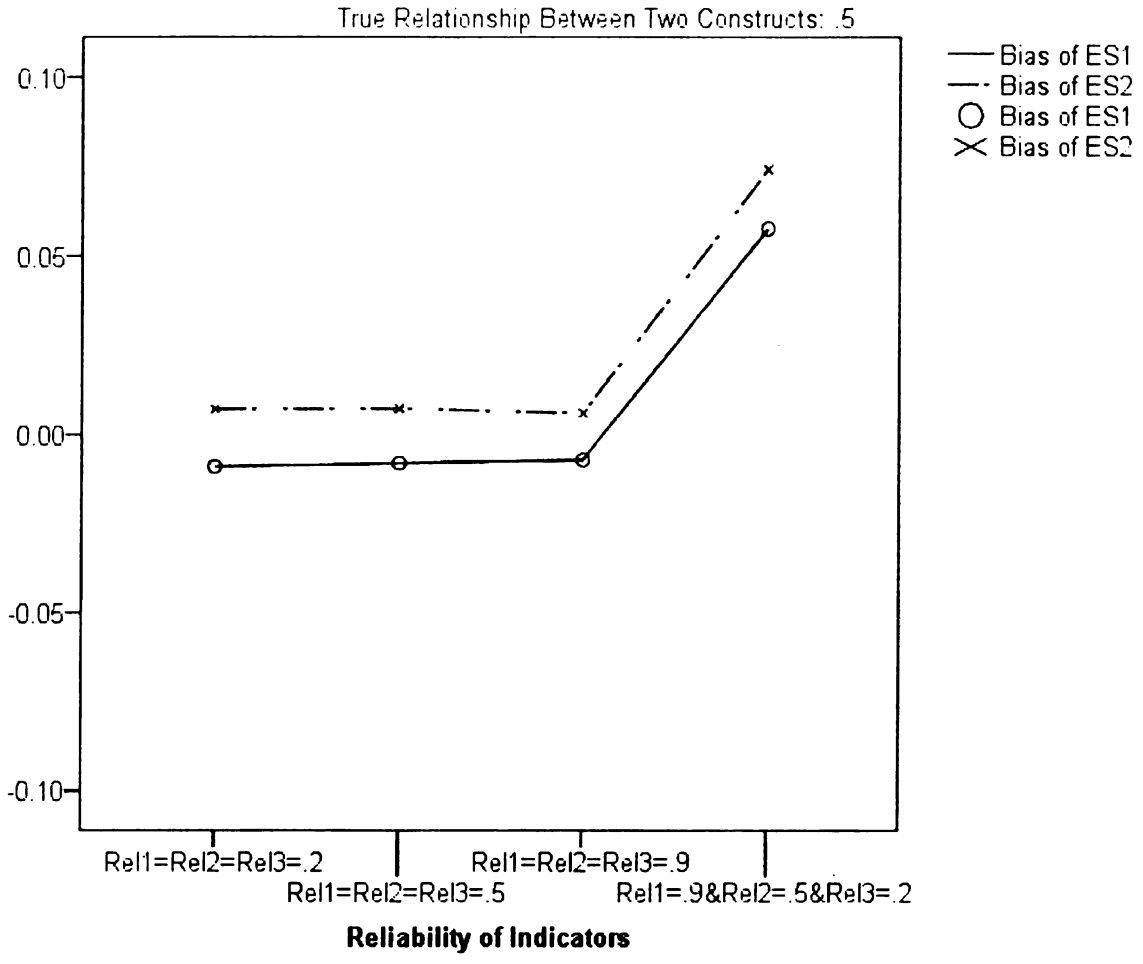


Figure 4.6

Bias of two estimators depending on the reliabilities of indicators when γ is set to .5

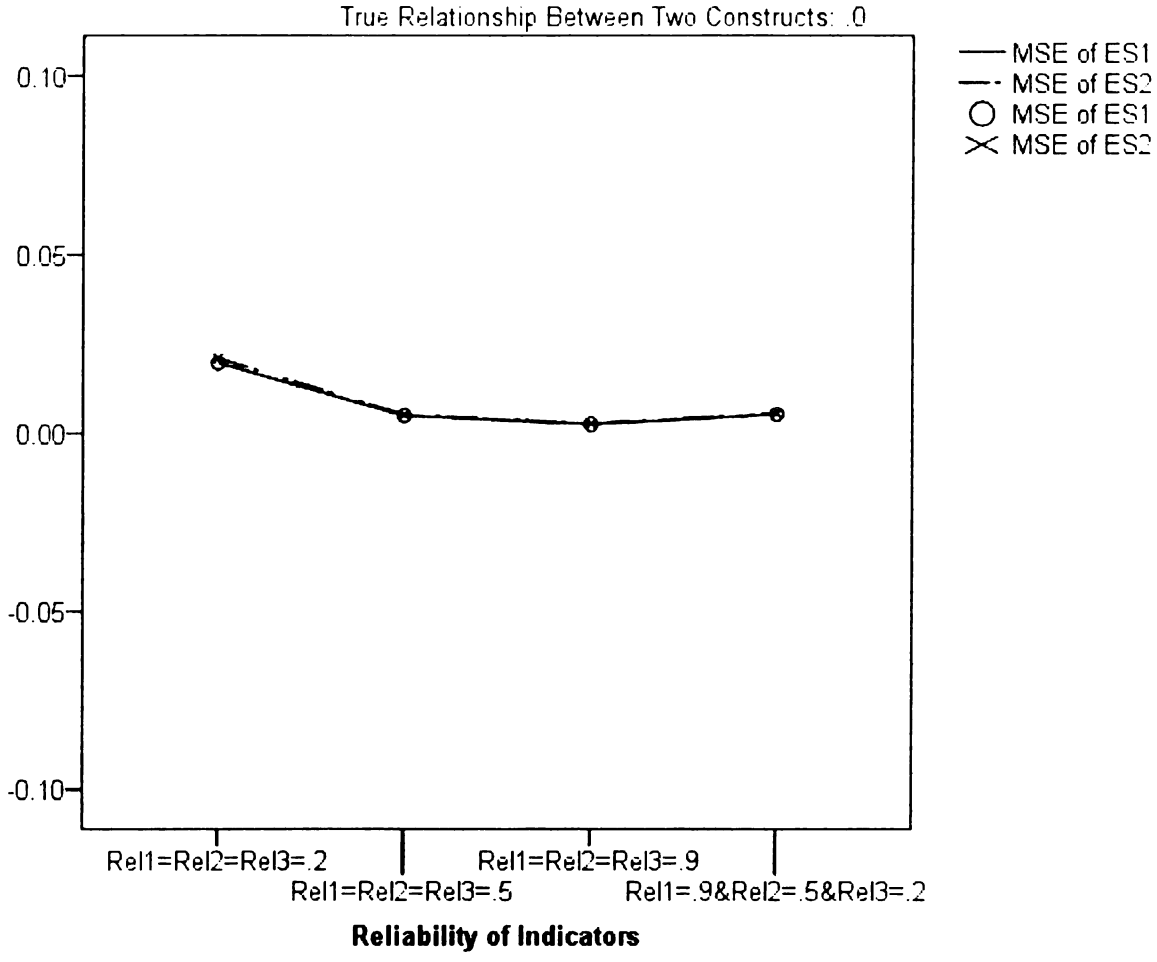


Figure 4.7

MSEs of two estimators depending on the reliabilities of indicators when γ is set to 0

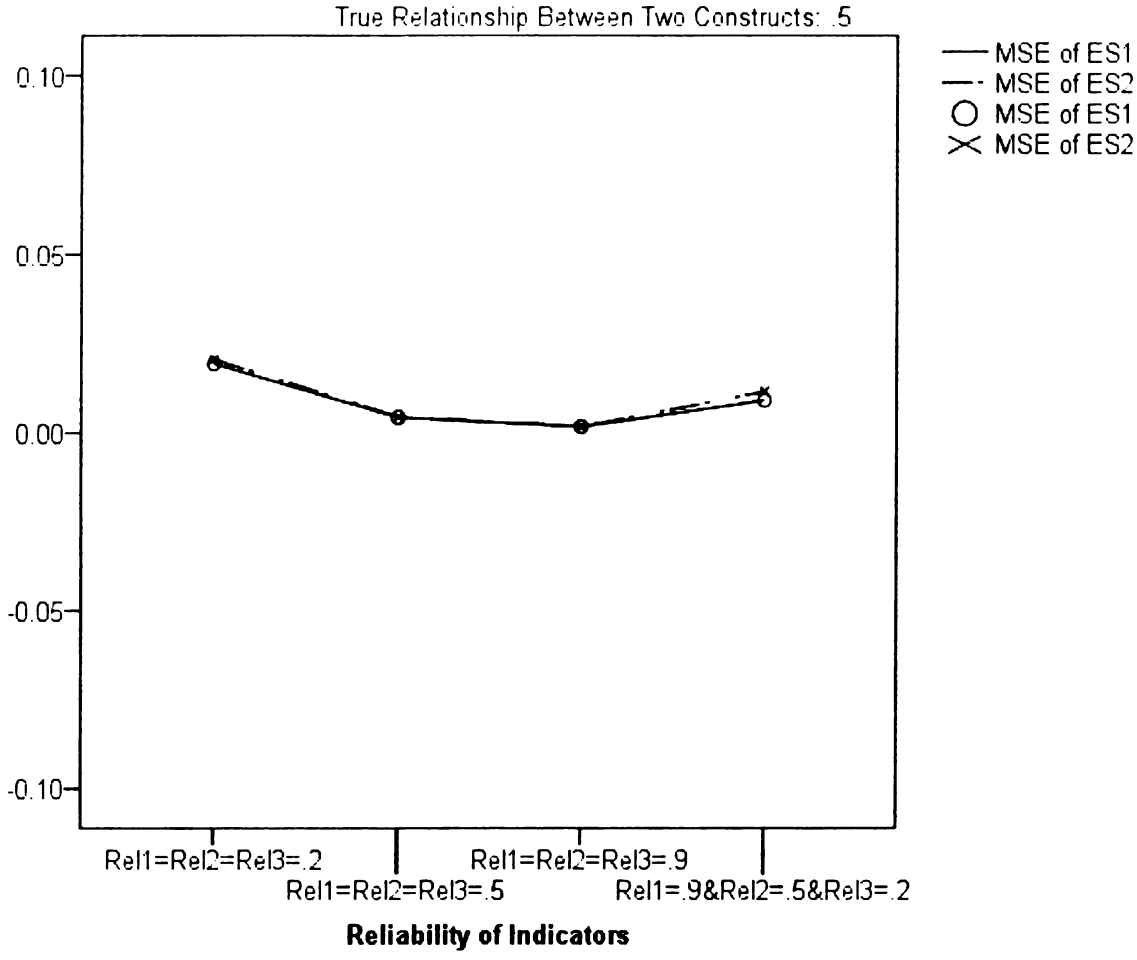


Figure 4.8

MSEs of two estimators depending on the reliabilities of indicators when γ is set to .5

