COMPARING META-ANALYTIC STRUCTURAL EQUATION MODELING APPROACHES ACROSS MODEL ASSUMPTIONS USING AN EMPIRICAL EXAMPLE

A Dissertation

by

ASHLEY HANNAH MAJZUN

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Christopher G. Thompson
Co-Chair of Committee,	Oi-man Kwok
Committee Members,	Wen Luo
	Lei-shih Chen
Head of Department,	Jeffrey Liew

May 2023

Major Subject: Educational Psychology

Copyright 2023 Ashley Majzun

ABSTRACT

Meta-analytic Structural Equation Modeling (MASEM) is the combination of meta-analysis (MA) and structural equation modeling (SEM). With new MASEM methodologies developed over the past few years, there is an opportunity to compare the past approaches with the new ones.

The purpose of this dissertation is two-fold. First, the parameter estimates, standard errors, confidence intervals, and heterogeneity measures are compared across 8 MASEM approaches (fixed-effect and random-effects univariate *r*, fixed-effect and random-effects univariate *z*, fixed-effect and random-effects Two-Stage SEM approach, and fixed-effect and random-effects One-Stage MASEM approach) using 25 studies relating to college persistence. Overall, results found only slight differences in estimates across methods (differences to two or three decimal places). The biggest difference was found in significant path estimates between univariate and multivariate approaches, which is primarily due to sample size differences.

The second purpose of this paper was to synthesize the current literature pertaining the relationships between student background characteristics, institutional characteristics, academic integration, and social integration on student success. Results indicate that student background characteristics, academic integration, and social integration had a direct significant impact on student success. Although institutional characteristics did not have a significant impact on student success directly, it had a significant impact on academic and social integration.

DEDICATION

To my mom, for always encouraging me to dream big.

To my husband, Zach, for helping me fulfill those big dreams.

ACKNOWLEDGEMENTS

I would like to first and foremost thank my chair, Dr. Chris Thompson, for being so supportive throughout this journey. You haven't stopped answering my questions since I first took your meta-analysis class in 2018, and I am grateful to have shared many thoughtful conversations about meta-analysis, MASEM, and mathematics/statistics in general over the years. I would also like to thank my co-chair, Dr. Oi-man Kwok, and committee members, Dr. Wen Luo and Dr. Lei-Shih Chen, for their advice, guidance, and mentorship since I first entered the program. I would like to specially thank Dr. Suzanne Jak for her help and advice with the ML MASEM (i.e., the fixed-effect OSMASEM) approach.

To my friends – Gaby, Fatma, Aditi, Geoff, Brandie, Lucy, and Maria – thank you for your friendship throughout the years and always bringing me joy.

To my family – mom, dad, sister Sarah, and cousin Jung– for all your support and humor throughout this journey.

To my husband, Zach, and children, Allison and Ethan - thank you so much for your patience, your attention when I talked endlessly about meta-analysis and SEM, and your support when I had to write. You all motivate me to work harder and make the world a better place.

Lastly, to all the students I have ever taught, tutored, and mentored – thank you for always inspiring me and reminding me of the excitement to learn.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Dr. Christopher G. Thompson (co-chair), Dr. Oi-Man Kwok (co-chair), and Dr. Wen Luo of the Department of Educational Psychology as well as to Dr. Lei-Shih Chen of the Department of Health and Kinesiology. Dr. Suzanne Jak of the University of Amsterdam assisted with modified R syntax for the ML MASEM approach. All other work for the dissertation was completed independently by the student.

Funding Sources

There are no outside funding contributions to acknowledge related to this dissertation.

NOMENCLATURE

k	study identification index
K	total number of studies with $k = 1 \ldots, K \geq 2$
T_k	kth general effect size
v_k	variance of the kth effect size
θ	population effect size
σ_k^2	variance of the population effect size
e_k	univariate within-studies error
$\hat{\theta}$	estimated population effect size
$\widehat{w}_k$	weight for the fixed-effect model
$var_{\hat{\theta}}$	variance of the estimated population effect size
μ	mean of the population distribution of all effect sizes
u_k	between-studies error
τ^2	between-studies variance
$\hat{\tau}^2$	estimated between-study variation
$\hat{\mu}$	estimated mean effect size
$\widehat{w}_k^*$	weight for the random-effects model
p	number of effect sizes that are of interest in a multivariate meta-analysis
p_i	number of observed effect sizes in the ith study
$\boldsymbol{y}_i$	$p_i \times 1$ matrix of observed effect sizes
$\boldsymbol{X}_i$	$p_i \times p$ design matrix with 0's and 1's to select the observed effect size

$\boldsymbol{f}_i$ $p \times 1$ vector of population effect sizes

$\boldsymbol{e}_i$ $p_i \times 1$ vector of sampling errors

$\boldsymbol{V}_i$ known covariance matrix of $\boldsymbol{e}_i$

β_F population "true" effect for a multivariate meta-analysis

F_{GLS} function that is minimized using generalized least squares to determine the fixed-effect size

$\hat{\beta}_F$ estimated overall effect size in a multivariate fixed-effect model

$\widehat{\Omega}_F$ estimated asymptotic covariance matrix in a multivariate fixed-effect model

β_R average population effect size under a random-effects model in multivariate analysis

$\boldsymbol{Z}$ a matrix composed of 0's and 1's to select the random effects in multivariate analysis

u_i stacked random effects (between-studies error) for all studies in multivariate analysis

$\hat{\beta}_R$ multivariate random effects estimated overall effect

$\widetilde{\boldsymbol{V}}$ asymptotic sampling covariance matrix for random-effects multivariate analysis

$\widehat{\Omega}_R$ estimated asymptotic sampling covariance matrix for random-effects multivariate effect size

Q the weighted sum of squared deviations between individual studies and fixed effect

Q_{UNI} Q-statistic for the univariate case

w_i See definition for $\widehat{w}_k$

Q_{MUL} Q-statistic for the multivariate case

I^2 ratio of true heterogeneity to total variances across observed effect estimates

τ^2 true variance of effect sized (between-studies variance)

$\hat{\tau}^2$ estimated variance of the true effects (estimated between-studies variance)

T^2 see definition of $\hat{\tau}^2$

T see definition of $\hat{\tau}$

$\hat{\tau}$ estimated standard deviation of the true effects

$\tilde{v}$ typical within-study sampling variance when calculating I^2

$I^2_{Q(MUL)}$ multivariate I^2 index

df_{MUL} degrees of freedom for multivariate test of homogeneity

W_i study weight; see definition for $\widehat{w}_k$

$F_{ML}(\theta)$ maximum likelihood discrepancy function

n_i^{MG} sample size reported in the ith group for multi-group SEM

$F_i(\theta)$ fit function for the ith group

r sample correlation coefficient

ρ population correlation coefficient

z Fisher's z value

I total number of observed variables

Y_i ith observed variable in a study with $i = 1, \ldots, I$

n_i sample size for the ith study

r_{ist} sample correlation between variables Y_s and Y_t in the ith study

ρ_{st} population correlation between variables Y_s and Y_t

$\bar{r}_{FE\,st}$ overall mean of the correlations between Y_s and Y_t under fixed-effect model

w_{ist} inverse-variance weight for fixed-effect model with correlations

$s^2_{r\,ist}$ variance of r_{ist}

$\bar{r}_{RE\,st}$ overall mean of the correlations between Y_s and Y_t under random-effects model

w^*_{ist} weight for random-effects model with correlations

$\hat{\tau}^2_{st}$ estimated between studies variance for the correlation between Y_s and Y_t

z_{ist} the Fisher's z value between Y_s and Y_t for the ith study

$s^2_{z_{ist}}$ variance of z_{ist}

$\bar{z}_{FE_{st}}$ fixed effect weighted overall mean Fisher's z

w^+_{ist} weight for fixed-effect Fisher's z model

$\bar{z}_{RE_{st}}$ random effects weighted overall mean Fisher's z

w^{++}_{ist} weight for random-effects Fisher's z model

p^g number of observed variables in the gth study

$\mathbf{\Sigma}^g$ covariance matrix for the gth study

$\boldsymbol{D}^g$ $p^g \times p^g$ diagonal matrix of standard deviations in the gth study

$\boldsymbol{P}^g$ $p^g \times p^g$ correlation matrix in the gth study

I identity matrix

$\mathbf{\Lambda}^g$ factor loadings for the gth study

$\mathbf{\Phi}^g$ factor covariance for the gth study

$\mathbf{\Psi}^g$ error variance matrix

$\widehat{\boldsymbol{P}}$ estimated pooled correlation matrix using the fixed-effect TSSEM approach

$\widehat{\boldsymbol{V}}$ asymptotic covariance matrix of parameter estimates for the TSSEM approaches

N total sample size

n^g sample size for the gth study

$\widehat{\boldsymbol{p}}_R$ pooled average correlation matrix for the random-effects TSSEM approach

$\widehat{\boldsymbol{V}}_R$ asymptotic sampling covariance matrix for the random-effects TSSEM approach

$F(\boldsymbol{\gamma})$ the discrepancy function for stage 2 of the TSSEM approaches

$\boldsymbol{r}^*$ $p' \times 1$ vector of elements from the sample correlation matrix

$\boldsymbol{\rho}(\boldsymbol{\gamma})$ $p' \times 1$ vector of elements from the pooled correlation matrix (under fixed or random effects)

p_g see definition for p^g

q total number of variables

$\boldsymbol{D}_g$ see definition for $\boldsymbol{D}^g$

$\boldsymbol{P}_g$ see definition for $\boldsymbol{P}^g$

$\boldsymbol{M}_g$ $p_g \times q$ selection matrix composed of 0's and 1's

$\boldsymbol{P}_{model}$ model implied correlation matrix

$\boldsymbol{F}$ selection matrix composed of 1's for observed variables and 0's for latent variables

$\boldsymbol{I}$ Identity matrix

$\boldsymbol{S}$ symmetric matrix with variances and covariances in random-effects OSMASEM path model

$\boldsymbol{A}$ symmetric matrix with asymmetric paths used in random-effects OSMASEM path model

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

Two popular quantitative methods in applied research are meta-analysis and structural equation modeling. Structural equation modeling (SEM) aims to test a hypothesized causal model by fitting a model to a single data set and then determining how well the model fits the data. Meta-analysis (MA) is a method of combining quantitative results (namely effect sizes) across many studies. Both quantitative methods rely heavily on theoretical justification and are used in multiple disciplines (Borenstein et al, 2021; Kline, 2016). Since the 1970s, publications using either method has grown from virtually nonexistent to thousands each year (Cheung, 2015).