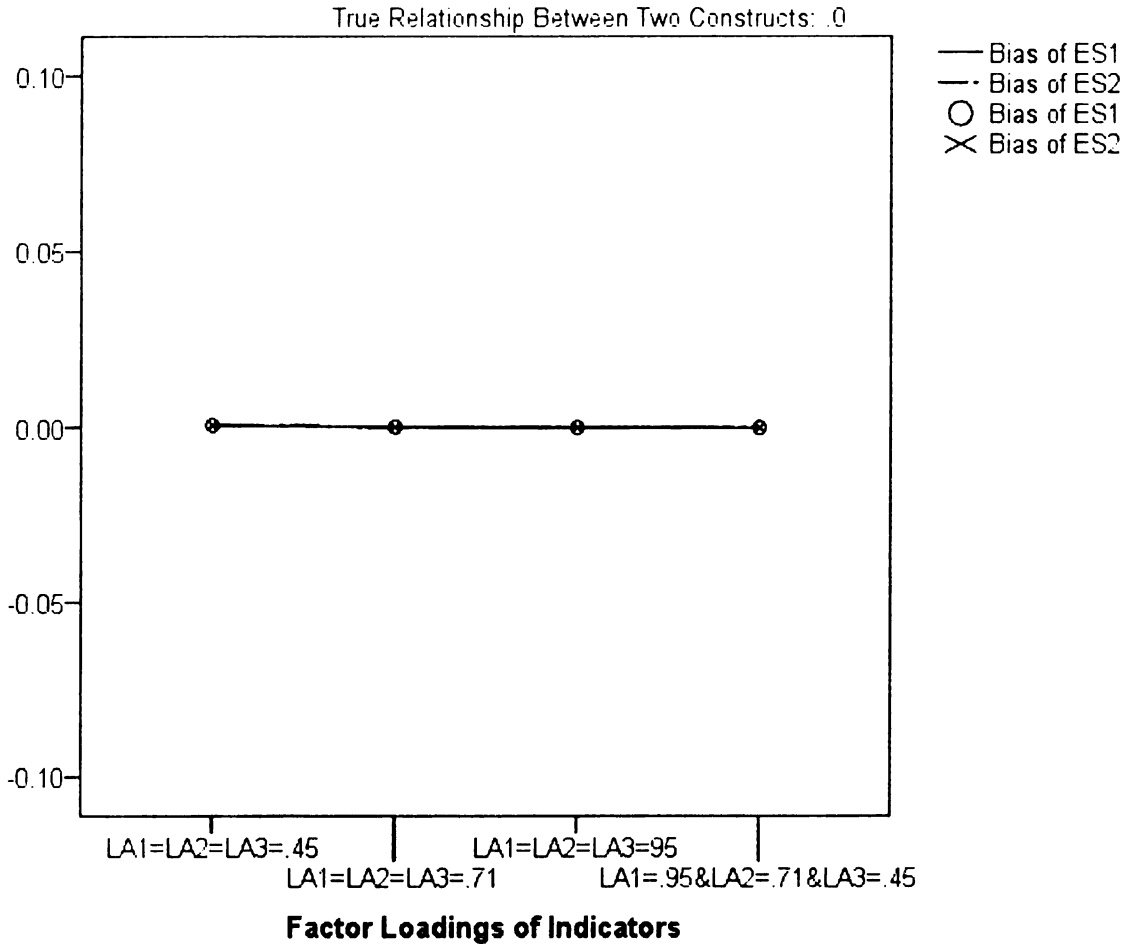


Figure 4.9

Biases of two estimators depending on the factor loadings of indicators when γ is set to 0

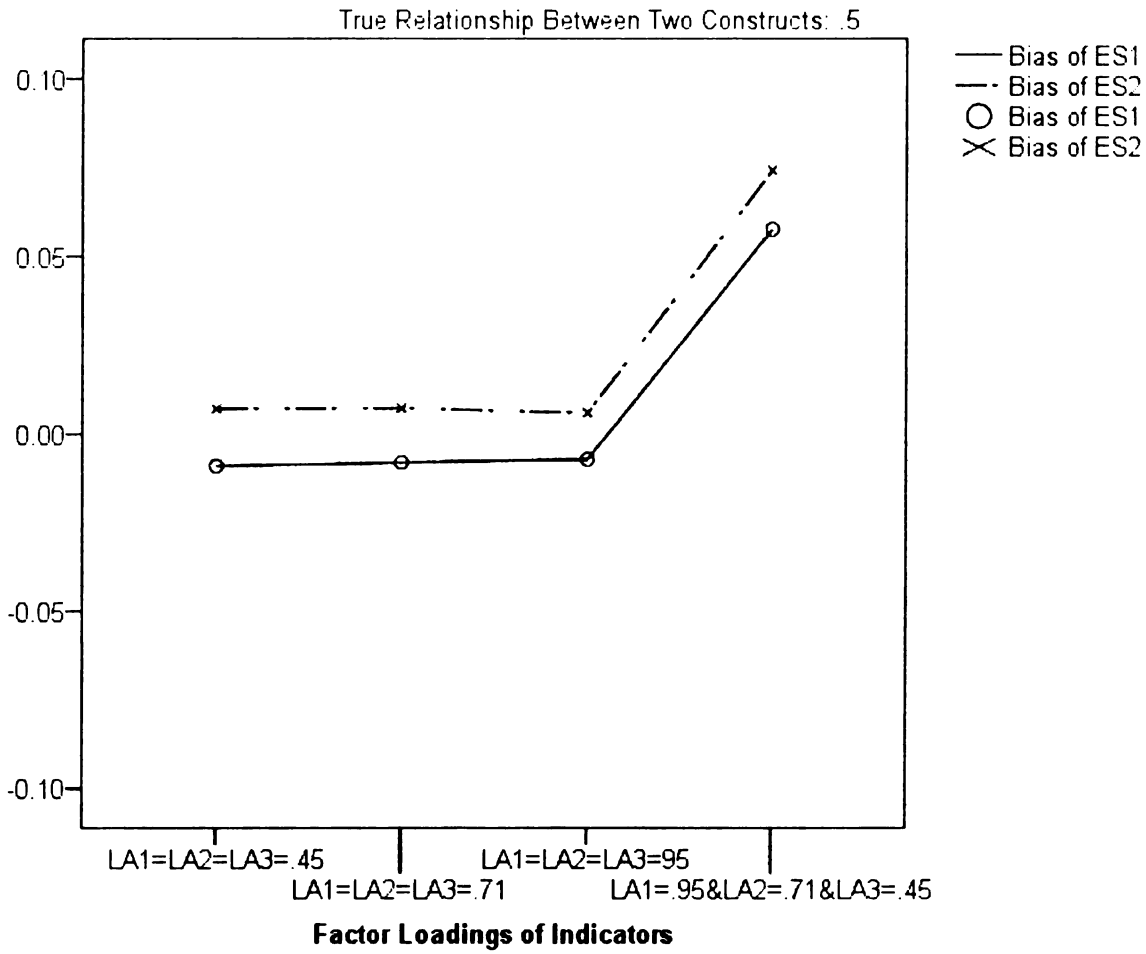


Figure 4.10

Biases of two estimators depending on the factor loadings of indicators when γ is set to .5.

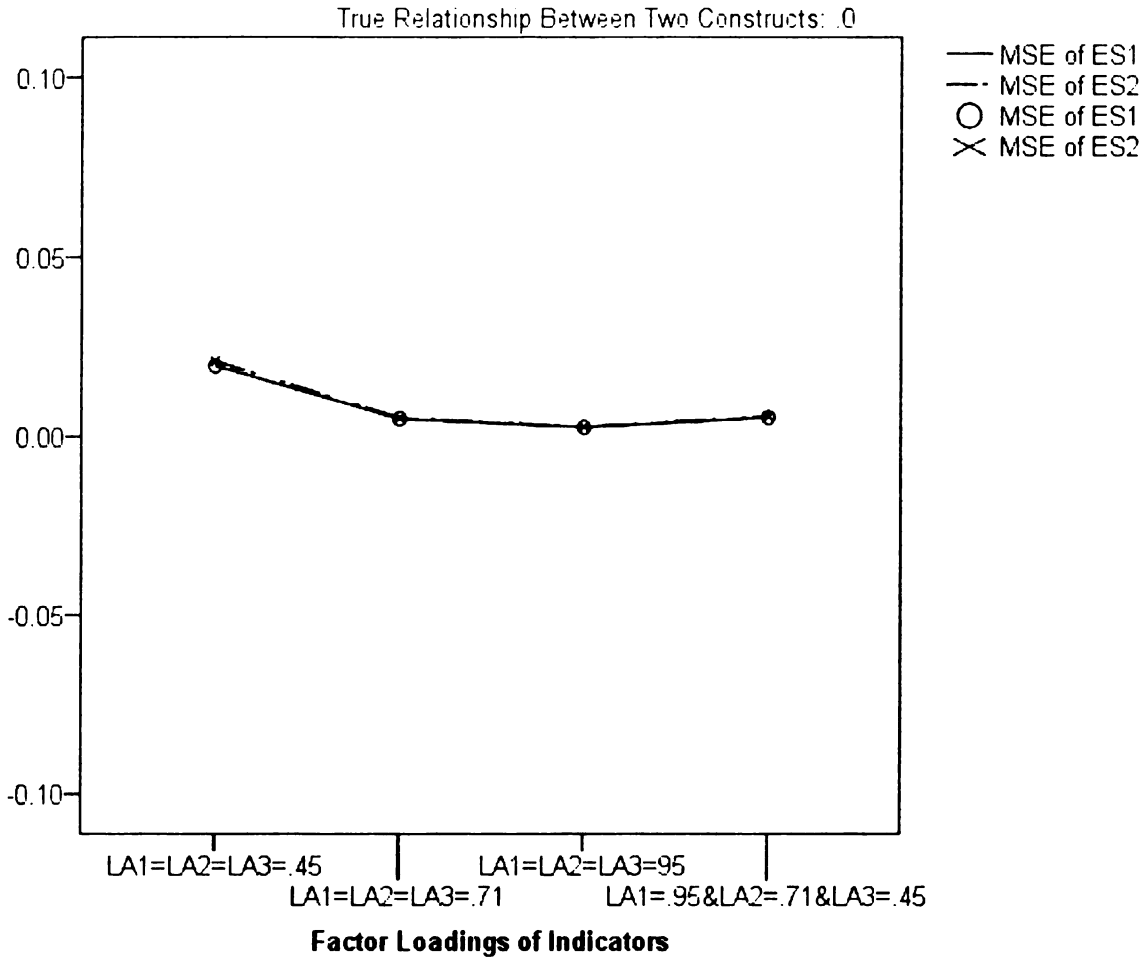


Figure 4.11

MSEs of two estimators depending on the factor loadings of indicators when γ is set to 0.

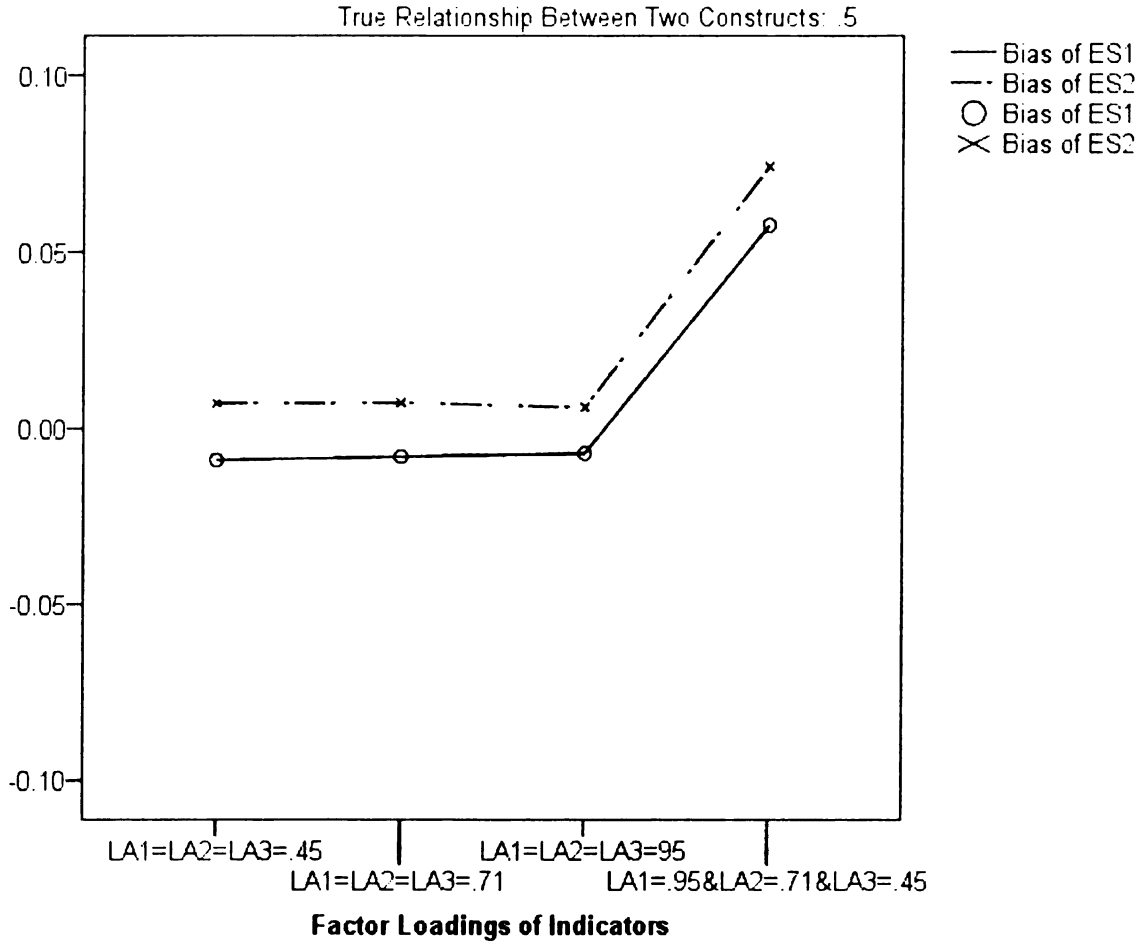


Figure 4.12

MSEs of two estimators depending on the factor loadings of indicators when γ is set to .5.