Although there has been a rise in popularity of structural equation modeling in recent years, researchers have been interested in synthesizing SEM studies since the 1980's (Brown & Peterson, 1993; Hom et al, 1992; Premack & Hunter, 1988; Wagner, 1988). As fitting a SEM often involves a correlation or covariance matrix, early works focused on how to combine correlation matrices (Becker, 1992; Becker, 1996; Hedges & Olkin, 1985; Hunter & Schmidt, 1982; Viswesvaran & Ones, 1995). Research has since expanded to more advanced methodological issues such as missing data, fixed-effect/random-effects models, univariate/multivariate models, and small sample sizes (Cheung & Chan, 2005; Cheung & Cheung, 2014; Cho, 2015; Furlow & Beretvas, 2005; Jak & Cheung, 2020; Sheng et al, 2016; Yuan, 2016; Zhang, 2011).

Cheung (2009) described three main advantages of MASEM: to address the generalizability of findings across settings, to identify potential moderators that influence the structure of the model, and to obtain more precise estimates by increasing sample size. Because of the ability to comprehensively synthesize research in addition to its flexibility as a methodology, researchers across disciplines where SEMs are of interest will also find MASEM of interest as well. In this dissertation, the aim is to add to the current body of MASEM literature while concurrently applying by comparing eight MASEM approaches using a data set related to two-year college students and their success based on social and academic integration, student background characteristics, and institutional characteristics.

MASEM Framework

There are three important factors to consider when determining which MASEM approach is most appropriate for analysis. Should a parameter-based approach be used or correlation-based approach? Are the effect sizes assumed to originate from a single population (fixed effect), or is there an expectation that the effect sizes originate from multiple populations (random effects)? Lastly, is it assumed that the effect sizes are independent or dependent of one another?

There are generally two main approaches to synthesizing SEMs: the parameter-based approach and correlation-based approach. Parameter-based MASEM aims to combine parameter estimates (e.g., factor loadings or path coefficients) across studies (See Figure 1). Cheung & Cheung (2016) commented that one of the main strengths of parameter-based MASEM is that the heterogeneity of parameter estimates (instead of correlation matrices in the correlation-based approach) can be calculated. However, limitations of this approach include the inability to handle missing parameter estimates and its inappropriate use in the case of some over-identified models (Cheung & Cheung, 2016). The most notable parameter-based approaches include the full-information MASEM (Yu et al., 2016), parameter-based MASEM (Cheung & Cheung, 2016), and Bayesian MASEM (Ke, et al, 2018).

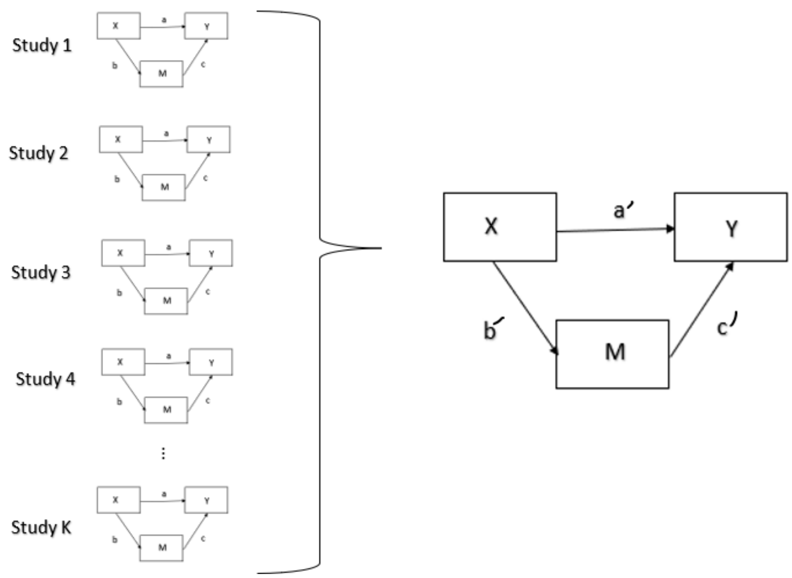

Figure 1: Illustration of Parameter-Based MASEM

In correlation-based MASEM, there are usually two stages. First, the homogeneity of correlation matrices is assessed and then correlations are combined or pooled. Then, the pooled correlation matrix is used to fit a SEM. The correlation-based approach has been the more popular approach as this method was developed sooner, can handle missing correlation coefficients and varying structural models can be tested and compared; however, one main limitation (like parameter-based MASEM) is the possible inappropriate use with over-identified models (Cheung & Cheung, 2016). The most well-known correlation-based approaches are the univariate-*r* approach (Viswesvaran & Ones, 1995), the generalized least squares (GLS) approach (Becker, 1992), and the two stage SEM approach (TSSEM) (Cheung, 2014). See Figure 2 for an illustration of correlation-based MASEM.

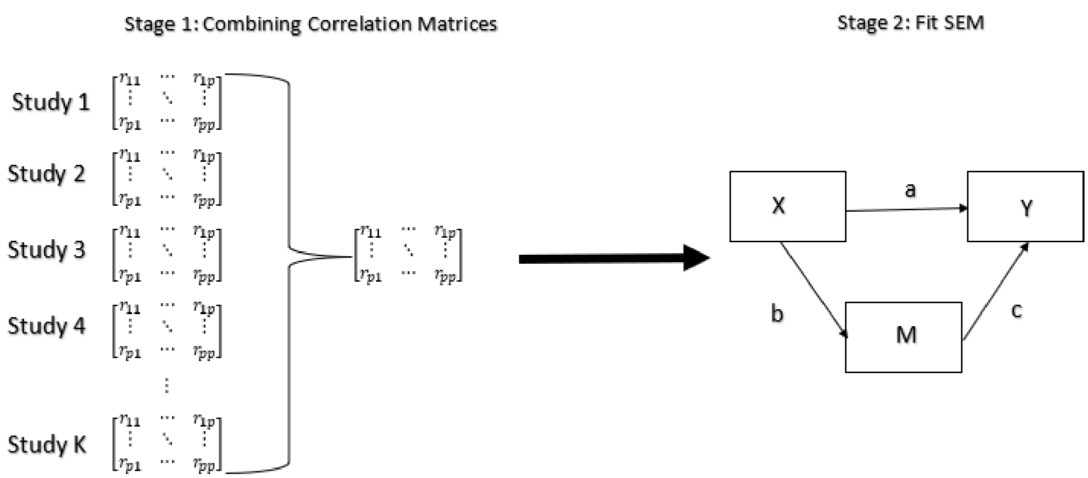

Figure 2: Illustration of Correlation-based MASEM

Two approaches that also use correlations that do not include two separate stages are the Maximum Likelihood MASEM approach (Oort & Jak, 2016) and the one-stage MASEM (OSMASEM) approach (Jak & Cheung, 2021). The former method is the fixed effect equivalent to the latter's random-effects method. In both approaches, correlations from individual studies are fitted directly to an SEM, without needing to pool correlation matrices first. Multigroup SEM is used to determine homogeneity.

In meta-analysis, it is generally assumed that effect sizes are either from the same population (fixed-effect model), or effect sizes originate from different populations (random-effects model). There are special cases of mixed-effect models, where the effect sizes are grouped into categories (random effects) then assumed to be fixed effect within each category as well. In MASEM, this is also the case. Fixed-effect models in MASEM assume that either the correlation matrices or parameter estimates are from the same population (homogenous), whereas random-effect models assume the effect sizes (correlation matrices or parameter estimates) are from different populations (heterogeneous). Univariate models assume that correlations or parameter estimates are independent of each other whereas multivariate models assume that correlation matrices or parameter estimates are dependent.

Motivation for this Dissertation

Comparing MASEM Approaches Across Model Assumptions

Since the development of MASEM, research has recommendations and best practices have been published to help guide researchers in choosing the most appropriate MASEM method (Cheung, 2009; Sheng et al, 2016; Yuan, 2016). Sheng et al (2016) found that out of 160 MASEM studies, 36.3% acknowledged and discussed at least one issue with using MASEM. Additionally, they reported that out of 160 MASEMs, only nine (5.6%) had used the fixed-effect TSSEM approach, a multivariate approach, even 10 years after it had been published.

Several simulation studies have been conducted comparing the various methodological approaches of MASEM (Becker, 1992; Cheung & Chan, 2005; Furlow & Beretvas, 2005; Oort & Jak, 2016; Jak & Cheung, 2020; Cho, 2015; Zhang, 2011). Although there are heterogeneity indices to provide evidence as to whether the effect sizes originate from a fixed-effect or random-effects model, it is not recommended to rely on these indices alone; a theoretical basis should be included in the analysis of which model best represents the effect sizes (Borenstein et al, 2021; Cooper, 2017). However, it is not uncommon for researchers to use these heterogeneity indices to justify their model selection. Therefore, it is not a trivial task to compare MASEM approaches across model assumptions.

Overall Effects of Academic and Student Integration on Two-Year Students

Interestingly, although enrollment into higher education has generally increased over the past two decades, persistence and retention rates have stayed relatively consistent. Enrollment into an undergraduate institution of higher education has increased almost 26% over the past 20 years from roughly 13.2 million students to 16.6 million students (Hussar et al., 2020). In the fall of 2018 alone, around 35 percent (or roughly 5.7 million students) enrolled into a two-year institution, and it is projected to increase over the next 10 years (Hussar et al., 2020). However, student retention has only seen an increase of roughly 5% from the past 15 years, with graduation rates increasing 7.6% for four-year institutions and 2.1% for two-year institutions (National Center for Education Statistics, 2019, Table 326.10; National Center for Education Statistics, 2019, Table 326.20; National Center for Education Statistics, 2019, Table 326.30).

Although obtaining at least some education from a two-year college is becoming more prevalent as there are less barriers to entry, retention rates and graduation rates are significantly lower amongst two-year students when compared to their four-year counterparts. In 2018, the percentage of first-time full-time undergraduates retained at the four-year institution was roughly 81% with whereas for two-year institutions the retention rate was 62% (Hussar et al., 2020). The retention rate for part-time students at two-year institutions is even lower at around 45% (National Center for Education Statistics, 2019, Table 326.30). The overall graduation rate within 150% of normal time for first-time full-time undergraduates at four-year institutions (6 years) was 62% in 2018, whereas 33% of full-time undergraduates at two-year institutions graduated within 150% of normal time (in 3 years) (Hussar et al., 2020). One of the main contributors to the difference in retention and graduation can be explained by the difference in student population that is served from a four-year institution to a two-year institution.

Retention theories over the past 100 years have aimed to try to explain the factors that impact student retention, generally focusing on full-time students attending four-year institutions (Crisp & Mena, 2012). Major contributing factors usually involve student background characteristics, characteristics about the institution itself, and measures of the student's experience while enrolled at the institution. Dolan (2019) conducted a MASEM based on the work of Tinto (1975). Although the "Stage 2" model fit was moderate to poor, there were significant relationships between academic integration, social integration, institutional commitment, and organization factors, whereas student background characteristics and external factors were not significant predictors of student persistence. Because there are few-to-no MASEMs focusing on two-year institutions, there is a need to investigate how these retention theories apply to two-year college students, as well as a need to synthesize the impacts of academic and social integration on two-year college students.

Statement of Purpose & Research Questions

There are two purposes for this dissertation. The primary purpose is to compare eight MASEM approaches (the fixed- and random-effects models of the univariate *r*, univariate *z*, TSSEM and OSMASEM approaches -Table 1 displays the methods to be compared) using a data set to fit a path model relating to academic and social integration and student retention (see Figure 3). As a byproduct of the first purpose of this dissertation, the second purpose to add to the current body of higher education literature by interpreting the results of the MASEMs in addressing students attending two-year institutions and the impact of social and academic integration on their retention and success.

Table 1: MASEM Approaches

Method	Authors	Fixed/Random Effects	Univariate/ Multivariate
Univariate r	Viswesvaran & Ones (1995)	Fixed	Univariate
Univariate r	Viswesvaran & Ones (1995)	Random	Univariate
Univariate z	Hedges & Olkin (1985)	Fixed	Univariate
Univariate z	Hedges & Olkin (1985)	Random	Univariate
TSSEM	Cheung & Chan (2005)	Fixed	Multivariate
TSSEM	Cheung & Chan (2014)	Random	Multivariate
OSMASEM	Oort & Jak (2016)	Fixed	Multivariate
OSMASEM	Jak et al (2020)	Random	Multivariate

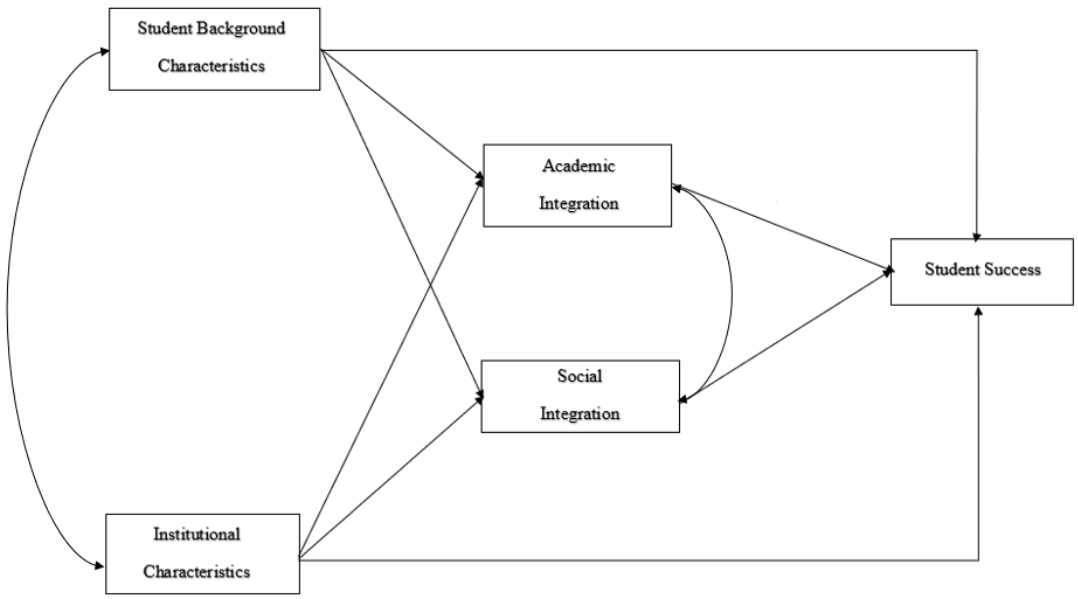

Figure 3: Proposed Path Model

The first purpose of this dissertation is paired with the following research questions:

RQ1a: How do the parameter estimates, standard errors, goodness-of-fit indices, and heterogeneity measures compare across all eight methods?

RQ1b: Are the findings in RQ1a consistent with current literature?

RQ1c: What are the implications for practitioners and researchers regarding the use of these methods?

The second purpose of this dissertation is paired with the model proposed below based on Yu (2015) along with the following questions regarding two-year college students:

RQ2a: What is the overall impact of student characteristics on academic integration?

RQ2b: What is the overall impact of student characteristics on social integration?

RQ2c: What is the overall impact of student characteristics on student retention?

RQ2d: What is the overall impact of institutional characteristics on academic integration?

RQ2e: What is the overall impact of institutional characteristics on social integration?

RQ2f: What is the overall impact of institutional characteristics on student retention?

RQ2g: What is the overall impact of academic integration on student retention?

RQ2h: What is the overall impact of social integration on student retention?

Significance of the Study

One of the unique aspects of this study is that it compares both fixed-effect and random-effects models in the same study. Most new MASEM approaches will compare their results within the same assumptions about the data (e.g., fixed effect or random effects). The results of this dissertation will contribute knowledge on the differences between approaches and across different model assumptions.

For higher education research, this dissertation will help provide an overall effect of social and academic integration, student background characteristics, and institutional characteristics on student retention for those attending two-year colleges. As two-year colleges have an increase in student attendance as well as face the diverse challenges to retain students, this dissertation can help provide some information on the average effect in addition to, more importantly, the variability of the effect. If the effect is extremely variable, further MASEMs can be conducted using subgroups of the population to determine for which groups academic and/or social integration are the most important (i.e., mixed effect approach).

CHAPTER II

LITERATURE REVIEW

Meta-Analysis

One of the primary goals of meta-analysis is to calculate an overall effect size by combining several effect sizes across studies using statistical methods. Gene Glass (1976) first used the term “meta-analysis” and defined it as: “…the analysis of analyses. I use it to refer to the statistical analysis of a large collection of analysis results from individual studies with the purpose of integrating the findings.” Although this term was first used in 1976, the interest in combining quantitative results started well before this time with Pearson (1904) and has only grown in interest across disciplines since then (Cooper, 2017).

There are some underlying assumptions about the effect sizes that need to be determined prior to conducting a meta-analysis. Generally, there are univariate approaches (which assumes independent effect sizes) and multivariate approaches (which consider the dependent effect sizes). Additionally, effect sizes can originate from a single population (fixed-effect model) or can originate from different populations (random-effects model). To assess heterogeneity across studies, content expertise and statistical measures can indicate which model best represents the effect sizes. Additionally, as maximum likelihood estimation is most often used to fit a SEM, the effect sizes are assumed to be multivariate normal (Kline, 2016). For univariate meta-analysis, effect sizes are assumed to be normally distributed (Borenstein et al, 2009)

Univariate Methods

As mentioned above, one assumption about the effect sizes is that they either are obtained from a single population (fixed-effect model) or obtained from several different populations (random-effects model). The variation in the study effect sizes from the population is due to sampling error (within studies variation) for a fixed-effect model. In a random-effects model, the difference between a study effect size and the mean population effect size can be due to sampling error (within studies error) as well as error due to belonging to different populations (between studies error).