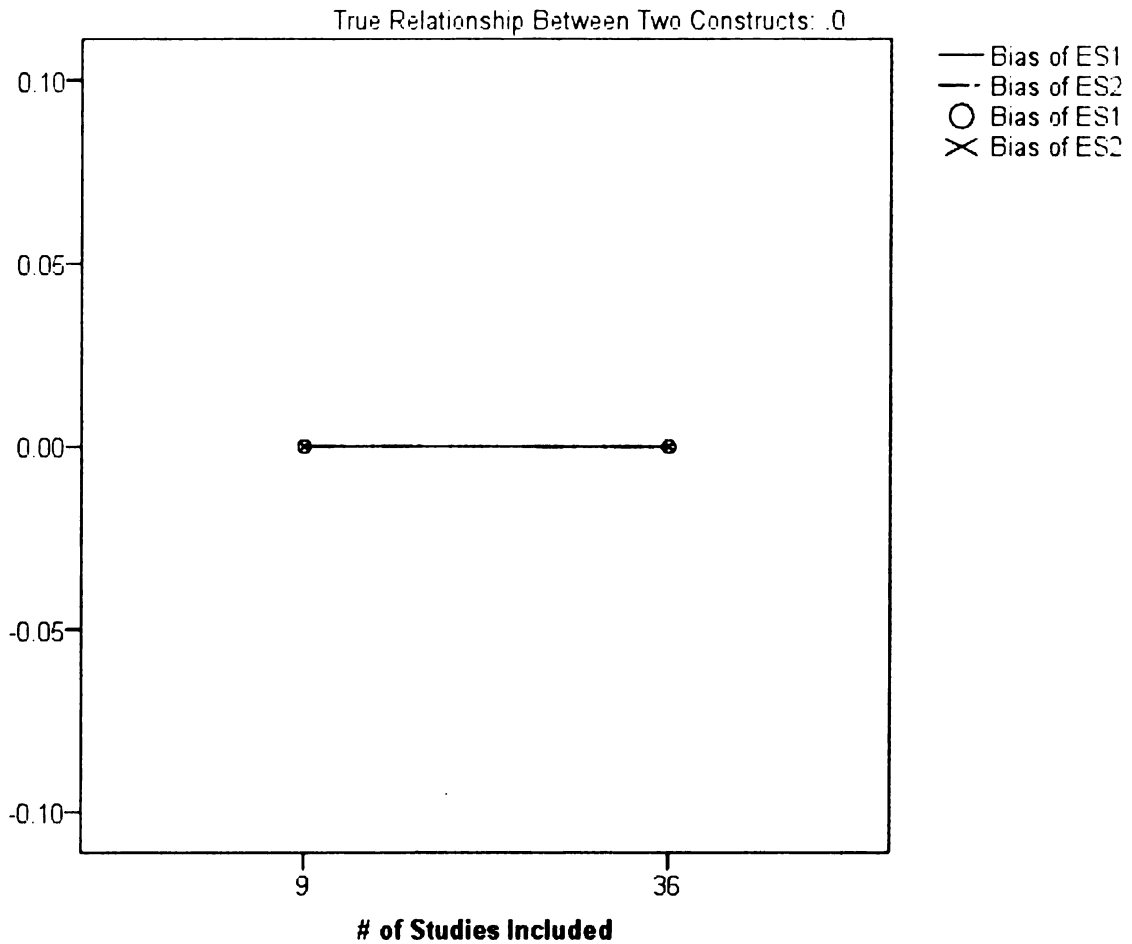


Figure 4.13

Biases of two estimators depending on k when γ is set to 0

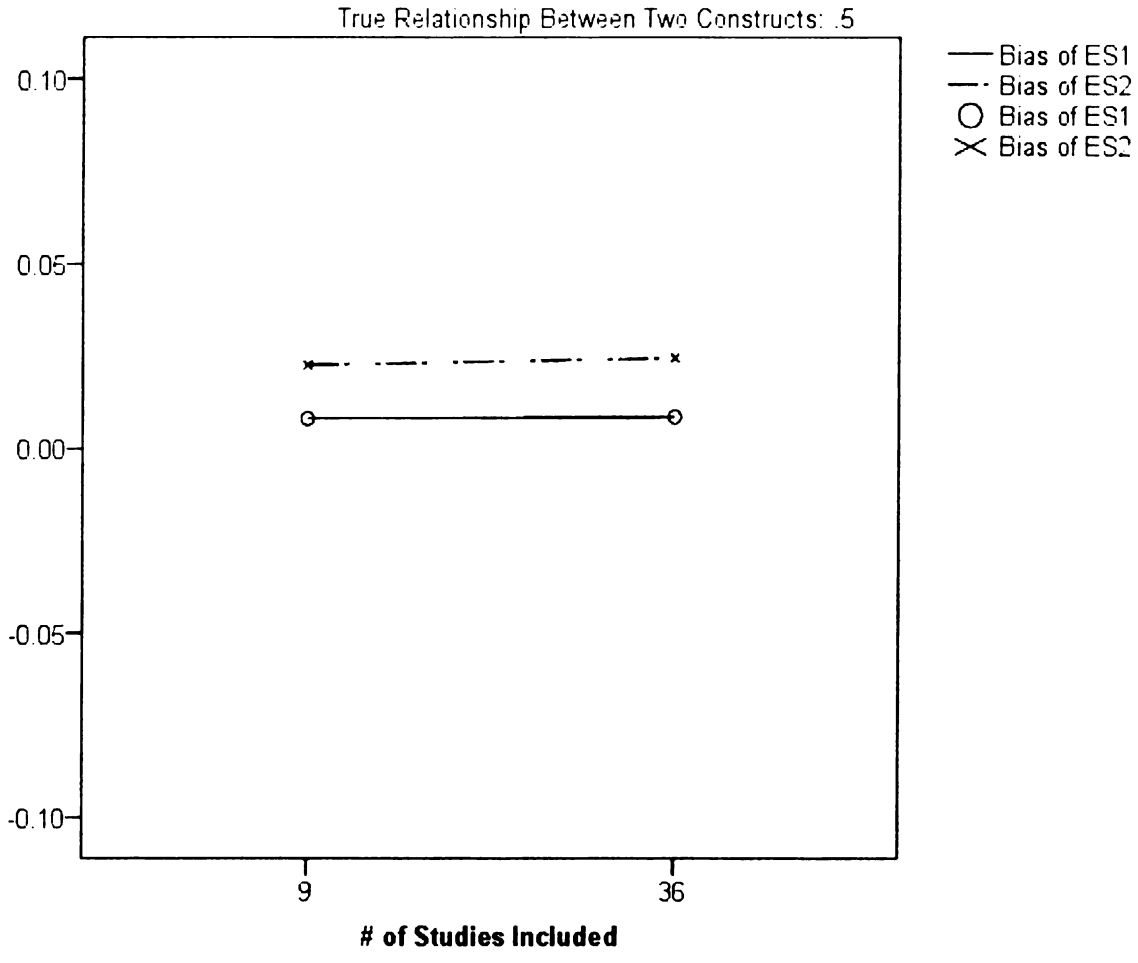


Figure 4.14

Biases of two estimators depending on k when γ is set to .5

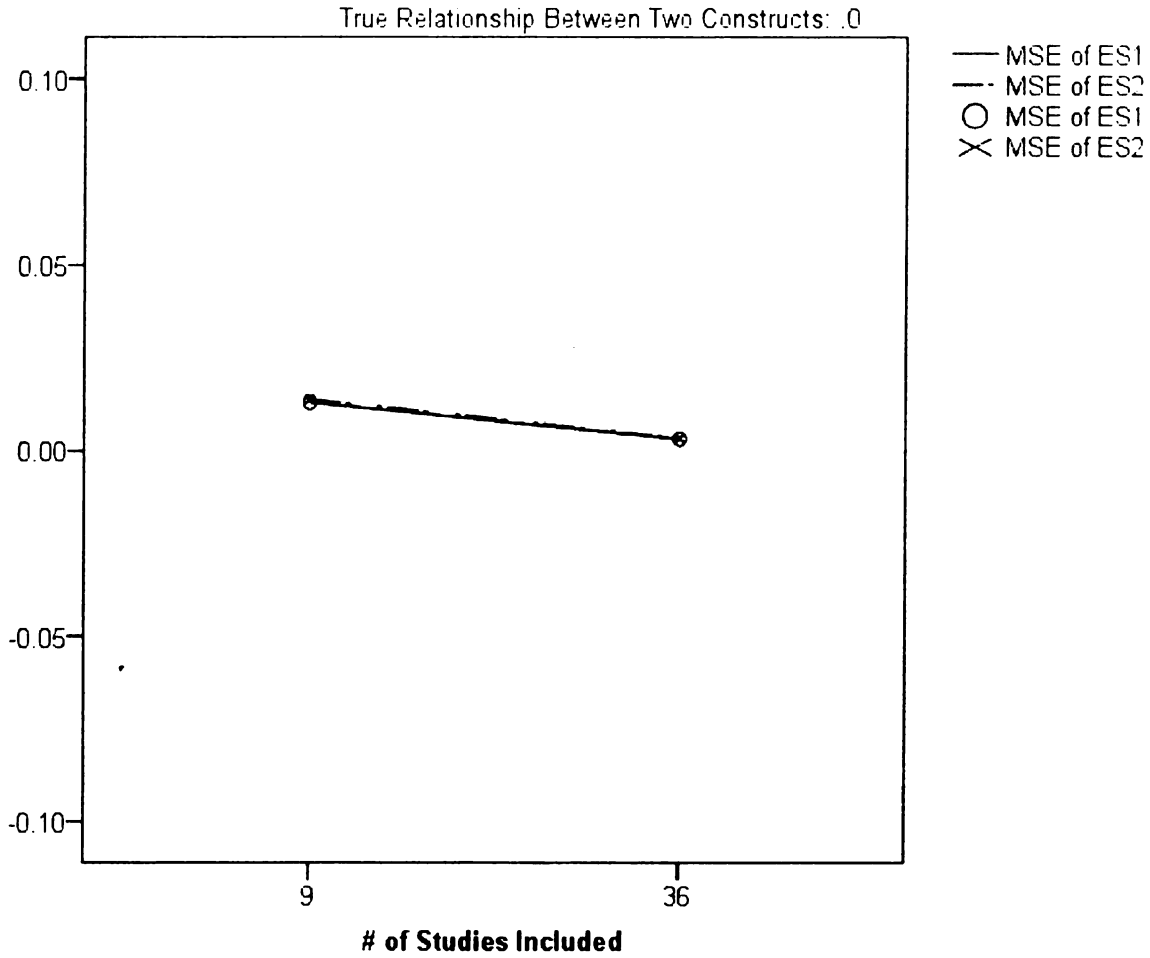


Figure 4.15

MSEs of two estimators depending on k when γ is set to 0

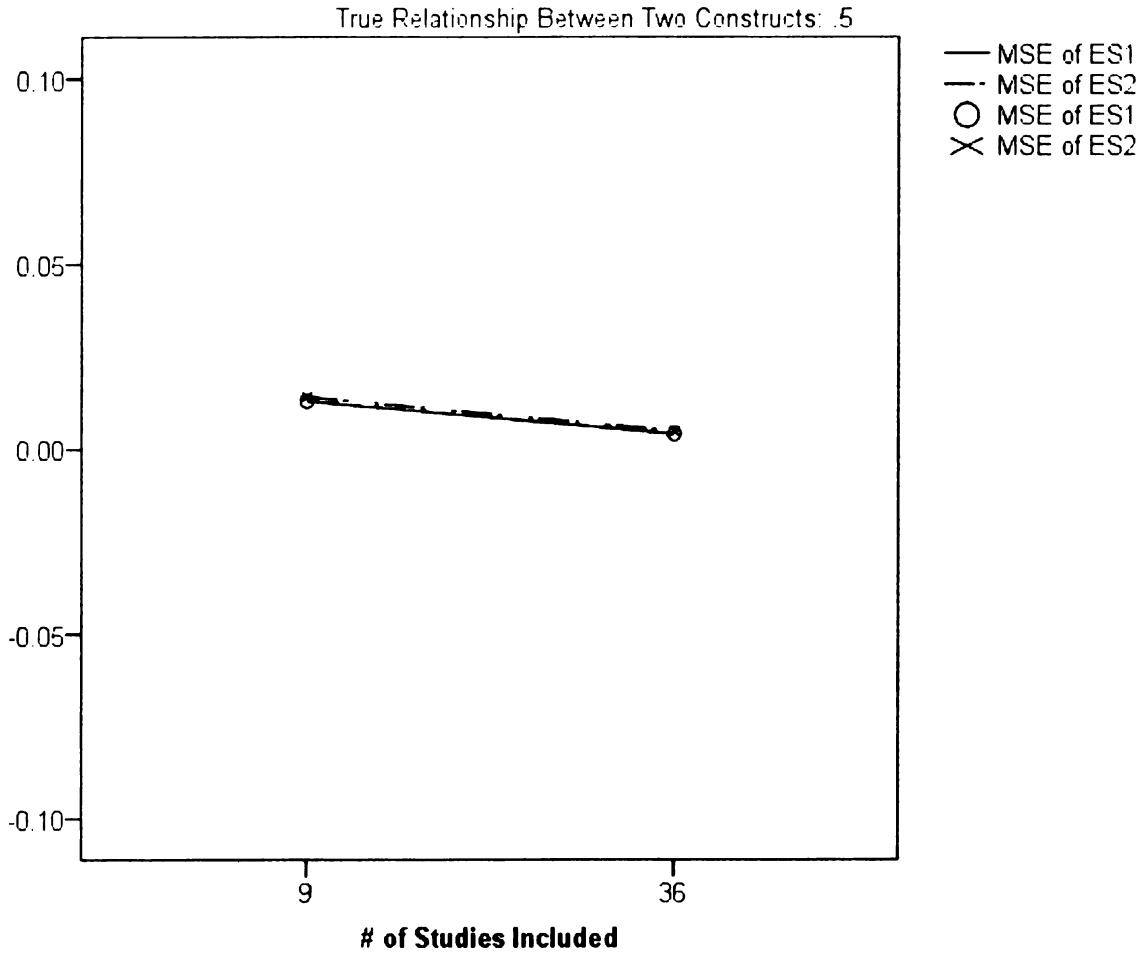


Figure 4.16

MSEs of two estimators depending on k when γ is set to .5

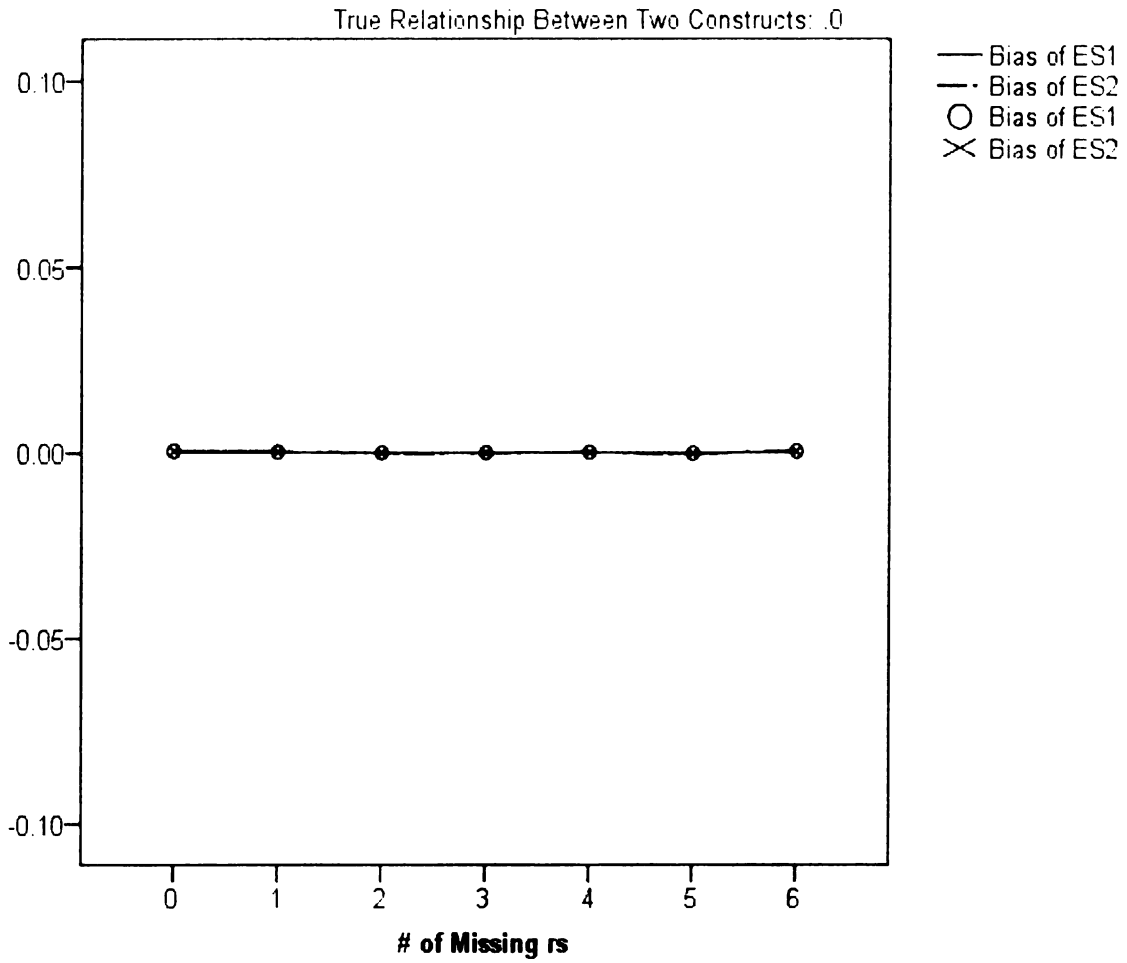


Figure 4.17

Biases of two estimators depending on the number of missing *r*s when γ is set to 0