The univariate approach in meta-analysis assumes that effect sizes are independent of each other. For K studies, let T_k represent the kth sample effect size and v_k represent the sample variance of the kth study. Then,

$$T_k = \theta + e_k,$$

where θ represents the population effect size and e_k represents the sampling error (within-studies error term). It is assumed that T_k is normally distributed with mean θ and variance σ_k^2.The estimated overall effect size under a fixed-effect model is calculated using the inverse-variance weighted mean:

$$\hat{\theta} = \frac{\sum_{k=1}^{K} \hat{w}_k T_k}{\sum_{k=1}^{K} \hat{w}_k},$$

with weight

$$\hat{w}_k = \frac{1}{v_k},$$

and

$$var_{\hat{\theta}} = \frac{1}{\sum_{k=1}^{K} \hat{w}} = \frac{1}{\sum_{k=1}^{K} \frac{1}{v_k}}.$$

For the random-effects model, there is an additional error term in the estimated effect size that is due to between-studies variation. That is,

$$T_k = \mu + u_k + e_k,$$

where μ represents the mean of the population distribution of all true effect sizes, u_k represents the between-studies error term and e_k represents the within-studies error term. It is assumed that both error terms are normally distributed with $e_k \sim N(0, \sigma_k^2)$ and $u_k \sim N(0, \tau^2)$ with T_k also normally distributed with mean θ_k and variance $\sigma_k^2 + \tau^2$. The random-effects model is calculated using the bivariate weighted mean:

$$\hat{\mu} = \frac{\sum_{k=1}^{K} \hat{w}_k^* T_k}{\sum_{k=1}^{K} \hat{w}_k^*},$$

with weight

$$\hat{w}_k^* = \frac{1}{v_k + \hat{\tau}^2},$$

where $\hat{\tau}^2$ represents the estimated between-study variation and

$$var_{\hat{\mu}} = \frac{1}{\sum_{k=1}^{K} \hat{w}_k^*} = \frac{1}{\sum_{k=1}^{K} \frac{1}{v_k + \hat{\tau}^2}}.$$

Note that the overall effect size of a fixed-effect model can be derived from the random-effects model ($\hat{\tau}^2 = 0$).

Multivariate Methods

Multivariate meta-analysis takes into consideration dependence of multiple effect sizes reported in the same study. Cheung (2015) described multivariate analysis with the following notation. Let p be the number of effect sizes that are of interest in a multivariate meta-analysis and p_i be the number of observed effect sizes in the ith study. It is assumed $p_i \leq p$, because every study will at most contain all the same variables across studies.

For multiple effect sizes, the univariate fixed-effect equation can be extended. Let the column vector $\boldsymbol{y}_i$ represent a $p_i \times 1$ matrix of observed effect sizes such that

$$\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{f}_i + \boldsymbol{e}_i,$$

where $\boldsymbol{X}_i$ is a $p_i \times p$ design matrix with 0 and 1 to select the observed effect sizes, $\boldsymbol{f}_i$ is a $p \times 1$ vector of population effect sizes, and $\boldsymbol{e}_i$ is a $p_i \times 1$ vector of sampling errors that is assumed to be multivariate normally distributed with mean 0 and known covariance matrix $\boldsymbol{V}_i$ for large sample sizes.

For the multivariate fixed-effect model, it is assumed that all true effects are the same; that is,

$$\boldsymbol{f}_1 = \boldsymbol{f}_2 = \cdots = \boldsymbol{f}_k = \beta_F.$$

Therefore, the fixed-effect model for the ith study is

$$\boldsymbol{y}_i = \boldsymbol{X}_i \beta_F + \boldsymbol{e}_i,$$

To estimate the overall fixed-effect size, the following function is minimized using generalized least squares (GLS) and known covariance matrix $\boldsymbol{V}$:

$$F_{GLS} = (\boldsymbol{y} - \boldsymbol{X}\beta_F)^T \boldsymbol{V}^{-\mathbf{1}} (\boldsymbol{y} - \boldsymbol{X}\beta_F).$$

The estimate for the overall effect size and asymptotic sampling covariance in a fixed-effect model is

$$\hat{\beta}_F = (\boldsymbol{X}^T \boldsymbol{V}^{-\mathbf{1}} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{V}^{-1} \boldsymbol{y}$$

$$\hat{\Omega}_F = (\boldsymbol{X}^T \boldsymbol{V}^{-1} \boldsymbol{X})^{-1}.$$

For the random-effects model for the ith study is:

$$\boldsymbol{y_i} = \boldsymbol{X_i}\beta_R + \boldsymbol{Z}\boldsymbol{u}_i + \boldsymbol{e_i},$$

where $\boldsymbol{Z} = diag(\boldsymbol{Z}_1, \boldsymbol{Z}_2, \dots, \boldsymbol{Z}_k)$ is a selection of 1s and 0s to select the random effects, $\boldsymbol{u} = [\boldsymbol{u}_1^T | \boldsymbol{u}_2^T | \dots | \boldsymbol{u}_k^T]^T$ is the stacked random effects for all studies, β_R is the average population effect sizes under the random-effects model, and $\boldsymbol{y}_i$, $\boldsymbol{X}_i$, and $\boldsymbol{e}_i$ are defined the same as for the fixed-effect model.

Similar to the fixed-effect model, we can minimize the function,

$$F_{GLS} = (\boldsymbol{y} - \boldsymbol{X}\beta_R)^T \widetilde{\boldsymbol{V}}^{-\mathbf{1}} (\boldsymbol{y} - \boldsymbol{X}\beta_R),$$

using GLS to estimate $\hat{\beta}_R$ for a random-effects model with known covariance matrix,

$$\widetilde{\boldsymbol{V}} = \boldsymbol{Z}_i \widehat{\boldsymbol{T}}^2 \boldsymbol{Z}_i^T + \boldsymbol{V}_i,$$

where $\boldsymbol{Z}_i$ is used to select the random effects and $\widehat{\boldsymbol{T}}^2$ as the random effects.

The random-effects overall effect and asymptotic sampling covariance matrix can be estimated by

$$\hat{\beta}_R = \left(\boldsymbol{X}^T \widetilde{\boldsymbol{V}}^{-\mathbf{1}} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \widetilde{\boldsymbol{V}}^{-1} \boldsymbol{y}$$

and

$$\hat{\Omega}_R = \left(\boldsymbol{X}^T \widetilde{\boldsymbol{V}}^{-1} \boldsymbol{X}\right)^{-1}.$$

Assessing Homogeneity

Because the underlying model for the effect size depends on whether the effect size does or does not originate from a single population, it is a common question of which model should be used. Borenstein et.al (2009) described four common statistical indicators to assess heterogeneity; the *Q*-statistic (a measure of weighted squared deviations from the fixed effect), the *p*-value (test of homogeneity using *Q*), the I^2 index (the magnitude of heterogeneity), τ^2(estimated between studies variance), and τ (estimated between studies standard deviation). Each of the measures of heterogeneity mentioned above provide a unique measure of heterogeneity, however Borenstein et.al (2009) emphasize that these indicators should not be relied on alone to determine whether a random-effects or fixed-effect model should be used as each indicator has its limitations (such as sensitivity to the number of studies or effect size).

The Test for Homogeneity

Two of the statistical heterogeneity measures are the *Q*-statistic and the *p*-value from the test of homogeneity (Cochran, 1954). The idea behind the *Q*-statistic and corresponding test is to determine if the effect sizes differ significantly from the fixed-effect model. If the *Q*-statistic is large, then there is a possibility a random-effects model should be considered since the effect sizes are far from the fixed effect. A small *Q*-statistic would suggest that the effect sizes are close to the fixed effect. Referring back to notation by Cheung (2015), in the univariate case, the null hypothesis for the test of homogeneity is

$$H_0: \beta_F = f_1 = f_2 = \cdots = f_k.$$

The formula for the *Q*-statistic using a univariate approach is,

$$Q_{UNI} = \sum_{i=1}^{k} w_i \left(y_i - \hat{\beta}_F\right)^2,$$

where the *Q*-statistic has a chi-square distribution with $(k - 1)$ degrees of freedom.

In the multivariate case, the null hypothesis for the test of homogeneity is,

$$H_0: \beta_F = \boldsymbol{X}_1 \boldsymbol{f}_1 = \boldsymbol{X}_2 \boldsymbol{f}_2 = \cdots = \boldsymbol{X}_k \boldsymbol{f}_k \,.$$

The formula for the *Q*-statistic using a multivariate approach is,

$$Q_{MUL} = \left(\boldsymbol{y} - \boldsymbol{X}\hat{\beta}_F\right)^T \boldsymbol{V}^{-1} (\boldsymbol{y} - \boldsymbol{X}\hat{\beta}_F),$$

where the *Q*-statistic is approximately distributed as a chi-square distribution with $\sum_{i=1}^{k} p_i - p$ degrees of freedom in large samples. A statistically significant result (rejection of the null hypothesis) may indicate that the studies are not from the same population, however the *p*-value to this test does not quantify the magnitude of heterogeneity between studies (only that they are not the same).

Quantifying the Percentage of Variance

To quantify the percentage of heterogeneity between studies, the I^2 index is one of the most common measures, which was defined by Higgins and Thompson (2002).

The general formula is,

$$I^2 = \frac{\hat{\tau}^2}{\hat{\tau}^2 + \tilde{v}} = \sum \frac{Q - df}{Q} \times 100\%,$$

where $\tilde{v}$ is a typical within-study sampling variance.

In the multivariate case,

$$I^2_{Q(MUL)} = 1 - \frac{df_{MUL}}{Q_{MUL}},$$

where Q_{Mul} and df_{Mul} are the *Q*-statistic and its degrees of freedom in testing the homogeneity of effect sizes.

I^2 can be interpreted as the proportion of the total variation of the effect size that is due to the between-study heterogeneity. Another interpretation is the percent of effect size variability which is not explained by sampling error. The scale of I^2 has a range of 0-100% and values of 25%, 50%, and 75% have been considered as low, moderate, and high levels of heterogeneity (Higgins et al, 2003).

Quantifying Between-Studies Variance

The last common measure of heterogeneity is τ^2, which represents the variance of the true effect sizes (i.e., the between studies variance). The scale of τ^2 can be interpreted in the same units as the effect size and as the average squared deviation.

To estimate τ^2, we can use the formula,

$$\tau^2 = \frac{Q-df}{C},$$

where

$$C = \sum W_i - \frac{\sum W_i^2}{\sum W_i}.$$

There have been several methods developed to estimate τ^2; the most common ways being the DerSimonian and Laird (1986) method, restricted maximum likelihood (REML) method, and the Paule-Mandel (1982) method. Stijnen, White, and Schmidt (2020) found from several simulation studies that the REML method is generally the recommended method for continuous outcomes.

Structural Equation Modelling

Kline (2016) described six basic steps in most SEM analyses: 1) specifying the model; 2) evaluating model identification; 3) selecting the measures and collect, prepare, and screen the data; 4) estimating the model; 5) respecifying the model, which is assumed to be identified; and 6) reporting the results.

When specifying the model, there are generally three types: path models, measurement models, and structural models. Path models are used to specify relationships among observed variables. Measurement models or confirmatory factor models (CFAs) are used to specify how observed variables are related to latent variables. Structural equation models are used to specify how latent variables are related to other latent variables.

After specifying a model, the model needs to be identified before it can be estimated. A model is identified if it is theoretically possible to estimate every model parameter (Kline, 2016). If the model is not identifiable, parameter estimates cannot be obtained therefore the model needs to be respecified. For path analysis, there are a few rules to help determine if a path model is identifiable. There are three identification rules for path models: the t-rule, the Null-B rule, and the Recursive rule (including the counting rule which applies to all SEMs).

To obtain parameter estimates and test statistics, summary statistics (such as means, covariances, correlations, and standard deviations) can be used or the raw data can be used. The estimation is usually based on a discrepancy function (Browne, 1982). A discrepancy function returns a scalar value of the difference between the sample covariance matrix (and the means) and the model-implied covariance matrix (and means). If the discrepancy is zero, the model-implied covariance matrix and sample covariance are the same. The parameters are then estimated by minimizing the discrepancy function (Cheung, 2015).

The most popular estimation procedures include maximum likelihood estimation, weighted least squares estimation, and generalized least squares estimation. Maximum likelihood (ML) estimation is used the most when the data are distributed multivariate normal. The parameter estimates of the covariance structure can be obtained my minimizing the ML discrepancy function $F_{ML}(\theta)$:

$$F_{ML}(\theta) = \log|\sum(\theta)| + tr(S\Sigma(\theta)^{-1}) - \log|S| - p,$$

where $tr(X)$ is the trace of X that takes the sum of the diagonal elements of X and p is the number of variables in the model. If the data is not normally distributed, weighted-least squares (WLS) estimation can be used.

Once the model is estimated, there are several indices to determine how well the model fits the data. These include the chi-square test, the comparative fit index (CFI) (Bentler, 1990), root mean square error of approximation (RMSEA) (Steiger, 1990), and the standardized root mean square residual (SRMR).

The results of the chi-square test for overall fit indicate whether the model fits the data exactly. However, one limitation to this statistic is that it is sensitive to large sample sizes; therefore, the chi-square test will reject the null hypothesis every time if a sample size is very large. Thus, other goodness-of-fit indices (such as CFI, RMSEA, SRMR) can be used in addition to the chi-square test to determine the type of model fit. The CFI index compares the specified model with the worst, most restricted model, whereas the RMSEA and SRMR indices compare the specified model the saturated model. Generally, an RMSEA and SRMR value of less than 0.05 indicate good model fit, while values less than 0.08 indicate fair fit; a CFI value greater than 0.95 indicates good model fit (Browne & Cudeck, 1993).

If the model fit is poor, the model can be respecified by adding or dropping paths using modification indices or expected parameter changes, however the decision to respecify should be theoretically sound and within the context the field.

Multigroup SEM

The above description refers to a single group analysis. However, it is not uncommon for a SEM to be different for various samples. As an example, a simple path model consists of a child's age and its effect on height in K-12. This path estimate may be different by gender. One way to assess this is to fit the model by separating the dataset by gender, fitting the model on both data sets, and then using the chi-square difference test to determine if the estimates parameters are the same.

The fit function for multi-group SEM as written in Cheung (2015) to be minimized is:

$$F_{MG}(\theta) = \frac{\sum_{i=1}^{k}(n_i^{MG}-1)F_i(\theta)}{\sum_{i=1}^{k}(n_i^{MG}-1)},$$

where $F_i(\theta)$ is the fit function for the ith group.

MASEM Approaches

As previously stated in Chapter 1, there are two general approaches to MASEM: parameter-based and correlation-based (see Figures 1 and 2 in Chapter 1). In parameter-based MASEM, the goal is to combine parameter estimates (e.g., factor loadings or path coefficients) across studies. In correlation-based MASEM, the goal is to first combine correlation matrices and then use the correlation matrix to fit a SEM. Correlation-based

MASEM has several advantages over parameter-based, which include are ease of use and the ability to quantify heterogeneity between studies. The most common MASEM approaches are the univariate *r*, univariate *z*, generalized least squares (GLS), and Two-Stage SEM (TSSEM).

Univariate Approaches

The univariate approaches to MASEM are in two stages: pooling the correlation matrices across studies by using univariate methods, and then treating the pooled correlation matrix as the observed covariance matrix to fit a SEM. There are two commonly used univariate approaches; one approach synthesizes correlations (r), whereas the second approach synthesizes Fisher's z values.

Stage 1: Combining Correlation Matrices using the Univariate R Approach

Becker et al (2020) describe the univariate and multivariate approaches of combining correlation studies. For the univariate approach, assume that I represents the total number of observed variables within a study and $Y_1, \ldots, Y_I$ represent the observed variables which are assumed to be normally distributed and statistically independent. The estimated correlation for the population correlation ρ_{ist} between variables Y_s and Y_t in the ith study is represented by r_{ist}. The sample size for the ith study is represented by n_i.

The assumption for the fixed-effect model is,

$$r_{ist} = \rho_{st} + \varepsilon_{ist},$$

where ε_{ist} represents within study error. The inverse-variance weighted mean correlation between Y_s and Y_t is calculated by:

$$\bar{r}_{FE_{st}} = \frac{\sum_{i=1}^{I} w_{ist} r_{ist}}{\sum_{i=1}^{I} w_{ist}},$$

where $w_{ist} = \frac{1}{s^2_{r_{ist}}}$ and $s^2_{r_{ist}} = \frac{\left(1-r^2_{ist}\right)^2}{n_i - 1}$.

The assumption for the random-effects model is,

$$r_{ist} = \rho_{st} + \varepsilon_{ist} + u_{ist},$$

where u_{ist} represents the between study error. The bivariate-weighted mean correlation is calculated by:

$$\bar{r}_{RE_{st}} = \frac{\sum_{i=1}^{I} w^*_{ist} r_{ist}}{\sum_{i=1}^{I} w^*_{ist}},$$

where $w_{ist} = \frac{1}{s^2_{r_{ist}} + \hat{\tau}^2_{r_{st}}}$ with $\hat{\tau}^2_{st}$ as the estimated between studies variance for the correlation between Y_s and Y_t, and $s^2_{r_{ist}} = \frac{\left(1-r^2_{ist}\right)^2}{n_i - 1}$.

Stage 1: Combining the Correlation Matrixes Using the Fisher's z Approach

The correlation coefficient, r_{ist}, is usually not used to calculate an overall effect size because the variance depends on the size of the correlation coefficient (Cooper et al, 2019). Therefore, correlations are transformed into the Fisher's z scale, analysis is conducted, and then the result can be converted back into a correlation coefficient for interpretation. This approach was developed by Hedges and Olkin (1985).

The Fisher's z transformation is,

$$z_{ist} = 0.5 \ln\left(\frac{1+r_{ist}}{1-r_{ist}}\right),$$

with variance

$$s_{z_{ist}}^2 = \frac{1}{n_i - 3}$$

For fixed-effect models, the weighted mean correlation is calculated by:

$$\bar{z}_{FE_{st}} = \frac{\sum_{i=1}^{I} w_{ist}^{+} z_{ist}}{\sum_{i=1}^{I} w_{ist}^{+}},$$

where $w_{ist}^{+} = \frac{1}{s_{z_{ist}}^2}$and $s_{z_{ist}}^2 = \frac{1}{n_{ist}-3}$.

For random-effect models, the weighted mean correlation is calculated by,

$$\bar{z}_{RE_{st}} = \frac{\sum_{i=1}^{I} w^{++}{}_{ist} z_{ist}}{\sum_{i=1}^{I} w_{ist}^{++}},$$

where $w_{ist}^{++} = \frac{1}{s_{z_{ist}}^2 + \hat{\tau}_{z_{st}}^2}$ and $s_{ist}^2 = \frac{1}{n_{ist}-3}$.

To transform the Fisher's z back to the correlation coefficient r,

$$r = \frac{e^{2z}-1}{e^{2z}+1}.$$

Stage 2: Fitting the SEM for the Univariate Approaches

To fit the SEM, the pooled correlation matrix is treated as the observed covariance matrix and maximum likelihood is used for estimation . If there are no missing correlations, the sum of the sample sizes can be used; however, if correlations are missing, averages like the harmonic mean, arithmetic mean and median have been used (Cheung, 2015).