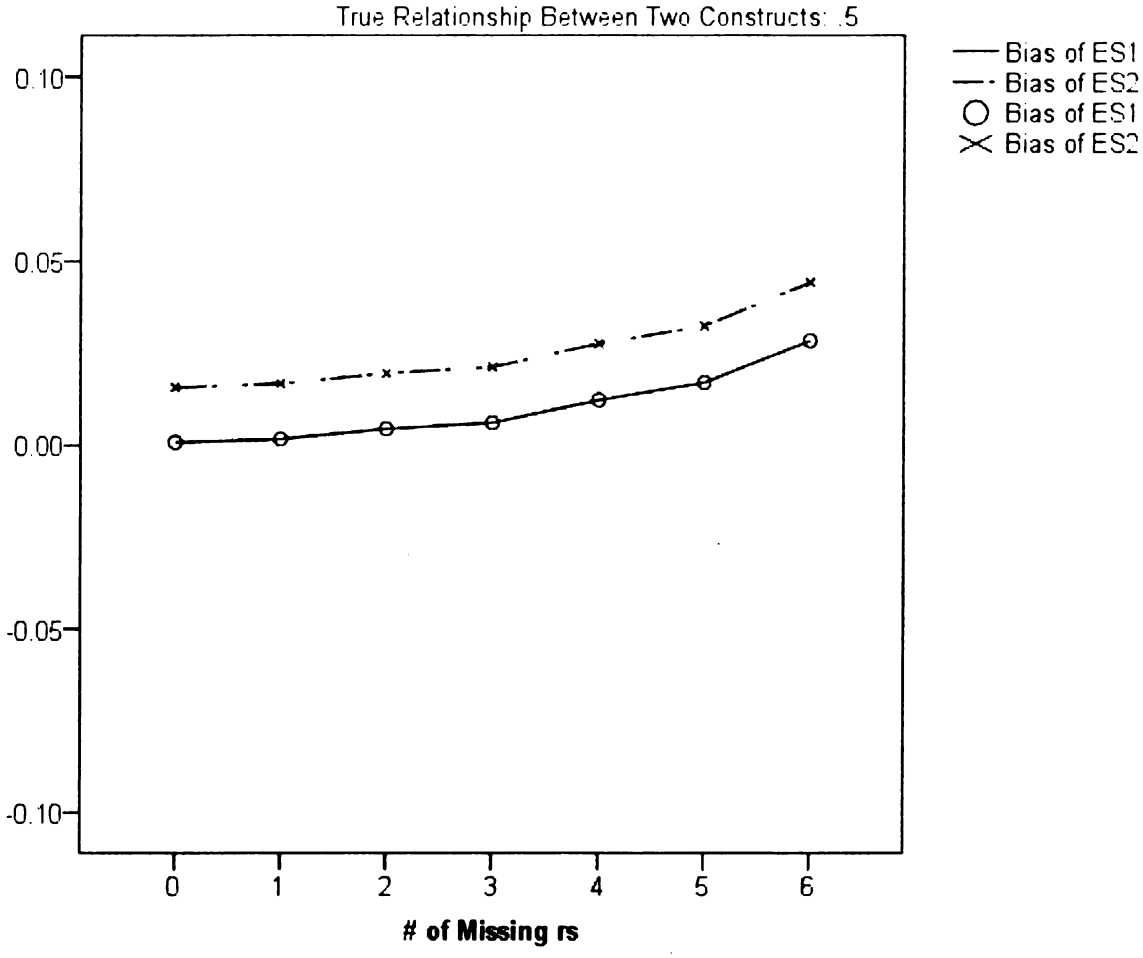


Figure 4.18

Biases of two estimators depending on the number of missing *r*s when γ is set to .5

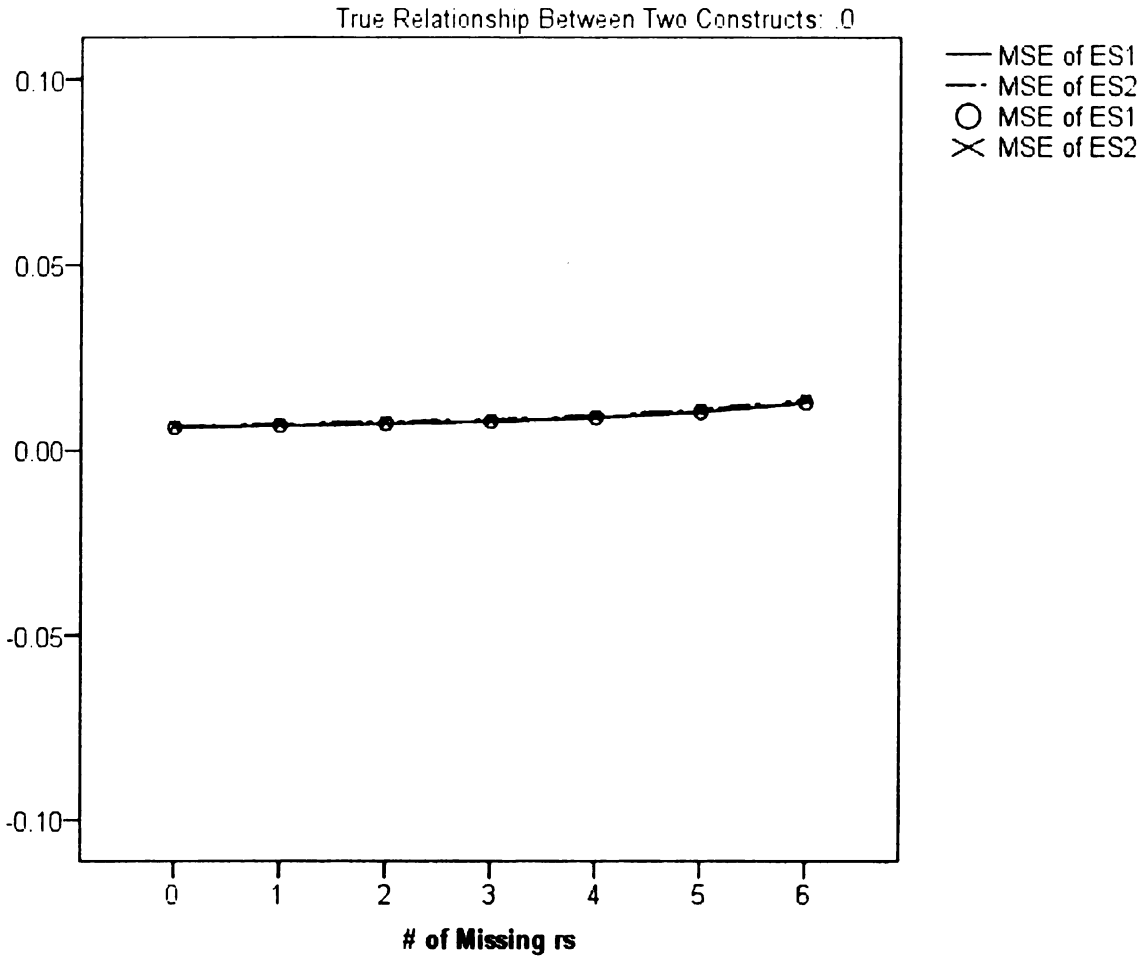


Figure 4.19

MSEs of two estimators depending on the number of missing *r*s when γ is set to 0

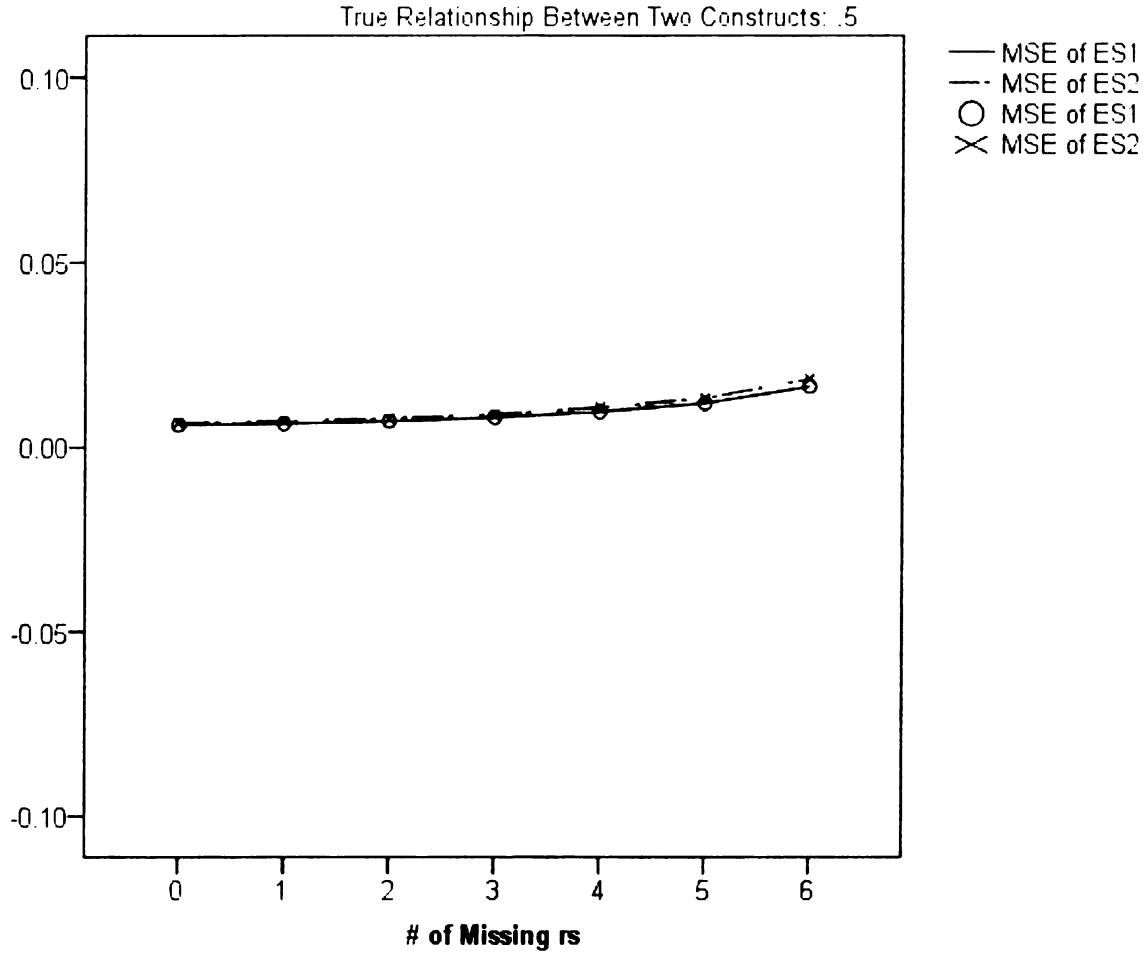


Figure 4.20

MSEs of two estimators depending on the number of missing *r*s when γ is set to .5

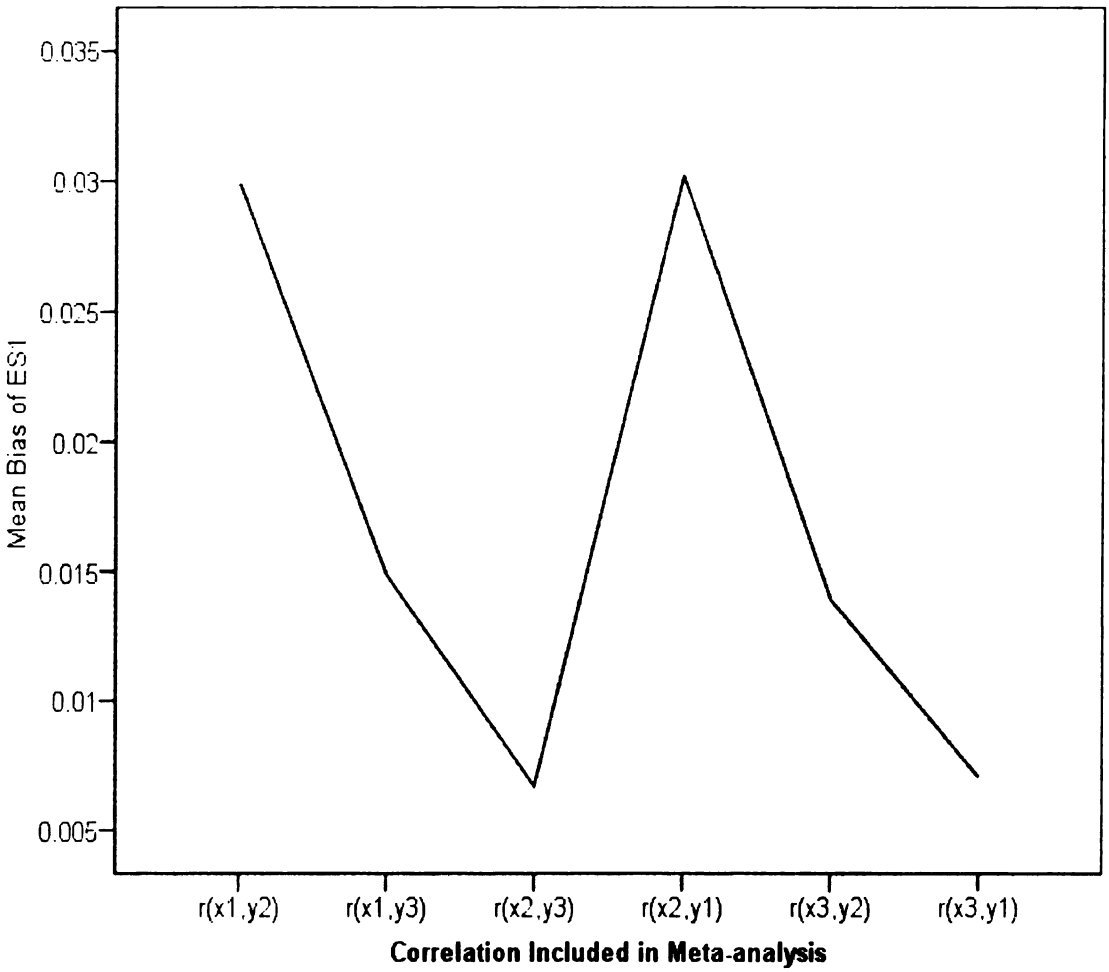


Figure 4.21

Bias of ES1 depending on which correlation is included with γ of .5

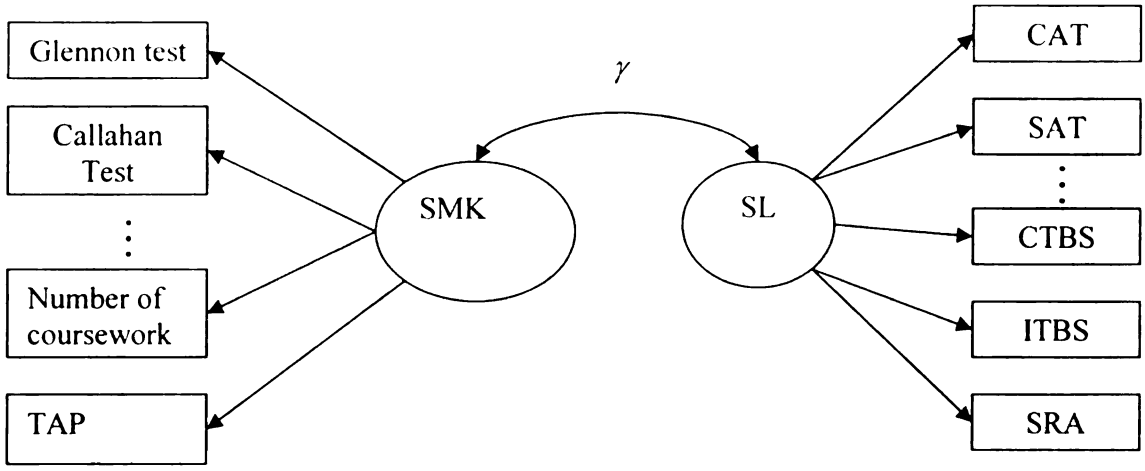


Figure 5.1.

A model for meta-analysis investigating teachers' subject matter knowledge (SMK) and student learning in mathematics

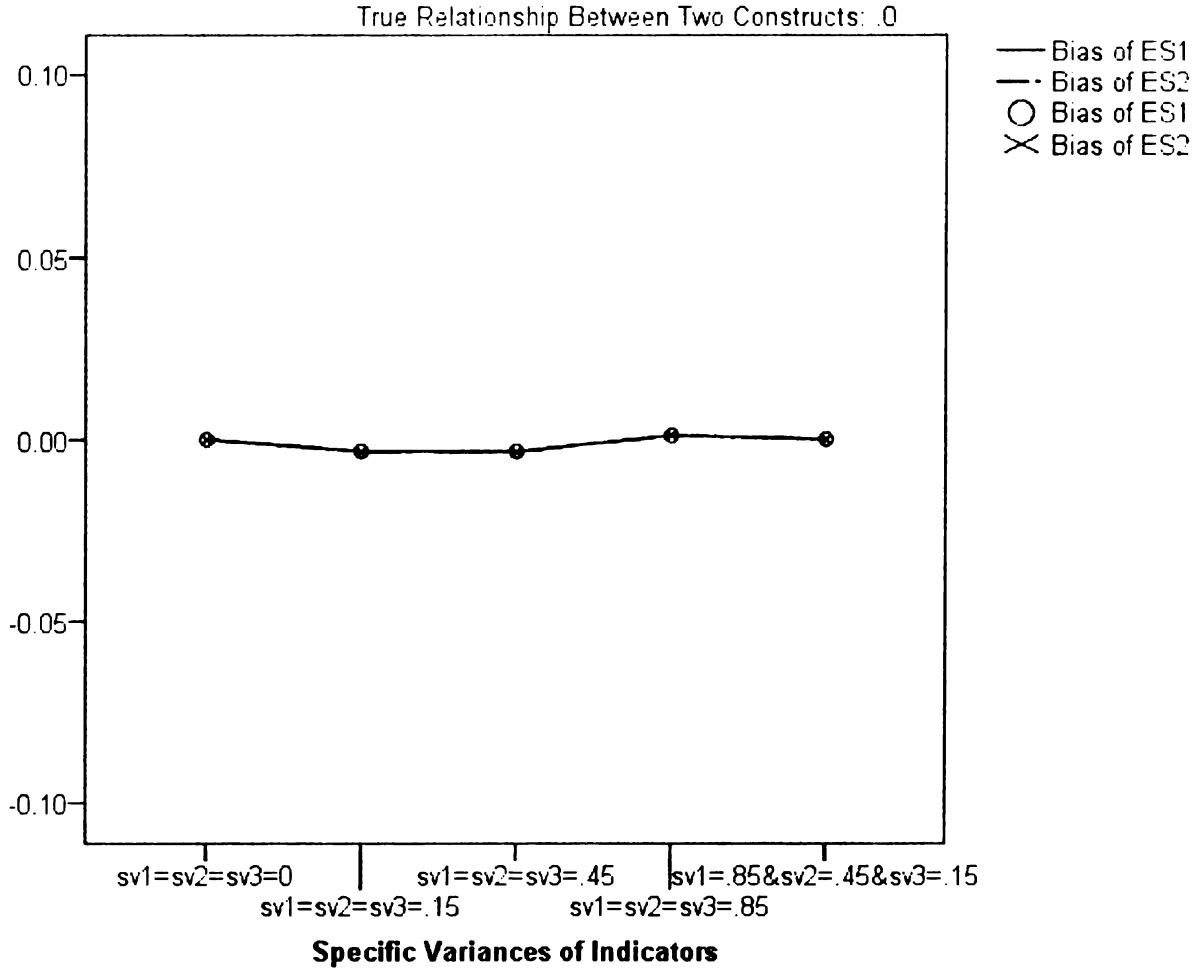


Figure 6.1

Biases of two estimators depending on specific variances of indicators when γ is set to 0