The Two-Stage SEM Approach

The Two-Stage Structural Equation Model, developed by Cheung & Chan (2005), uses SEM in both steps of the MASEM process. This is a unique approach that models the meta-analysis with a SEM in addition to using SEM in the second stage. To obtain the correlation matrix, a multi-group SEM is conducted to first determine if there is homogeneity between studies. To test for homogeneity, the Likelihood-Ratio (LR) statistic and goodness-of-fit indices in SEM are used. If the test of homogeneity suggests that the correlation matrices are the same, this implies a fixed-effect model may be appropriate and the corresponding correlation matrix where the SEM paths were contained to be equal is used for stage 2.

Stage 1: Pooled Correlation Matrix in Fixed-Effect TSSEM

From Cheung (2005), the covariance matrix can be decomposed into the matrices of standard deviations and correlations such that,

$$\boldsymbol{\Sigma}^g = \boldsymbol{D}^g \boldsymbol{P}^g \boldsymbol{D}^{g^T},$$

and $Diag[P^g]$ are 1's where D^g is the $p^g x p^g$ diagonal matrix of standard deviations and the P^g is the $p^g x p^g$ correlation matrix in the gth study, respectively. Equation 38 is equivalent to a CFA model with,

$$\sum(\boldsymbol{\theta})^g = \boldsymbol{\Lambda}^g \boldsymbol{\Phi}^g \boldsymbol{\Lambda}^{g^T} + \boldsymbol{\Psi}^g,$$

where $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}$, and $\boldsymbol{\Psi}$ are the factor loadings, factor covariance, and error variance matrices, respectively. Then using the previous equation, $\boldsymbol{\Lambda}^g$ is a $p^g \times p^g$ diagonal matrix (the standard deviation matrix $\boldsymbol{D}^g$), $\boldsymbol{\Phi}^g$ is a $p^g \times p^g$ standardized matrix (the correlation matrix $\boldsymbol{P}^g$), and $\boldsymbol{\Psi}^g$ is a $p^g \times p^g$ zero matrix.

To obtain the estimated pooled correlation matrix, all the factor correlation matrices $\boldsymbol{\Phi}^g$ must be equal. To test whether the correlation matrices are the same, suggesting a fixed-effect model, a multigroup SEM can be conducted. The number of constraints imposed is,

$$\sum_{g=1}^{K} \frac{p^g(p^g-1)}{2} - \frac{p(p-1)}{2}.$$

A chi-square difference test can be used to evaluate the equality constraints by comparing the model with constraints on the equality of correlation matrices against the model without constraints. The test statistic is asymptotically distributed as a chi-square with degrees of freedom

$$\sum_{g=1}^{K} \frac{p^g(p^g-1)}{2} - \frac{p(p-1)}{2}.$$

Moreover, goodness of fit indices can also be used to evaluate the model fit. The estimate $\hat{P}$ is the pooled correlation matrix and the asymptotic covariance matrix of parameter estimates $\hat{V}$ is the asymptotic covariance matrix of the pooled correlation matrix.

When the hypothesis of homogeneity of the correlation matrices is not rejected, we can use the estimate of the pooled correlation matrix $\hat{P}$ to fit SEM in stage 2. If the studies are heterogeneous, then random-effects TSSEM should be used.

Stage 1: Pooled Correlation Matrix in Random-Effects TSSEM

For the random-effects TSSEM approach, Cheung (2015) uses multivariate meta-analysis as the first step to combine correlation matrices, then fits a SEM with the pooled correlation matrix using the same steps as the fixed-effect TSSEM in stage 2.

The multivariate random-effects model can be calculated either using a GLS approach (Becker, 1992) or can be calculated using an SEM approach (Cheung, 2015).

Therefore, for the first stage of random-effects TSSEM, the known sampling covariance matrix V_i in each study needs to be estimated using the formula

$$Cov(r_{ij}, r_{kl}) = \frac{0.5\rho_{ij}\rho_{kl}\left(\rho_{ik}^2+\rho_{il}^2+\rho_{jk}^2+\rho_{jl}^2\right)+\rho_{ik}\rho_{jl}+\rho_{il}\rho_{jk}-(\rho_{ij}\rho_{ik}\rho_{il}+\rho_{ij}\rho_{jk}\rho_{jl}+\rho_{ik}\rho_{jk}\rho_{kl}+\rho_{il}\rho_{jl}\rho_{jl})}{n}.$$

Then, a multivariate meta-analysis can be conducted using the correlation matrices and estimated sampling covariance matrix.

The average correlation matrix $\hat{\boldsymbol{p}}_{\boldsymbol{R}}$ based on a random-effects model and its asymptotic sampling covariance matrix $\hat{\boldsymbol{V}}_{\boldsymbol{R}}$ are estimated. Multivariate heterogeneity indices can be used to determine if the correlation matrices are the same across studies.

Stage 2: Fitting the SEM for the TSSEM approach

The second stage of fitting an SEM on the pooled correlation matrices is the same for the fixed-effect and random-effects TSSEM approaches. For the second stage, the weighted least squares (WLS) estimation method is used. After analysis from Stage 1, we have the estimate of the pooled $p \times p$ correlation matrix with its

$$\frac{p(p-1)}{2} \times \frac{p(p-1)}{2}$$

asymptotic covariance matrix of parameter estimates $\hat{\boldsymbol{V}}$ and the total sample size N, which equals the sum of all sample sizes, that is,

$$N = \sum_{g=1}^{K} n^g.$$

The discrepancy function is,

$$F(\boldsymbol{\gamma}) = \left(\boldsymbol{r}^* - \boldsymbol{\rho}(\boldsymbol{\gamma})\right)^T \widehat{\boldsymbol{V}}^{-1}(\boldsymbol{r}^* - \boldsymbol{\rho}(\boldsymbol{\gamma}))$$

where $\boldsymbol{r}^*$ and $\boldsymbol{\rho}(\boldsymbol{\gamma})$ are the $p' \times 1$ vectors of $p' = \frac{p(p-1)}{2}$ elements obtained by straining out the lower triangular elements, excluding the diagonals in the sample and the implied correlation matrices R and $P(\gamma)$, respectively. $\widehat{\boldsymbol{V}}$ is the $p' \times p'$ weight matrix estimated from the first stage and γ is a structural parameter vector.

The One-Stage MASEM Approach

Fixed-Effect One-Stage MASEM

Oort and Jak (2016) proposed a fixed-effect ML approach to MASEM that Jak and Cheung (2020) refer to as a fixed-effect one-stage MASEM (OSMASEM). Although their approach is like the first stage of TSSEM, one notable difference is that the study correlation matrices are used throughout the analysis, instead of a pooled correlation matrix being obtained in "stage 1". Additionally, ML estimation is used throughout their approach, whereas WLS estimation was used in the second stage of TSSEM.

The OSMASEM approach compares three nested SEMs: 1) a saturated model, 2) a multi-group model, and 3) a multi-group model that uses the model implied correlation matrix from 2). Let p_g be the number of observed variables in the gth study. Further, let q represent the number of all variables. Some studies may be missing some variables and correlations, therefore $p_g \leq q$. The saturated model is,

$$\mathbf{\Sigma}_g = \boldsymbol{D}_g \boldsymbol{P}_g \boldsymbol{D}_g^T,$$

where $\boldsymbol{P}_g$ is a $p_g \times p_g$ correlation matrix for the gth study and D_g is a $p_g \times p_g$ matrix that accounts for differences in scaling of the variables across G studies.

The multi-group model used to test for homogeneity is,

$$\mathbf{\Sigma}_g = \boldsymbol{D}_g \left(\boldsymbol{M}_g \boldsymbol{P} \boldsymbol{M}_g^T \right) \boldsymbol{D}_g,$$

where $\boldsymbol{P}$ is a $q \times q$ population correlation matrix with diagonals equal to 1, $\boldsymbol{M_g}$ is a $p_g \times q$ selection matrix to obtain the observed correlations, and $\boldsymbol{D}_g$ is a $p_g \times p_g$ matrix that accounts for differences in scaling of the variables across G studies.

The third model is to test the model fit of the SEM is,

$$\mathbf{\Sigma}_g = \boldsymbol{D}_g \left(\boldsymbol{M_g} \boldsymbol{P_{model}} \boldsymbol{M}_g^T \right) \boldsymbol{D_g},$$

where $\boldsymbol{P}_{model}$ is the model implied correlation matrix.

Therefore, the test for homogeneity uses a chi-square difference test and LR test between the first and second models. If there is evidence to suggest homogeneity across correlation matrices, then the second and third models are used in a chi-square difference test and likelihood ratio test to determine overall model fit.

Random-Effects One-Stage MASEM

In one-stage MASEM, Jak and Cheung (2020) consider the correlation coefficients as the “variables” and the studies as “subjects” in the data set. OSMASEM fits the SEM by restricting the pooled correlations in the multivariate random-effects model. In its simplest form, the random-effects model decomposes the vector r_i of observed correlation coefficients for a study i into three parts:

$$\boldsymbol{r_i} = \boldsymbol{\rho_R} + \boldsymbol{u_i} + \boldsymbol{\varepsilon}_i,$$

where $\boldsymbol{\rho_R}$ indicates the mean vector of the correlation coefficients, $\boldsymbol{u_i}$ is a vector of deviations of study $i's$ population correlation coefficients from $\boldsymbol{\rho_R}$, and $\boldsymbol{\varepsilon_i}$ is a vector with the sampling error of study i. The term $Cov(u_i) = T^2$ denotes the between-studies covariance matrix that must be estimated, and $Cov(\varepsilon_i) = V_i$ denotes the sampling covariance of the correlation coefficients, which is usually treated as known in a meta-analysis.

For example, a path model is nested under the previous equation by restricting $\boldsymbol{\rho_R}$:

$$\boldsymbol{\rho_R} = vechs(\boldsymbol{F}(\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{S}(\boldsymbol{I} - \boldsymbol{A})^{-1^T}\boldsymbol{F}^T)$$

where using the RAM formulation, $\boldsymbol{I}$ is an identity matrix, $\boldsymbol{F}$ is a selection matrix with 1's for observed variables and 0's for latent variables, $\boldsymbol{A}$ is a square matrix with asymmetric paths, $\boldsymbol{S}$ is a symmetrical matrix with variances and covariances, and $vechs(\cdot)$ vectorized the lower diagonal of its argument.

In general, for models with p observed variables, $\boldsymbol{\rho_R}$ will be a $p \times (p-1)/2$ dimensional column vector, and $\boldsymbol{A}$, $\boldsymbol{S}$, and $\boldsymbol{F}$, will be of dimension $p \times p$. Because we are using correlation matrices as inputs, the variances of the exogeneous variables are fixed at 1 and the diagonal elements of the model implied correlation matrix in the above equation should always be equal to one during estimation.

All model parameters are estimated with full information maximum likelihood (FIML). A test statistic of the hypothesized model can be obtained by performing a likelihood ratio test with the saturated model, similar to Oort and Jak (2016).

Comparison of Approaches

Overall, several simulation studies suggest that when within study sample size is large, parameter estimates, and their standard errors are similar. However, if the within study sample size is not large, there are notable discrepancies between each approach's performance.

Cheung and Chan (2005) compared the univariate *r*, univariate *z*, generalized least squares (GLS), and Two-Stage SEM (TSSEM) fixed-effect approaches in a simulation study with an empirical example. They ultimately concluded that the univariate *r*, univariate *z*, and TSSEM approaches performed well enough that any of the three methods could be used in stage 1; however, they recommended that the TSSEM approach be used in stage 2 as the goodness-of-fit indices performed better compared to the univariate and GLS approach.

Jak and Cheung (2020) compared a subgroup TSSEM approach (Jak & Cheung, 2018) and OSMASEM approach using an empirical example and by conducting a simulation analysis. In their empirical analysis, they found that parameter estimates, and standard errors were the same for each approach when moderators were not accounted for. Additionally, the τ^2 values were the same as well. For factor analysis, the model fit was good with almost identical fit-statistics, parameter estimates, and standard errors for a bi-factor model.

Oort and Jak (2016) conducted a simulation study comparing the fixed-effect OSMASEM and fixed-effect TSSEM. Overall, they found that the OSMASEM and TSSEM generally yielded very similar results, although the OSMASEM parameter estimates were significantly less biased when sample sizes are small. Test statistics, confidence intervals, false positive rates, and true positive rates did not differ between the two methods.

Cho (2015) using simulations to compare the factor loadings and standard errors of four fixed-effect univariate approaches and two fixed-effect multivariate approaches. The main takeaways from this study were that when sample sizes were large (within studies sample size of $n > 150$), the results across all seven approaches seemed to be similar; however, this was not the case when sample sizes were small.

Zhang (2011) compared three different multivariate approaches; the traditional GLS approach, a modified GLS approach, and the TSSEM approach for both fixed and random effects using simulation analysis. The results show that the modified GLS approach performs as well as or better than the TSSEM approach in both the first and second stages for both fixed-effect and random-effects data. The original GLS only performs well when the within study sample size is large enough. Both the modified GLS approach and the TSSEM approach produce equivalent parameter estimates across all conditions, however the standard errors from the TSSEM approach seem to be over-estimates under certain conditions.

Applying MASEM to Two-Year College Student Success

Student success is a term used at every level of education. Helping students progress through their educational programs and attain degrees results not only in better quality of life for students, but benefits society as well (Chen et al., 2020; Hussar et al., 2020; Ma et al., 2019). In higher education, persistence and retention are terms often used synonymously to describe the progress a student makes throughout their educational journey. Although other metrics such as human capital and student learning outcomes are used to measure student success, the most popular metrics used by institutions to measure persistence or retention are by the number of credits enrolled in each semester, enrollment from semester to semester, degree attainment, and graduation rates (Baldwin, Bensimon, Dowd & Kleiman, 2011; Mortenson, 2005; Mullin, 2012). Overall, these metrics aim to identify the students who are progressing well throughout their educational journeys, as well as those who are not progressing well with the intent to identify and develop interventions to help these students succeed. This can prove to be a complex task as students do not always move linearly throughout their educational careers; some students may transfer, or some students may need a break from school only to return years later.

Because two-year colleges generally have an "open-door" admissions process, it is expected that the student populations between a two- and a four-year institutions will not be the same. Two-year institutions were comprised of almost 63% of part-time students, whereas four-year institutions were comprised of 29% of part-time students in 2018 (National Center for Education Statistics, 2019, Table 303.30). Additionally, two-year colleges generally serve a higher proportion of first-generation students, underrepresented minority students (such as African American and Hispanic students), and students who achieved a high school grade point average less than 3.0 (Crisp & Mena, 2012). Another notable difference in the student populations served are their goals. Two-year colleges enroll more transfer students, vocational students, developmental student, community education students, dual enrollment students and English as a Second Language students compared to four-year institutions (Crisp & Mena, 2012).

In summary, when compared to four-year students, 2-year students are more likely to be: African American or Hispanic; financially independent; first-generation college students; less academically prepared; working part-or full-time during college; having lower degree aspirations; attending part-time; delaying enrollment into college following high school; receiving less financial aid; and earning a lower GPA during the first year of college (Astin and Oseguera, 2012). Because of these differences, it can be hypothesized that academic and social integration will have different impacts on students compared to the four-year student which most research is based on.

Student Background Characteristics and Institutional Characteristics

Researchers have repeatedly found that students' chances of degree attainment are substantially impacted by student background characteristics too. Astin and Oseguera (2012) conducted a longitudinal analysis to determine how pre-college characteristics, the characteristics of the college, and the college environment impacted the chances of degree completion. They found that high school grade average to be the strongest pre-college predictor of the student's chances of completing a bachelor's degree within four or six years after starting college. The institutional characteristic with the strongest effect was selectivity with the more selective the institution, the better the student's chances of finishing. This of course creates a unique challenge for two-year institutions who have an open-door policy.

Research has also looked heavily into institutional characteristics that impact student retention and graduation, as well as student characteristics. This includes researchers who study the individual effectiveness of programs and practices (Tinto, 2012a; Ziskin et al., 2012). For example, it is well known that first-year retention is a significant predictor in graduation (Schneider, 2010; Tinto, 2012a). Therefore, interventions focused on first-year experiences, such as freshman seminars and learning communities, tend to increase retention by assisting students in their transition to college both socially and academically (Braunstein & McGrath, 1997; Mertes & Hoover, 2014; Tinto, 2012a). Tinto (2012b) theorized that support, expectations, involvement, and feedback are essential elements that together contribute to student success. These four elements, in addition to student background characteristics, are the core of his Model of Institutional Action (Tinto, 2012a). Each of these elements can be found at varying levels at an institution, from administrative departments to classrooms. For two-year college students, most of these criteria are focused on the classroom level (where most student experiences are formed, especially for two-year college students), however they can be extended to the administration level as well, such as advising and financial support.

Academic and Social Integration on Student Success

Institutions were mainly concerned with enrollment and curriculum development up until the 1930s, when retention research began. (Berger & Lyon, 2005, as cited in Morrison & Silverman, 2012). Seminal theories regarding student retention include the work of Spady (1970), Tinto (1975, 1993), Bean and Metzner (1985), and Seidman

(2005). These researchers identified and formalized concepts of social and academic integration and their relation to student persistence and retention, as well as extending student retention research to non-traditional students and institutional program evaluation (Morrison & Silverman, 2012). Academic integration can be defined as student interactions and experiences relating to academic and intellectual development, particularly with faculty and staff (Henningsen, 2003). Social integration are student interactions and experiences with peers and faculty (generally informal interactions) (Henningsen, 2003).

Spady (1970) and Tinto (1975, 1993) posited that the social and academic experiences of students shaped whether they decided to persist or to drop out. In early retention research, institutional and student characteristics from the institution were frequently used (Morrison & Silverman, 2012). However, Spady (1970) introduced a model that related academic systems (such as grade performance and academic potential) with social systems (such as social integration and friendship support) and examined their impacts on institutional commitment and ultimately on a dropout decision. Most notable about his work was how he used an interdisciplinary approach, namely the framework from Durkheim (1961)'s theory of suicide, to examine student departure. Bean and Metzner (1985) studied non-traditional students (defined as older, part-time, and commuter students) and the relationship between student background variables, academic variables, environmental variables, and psychological outcomes on dropout. Seidman (2005) discussed the impact of early intervention using the Seidman formula, which aims to identify students who need academic or social assistance as soon as possible to increase retention. Tinto (1993) extended Durkheim (1961)'s theory of suicide and Van Gennep (1960)'s three stages of establishing membership in traditional societies to develop one of the most cited theories of student persistence. His model provided a longitudinal perspective and process that highlighted the impact of social and academic integration on student departure.