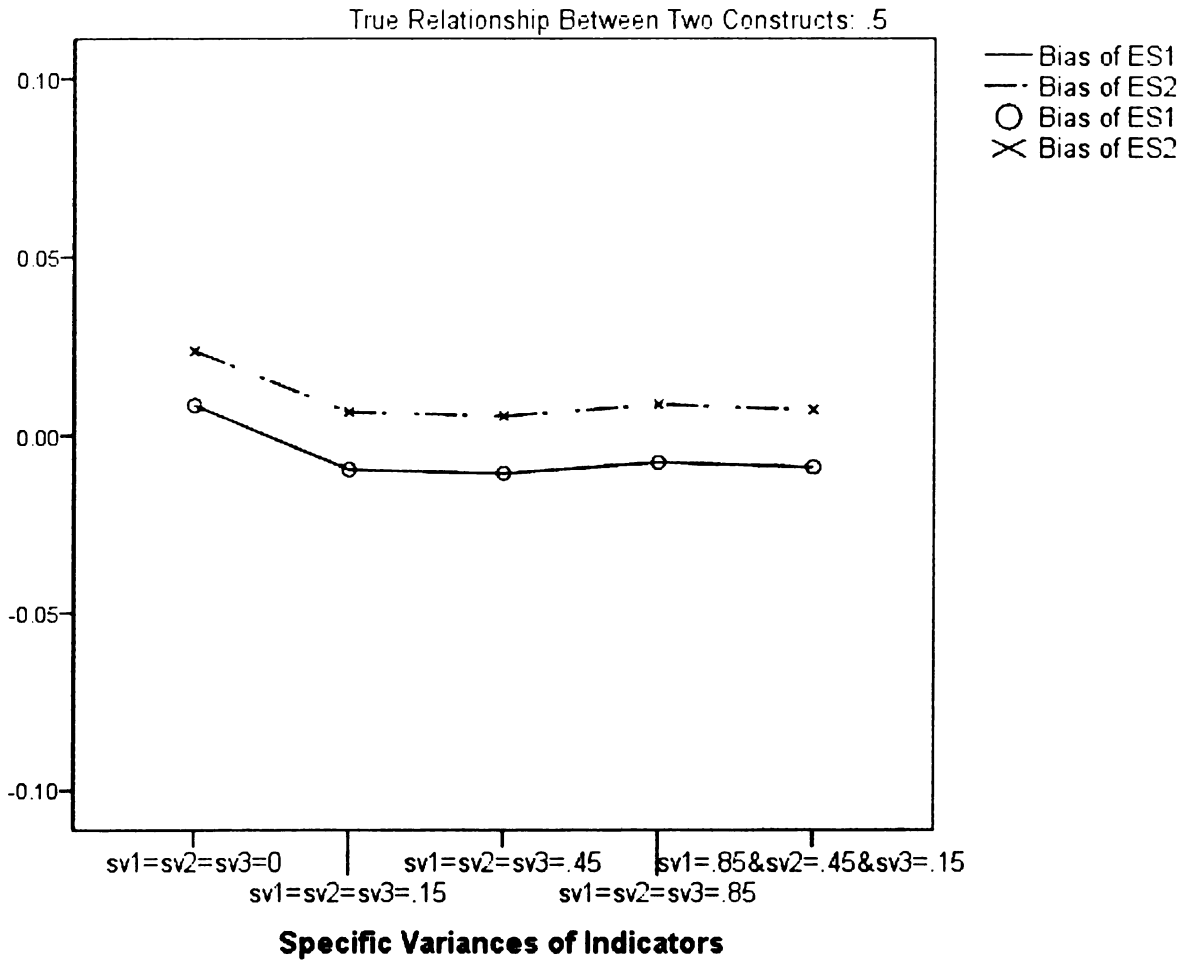


Figure 6.2

Biases of two estimators depending on specific variances of indicators when γ is set to .5

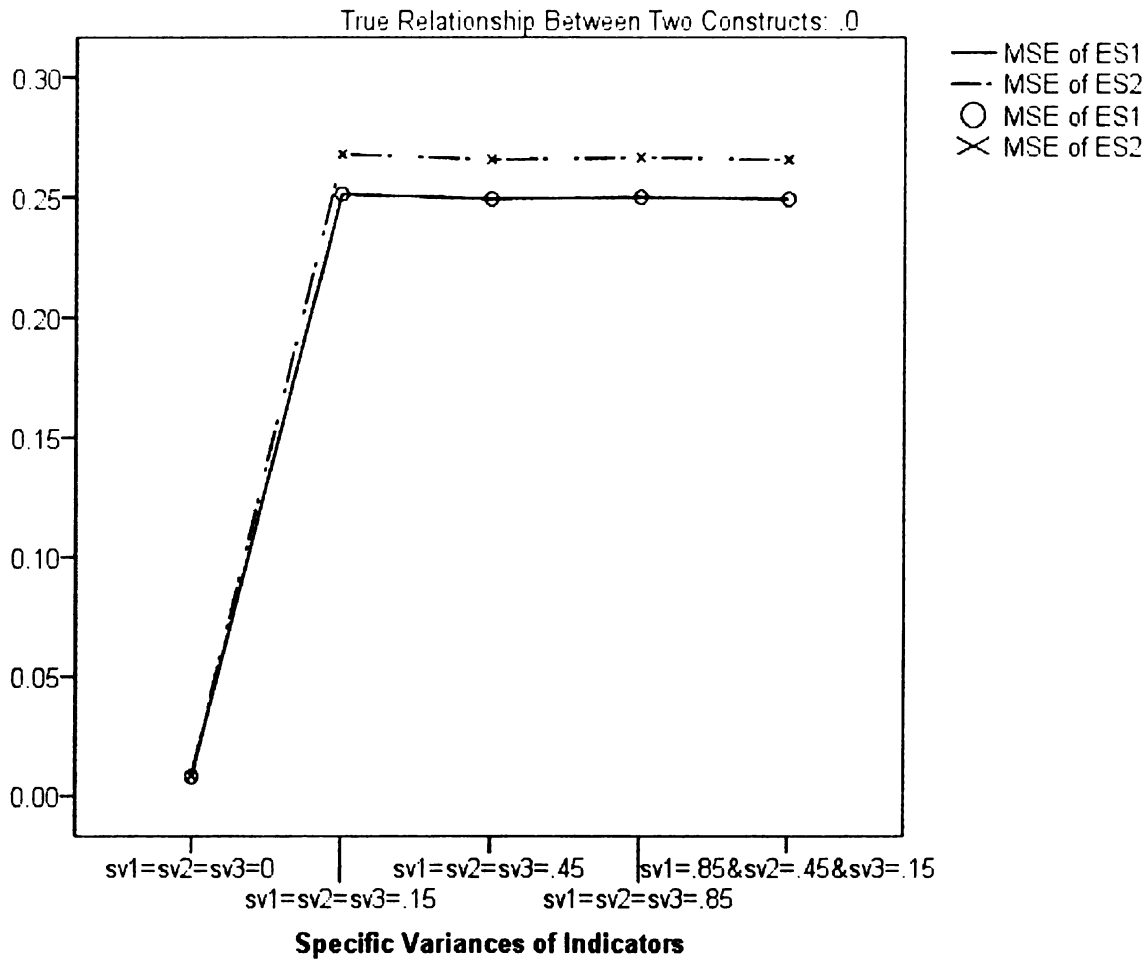


Figure 6.3

MSEs of two estimators depending on specific variances of indicators when γ is set to 0

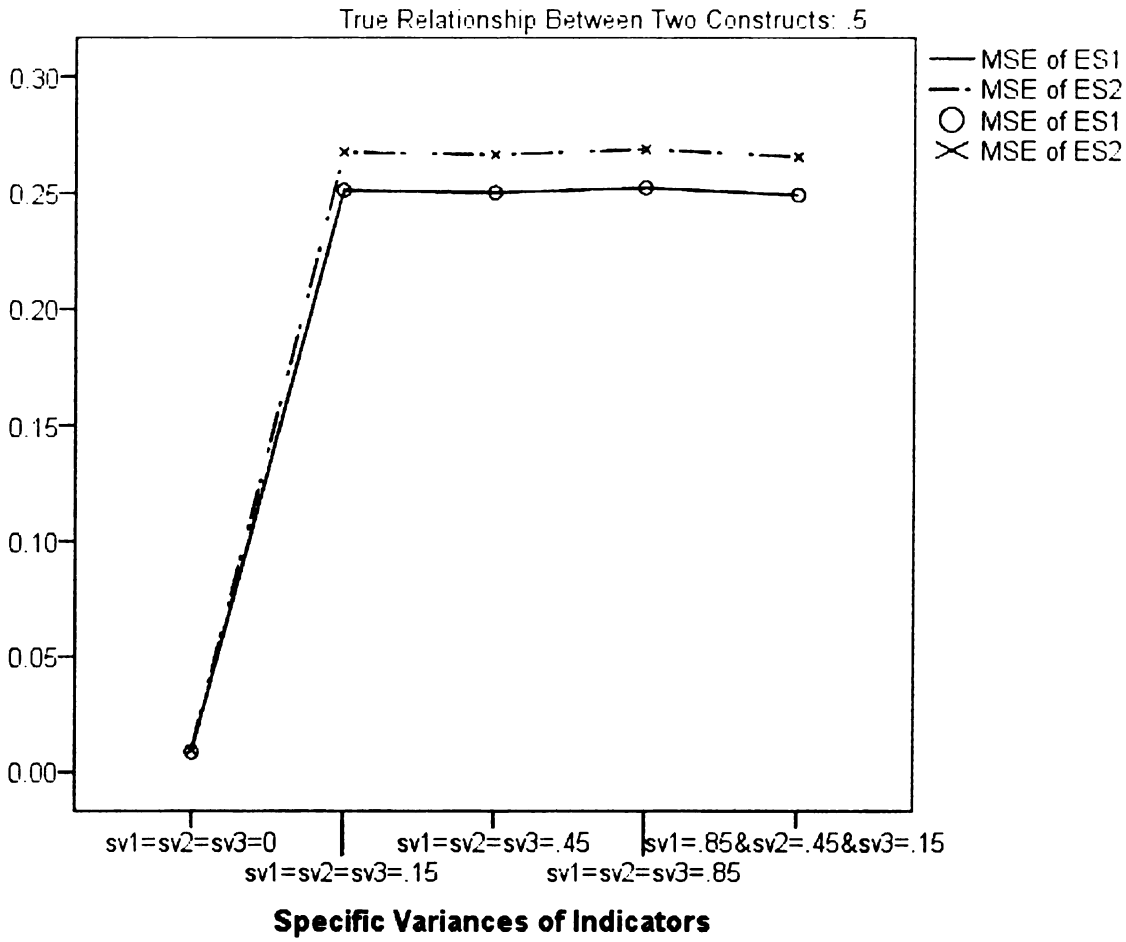


Figure 6.4

MSEs of two estimators depending on specific variances of indicators when γ is set to .5

BIBLIOGRAPHY

BIBLIOGRAPHY

Ahn, S., & Choi, J. (2004). *Teachers' subject matter knowledge as a teacher qualification: A synthesis of the quantitative literature on students' mathematics achievement.* Paper presented at the annual meeting of American Educational Research Association, San Diego, CA.

Ahn, S. & Becker, B. J. (2005, February). *Incorporating quality scores in a meta-analysis: A simulation study.* Poster presented at the 5th Annual Campbell Collaboration Colloquium, Lisbon, Portugal.

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey: Brooks/Cole. 40-48.

Ball, D. L. (1990a). The mathematical understandings that prospective teachers bring to teacher education. *Elementary School Journal, 90*, 449-466.

Ball, D. L. (1990b). Prospective elementary and secondary teachers' understanding of division. *Journal for Research in Mathematics Education, 21*(2), 132-144.

Baugh, F. (2002). Correcting effect sizes for score reliability: A reminder that measurement and substantive issues are lined inextricably. *Educational and Psychological Measurement, 62*(2), 254-263.

Becker, B. J. (1992). Using results from replicated studies to estimate linear models. *Journal of Educational Statistics, 17,* 341-362.

Becker, B. J. (1997). Meta-analysis and models of substance abuse prevention. *NIDA Research Monograph, 170*, 96-119.

Becker, B. J. (2001). Examining theoretical models through research synthesis: the benefits of model-driven meta-analysis. *Evaluation & The Health Professions, 24*(2), 190-217.

Becker, B. J., & Schram, C. M. (1994). Examining explanatory models through research synthesis. In H. M. Cooper & L. V. Hedges (Eds). *The handbook of research synthesis* (pp. 357 - 381). New York: Russell Sage Foundation.

Becker, B. J., & Wang, Q. (2006, April 8). *Study indices based on slopes.* Paper presented at the annual meeting of American Educational Research Association, San Francisco, CA.

Becker, B. J., & Wu, M. (2007). The synthesis of regression slopes in meta-analysis. *Statistical Science, 22*(3), 414-429.

Becker, B. J., & Fahrbach, K. (1994, April). *A comparison of approaches to the synthesis of correlation matrices*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Berk, R. (2006). Statistical inference and meta-analysis. Retrieved paper on May, 2006 from http://repositories.cdlib.org/uclastat/papers/2006051601

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley-Interscience.

Borman, G. D. (2002). Experiments for educational evaluation and improvement. *Peabody Journal of Education*, 77(4), 7-27.

Brown, M. A. (1988). *The relationship between levels of mathematics anxiety in elementary classroom teachers, selected teacher variables, and student achievement in grades two through six.* Unpublished doctoral dissertation, The American University.

Brown, S. P., & Peterson, R. A. (1993). Antecedents and consequences of salesperson job satisfaction: Meta-analysis and assessment of causal effects. *Journal of Marketing Research, 30*, 63-77.

Casella, G., & Berger, R. L. (1990). *Statistical inference*. Duxbury Press, Belmont, CA.

Chalmers, I., Hedges, L. V., & Cooper, H. (2002). A brief history of research synthesis. *Evaluation & the Health Professions, 25*(1), 12-37.

Chaney, B. (1995). *Student outcomes and the professional preparation of eighth grade teachers in science and mathematics: NSF/NELS:88 Teacher transcript analysis (regression).* Rockford MD: Westat, Inc.

Chiang, F. -S. (1996). *Ability, motivation, and performance: A quantitative study of teacher effects on student mathematics achievement using NELS:88 Data.* Unpublished doctoral dissertation, University of Michigan, Ann Arbor.

Choi, J., & Ahn, S. (2003) *Measuring teachers' subject-matter knowledge.* Presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Choi, J., Ahn, S., & Kennedy, M. (Under review). Role of teacher's subject matter knowledge. Manuscript submitted to the Journal on June, 2007

Cheung, M. W. -L. & Chan, W. (2005). Meta-analytic structural equation modeling: A two-stage approach. *Psychological Methods, 10*(1), 40-64.

Cheung, S. F. (2000). Examining solutions to two practical issues in meta-analysis: Dependent correlations and missing data in correlation matrices. *Dissertation Abstracts International,61*, 8B.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* New York: Holt.

Crowl, A., Ahn, S., & Baker, J. (In press). A meta-analysis of developmental outcomes for children of same sex and heterosexual parents. *Journal of GLBT Family Studies.*

Dominici, F., Parmigiani, G., Reckhow, K. H., & Wolpert, R. L. (1997). Combining information from related regressions. *Journal of Agricultural, Biological, and Environmental Statistics, 2*(3), 313-332.

Farley, J. U., Lehman, D. R., & Ryan, M. J. (1981). Generalizing from "imperfect" replication. *The Journal of Business, 54*(4), 597-610.

Field, A. P. (2001). Meta-analysis of correlation coefficients: A monte carlo comparison of fixed- and random-effects methods, *Psychological Methods*, *6*(2), 161-180.

Furlow, C. F. (2003). *Meta-analytic methods of pooling correlation matrices for structural equation modeling under different patterns of missing data.* Unpublished doctoral dissertation, University of Texas, Austin.

Furlow, C. F., & Beretvas, S. N. (2005). Meta-analytic methods of pooling correlation matrices for structural equation modeling under different patterns of missing data. *Psychological Methods, 10*(2), 227-254.

Gelman, A. (1995). Methods of moments using monte carlo simulation. *Journal of Computational and Graphical Statistics, 4*(1), 36-54.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*, 3-8.

Gleser, L. J., & Olkin, I. (1994). Stochastically dependent effect sizes. In H. M. Cooper & L. V. Hedges (Eds). *The handbook of research synthesis* (pp. 339 - 356). New York: Russell Sage Foundation.

Greene, W. H. (1997). *Econometric analysis.* New York, N.Y.: Macmillian.

Hancock, G. R. (1997). Disattenuated for score reliability: A structural equation modeling approach. *Educational and Psychological Measurement, 57*(4), 598-606.

Hedges, L. V. (1983). Combining independent estimators in research synthesis. *The British Journal of Mathematical and Statistical Psychology, 36*, 123-131.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando: Academic Press.

Hill, H., Rowan, R., & Ball, D. (2005) Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42 (2), 371-406.

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal, 105*(1), 11-30.

Hom, P. W., Caranikas-Walker, F., Prussia, G. E., & Griffeth, R. W. (1992). A meta analytical structural equations analysis in a model of employee turnover. *Journal of Applied Psychology, 77,* 890-909.

Hunter, J.E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies*. Beverly Hills, CA: Sage Publication.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. New York: Russell Sage Foundation.

Hunter, J. E., & Schmidt, F. L. (1994). Correcting for sources of artificial variation across studies. In H. M. Cooper & L. V. Hedges (Eds). *The handbook of research synthesis* (pp. 323 - 336). New York: Russell Sage Foundation.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings (2nd ed.)*. New York: Russell Sage Foundation.

Kennedy, M. (2007). *Finding cause in a self-propelled world.* Unpublished manuscript. Retrieved on August, 2007 from http://www.msu.edu/user/mkennedy/TQQT/.

Kennedy, M., Ahn, S., & Choi, J. (2008). The value added by teacher education. In M. Cochran-Smith, S. Feiman-Nemser & J. McIntyre (Eds.), *Handbook of research on teacher education: Enduring issues in changing contexts* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Lambert, R. G., & Curlette, W. L. (1995, April). *The robustness of the standard error of summarized, corrected validity coefficients to non-independence and non-normality of primary data.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Le, H. A. (2003). *Correcting for indirect range restriction in meta-analysis: Testing a new meta-analytic method.* Unpublished doctoral dissertation, University of Iowa, Iowa City, IA.

Le, H., & Schmidt, F.L. (2006) Correcting for indirect range restriction in meta-analysis:

Testing a new meta-analytic procedure. *Psychological Methods, 11*, 416-438.

Lipsey, M., & Wilson, D. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

Maydeu-Olivares, A., & Bockenholt, U. (2005). Structural equation modeling of paired comparison and ranking data. *Psychological Methods, 10*(3), 285-304.

Messick, S. (1993). Validity. In Linn, R. L. (Ed.). *Educational measurement* (pp. 13-104). Phoenix: The Oryx Press.

Nugent, W. R. (2006). The comparability of the standardized mean difference effect size across different measures of the same construct: Measurement considerations. *Psychological Measurement, 66*(4), 612-623.

National Council of Teachers of Mathematics. (n.d.). Retrieved January 8, 2008, from http://standards.nctm.org/document/chapter6/index.htm.

Olkin, I., & Siotani, M. (1976). Asymptotic distribution of functions of a correlation matrix. In S. Ikeda (Ed.), *Essays in probability and statistics* (pp. 235-251). Tokyo: Shinko Tsusho.

Oswald, F. L., & Converse, P. D. (2005). *Correcting for reliability and range-restriction in meta-analysis.* Paper presented at the 20th annual conference of the Society for Industrial and Organizational Psychology, Los Angeles, CA.

Oswald, F. L., & Johnson, J. W. (1998). On the robustness, bias, and stability of statistics from meta-analysis of correlation coefficients: Some initial monte carlo findings. *Journal of Applied Psychology, 83*(2), 164-178.

Peterson, R. A., & Brown, S. P. (2005). On the use of beta coefficients in meta-analysis. *Journal of Applied Psychology, 90*(1), 175-181.

Premack, S. L., & Hunter, J. E. (1988). Individual unionization decisions. *Psychological Bulletin, 97,* 274-285.

Qu, Y., & Becker, B. J. (2003). *Does traditional teacher certification imply quality? A meta-analysis.* Paper presented at the annual meeting of American Educational Research Association, Chicago, IL.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Raju, N. S., Anselmi, T. V., Goodman, J. S., & Thomas, A. (1998). The effect of correlated artifacts and true validity on the accuracy of parameter estimation in validity generalization. *Personnel Psychology, 51*, 453-465.

Raju, N. S., & Brand, P. A. (2003). Determining the significance of correlations corrected for unreliability and range restriction. *Applied Psychological Measurement, 27,* 52-71.

Raju, N. S., Burke, M. J., Normand, J., & Langlois, G. M. (1991). A new meta-analytic approach. *Journal of Applied Psychology, 76*(3), 432-446.

Raju, N. S., Fralicx, R., & Steinhaus, S. D. (1986). Covariance and regression slope models for studying validity generalization. *Applied Psychological Measurement, 10*(2), 195-211.

Raudenbush, S. W., Becker, B. J., & Kalaian, K. (1988). Modeling multivariate effect sizes. *Psychological Bulletin, 103*(1), 111-120.

Rubin, D. B. (1992). Meta-analysis: Literature synthesis or effect-size surface estimation? *Journal of Educational Statistics, 17*(4), 363-374.

Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology, 85,* 112-118.

Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology, 71(3),* 432-439.

Shadish, W. R., & Haddock, C. K. (1994). Combining estimates of effect size. In H. M. Cooper & L. V. Hedges (Eds). *The handbook of research synthesis* (pp. 261 - 282). New York: Russell Sage Foundation.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14.

Slavin, R. E. (1984). Meta-analysis in education: How has it been used? *Educational Researcher, 13*(8), 6-15.

Slavin, R. E. (2008). What works? Issues in synthesizing educational program evaluations. *Educational Researcher, 37*(1), 5-14.

Teddlie, C., Falk, W., & Falkowski, C. (1983, April). *The contribution of principal and teacher inputs to student achievement.* Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada.

Thum, Y. M., & Ahn, S. (2007). *Challenges of meta-analysis from the standpoint of a Latent variable framework.* Paper presented at the 7th Campbell collaboration colloquium, London, England.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 79*, 281–299.

Timm, N. H. (2004). Estimating effect sizes in exploratory experimental studies when using a linear model. *The American Statistician, 58*(3), 213-217.

Towne, L., Wise, L. L., & Winters, T. M. (Eds.). (2005). *Advancing scientific research in education.* Washington, DC: National Academic Press. Available from http://www.NAP.edu.

Vanhonacker, W. R., Lehman, D. R., & Sultan, F. (1990). Combining related and sparse data in linear regression models. *Journal of Business & Economic Statistics, 8*(3), 327-335.

Whiteside, M. F., & Becker, B. J. (2000). Parental factors and the young child's postdivorce adjustment: a meta-analysis with implications for parenting arrangements. *Journal of Family Psychology, 14*(1), 5-26.

Wu, M. (2006a, April). *Applications of Generalized Least Squares and Factored Likelihood in Synthesizing Regression Studies.* Paper presented at the American Educational Research Association, San Francisco, CA.

Wu, M. (2006b). *Methods of meta-analyzing regression studies: Applications of Generalized Least Squares and Factored Likelihoods.* Unpublished Doctoral dissertation, Michigan State University.