The overarching idea is that students who are generally more integrated academically and socially tend to have better outcomes, such as persisting and degree completion. Several studies have found the impact of social and academic integration on student retention and graduation rates (Barbera et al., 2020; Braxton, 2004; Dolan, 2019; Fung, 2010; Tinto, 1993; Spady, 1970). For two-year institutions, it appears that academic integration is more impactful than social integration (compared to four-year students) as two-year institutions serve more part-time students, older students with familial duties, and working students who may not have time to interact socially (Bean & Metzner, 1985; Braxton, 2004; Yu, 2015). Subsequently, it has been found that faculty interactions and classroom experiences heavily impact student success for students attending two-year institutions (Tinto, 2012a).

CHAPTER III

METHODS

Literature Search and Database Development

The following databases were used for searching articles: Education Resources Information Center (ERIC) for education-related articles, PsychINFO for behavioral and social science related articles, and ProQuest to include non-peer reviewed studies.

The following keywords were used with the "AND" and "OR" functions. Each term within a column was searched using the "OR" function and "AND" was used between columns. A narrow search was first conducted using columns 1-7 from Table 3 in Appendix A. Then, column 7 was dropped, and columns 2 and 3 were dropped in two subsequent searches to broaden the search. One search was conducted using Table 2 in Appendix A using all key terms. Qualitative studies (key terms: qualitative methods OR qualitative research OR qualitative study OR interview) were excluded. Lastly, only linked full text studies were searched and included.

Inclusion/Exclusion Criteria & Coding Procedure

After all studies from the search were collected, titles were screened, and each study was briefly reviewed. To be included in the analysis, studies needed the following information: (1) contain at least 1 student success metric relating to retention or persistence; (2) contain a full correlation matrix; (3) publication type was either a journal article or doctoral dissertation; (3) published in English; (4) U.S. based study; (5) included at least either student background characteristics, or institutional characteristic measures; (6) included both academic and social integrations measures; and, (7) population was comprised of only two year college students.

Studies were excluded if they were not sufficient to the purpose of this dissertation using the following exclusion criteria: (1) qualitative studies; (2) population was comprised of four year students only or a mix of two year and four year students; (1) non-academic papers or articles (e.g., magazines, videos, audio, newspapers, wire feeds, blogs, podcasts); (2) did not include student success measures relating to persistence or retention; (3) using academic performance (such as GPA) as a success measure (4) did not contain student background, institutional characteristics, academic or social integration information; (4) studies that did not include a full correlation matrix; (5) international studies.

An initial search gathered 454 studies. There were 13 additional studies found by searching through references and other related studies. A total of 398 studies remained after duplicates were removed. After initial title and abstract screening, 351 studies were excluded, and 47 studies remained for full-text assessment. Of the 47 studies, a remaining 25 studies were determined to contain sufficient information to be coded. See Figure 4 for the PRISMA diagram (Page et.al, 2021).

For the coding procedure, variables were coded based on how each study classified each construct. If the authors did not clearly define constructs, definitions based on the literature review were used to code studies. Any disagreements between definitions were defaulted to author’s definitions in the study. From each study, the correlations between exogenous and endogenous variables were obtained. The arithmetic mean was used to combine correlations if multiple variables were reported to measure the exogenous/endogenous variables unless otherwise specified. If a study described a procedure for calculating the correlation between exogenous/endogenous variables, it was used instead of the arithmetic mean.

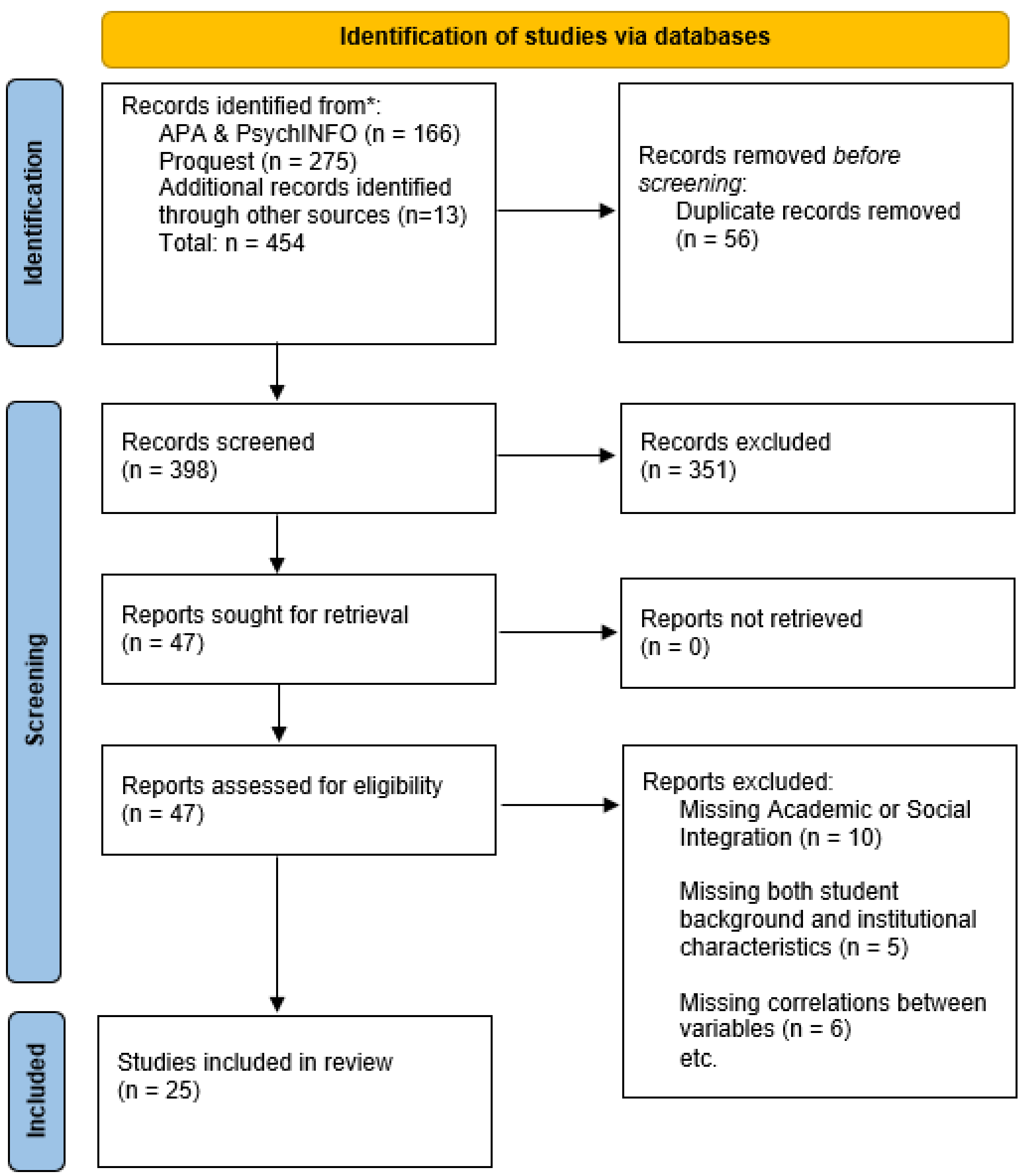

Figure 4: PRISMA diagram

Description of the Sample

A total of 25 studies were included in this MASEM. The year of publication ranged from 1985 to 2013, with the median and mean year of publication 2000 and 2001, respectively. The total sample size across studies was 15,274. There were 22 dissertations and 3 peer-reviewed journal articles.

Student Success

Of the 25 studies in the sample, 60% of studies (n = 15) reported subsequent semester and/or year enrollment (persistence) as a student success measure (Bengfort, 2012; Brown, 2007; Henningsen, 2003; Hillard, 1996; Jumpeter, 2005; Myers, 2001; Nakajima, 2008; Nippert, 2000; Nora, 1985; Pascarella et al, 1986; Pearl, 1993; Pietropaolo, 1994; Santos-George, 2012; Stryker, 1997; White, 1998).

36% of the studies (n = 9) used degree attainment/graduation as a student success measure (Barnhart, 2011; Bengfort, 2012; McNeil, 1997; Nakajima, 2008; Napoli & Wortman, 1998; Nippert, 2000; Pascarella et al 1986; Santos-George, 2012; Showalter, 2002). 8% of the studies (n = 2) used persistence within a course as a success measure (Aycock, 2011; Hoffman, 1998) and 8% (n=2) used transferring to another higher education institution as a success measure (Kraemer, 1993; Santos-George, 2012).

Student Background Characteristics

Several variables were reported to be classified as student background characteristics. The most common variables included were age (68% of studies, n = 17), gender (60% of studies, n = 15), and ethnicity/race (64% of studies, n = 16) (Aycock, 2011; Barnhart, 2011; Bengfort, 2012, Brown, 2007; Damon, 1996; Hackett, 2011;

Hillard, 1996; Hoffman, 1998; McNeil, 1997; Myers, 2001; Nakajima, 2008; Napoli & Wortman, 1998; Nippert, 2000; Nora, 1985; Pascarella et al, 1986; Pearl, 1993; Pietropaolo, 1994; Santos-George, 2012, Showalter, 2002; Stryker, 1997; Tovar, 2013; White, 1998).

The next most frequently reported variables were related to prior high school performance metrics and other pre-college entry characteristics. 48% of the studies (n = 12) included variables such as high school GPA, high school ranking, prior performance on English and Math assessments, prior skill and abilities, and high school diploma as student background characteristics (Aycock, 2011; Brown, 2007; Damon, 1996; McNeil, 1997; Nakajima, 2008; Napoli & Wortman, 1998; Nippert, 2000; Nora, 1985; Pascarella et al, 1986; Pearl, 1993; Santos-George, 2012). 44% of the studies (n = 11) reported pre-college entry characteristics such as expected degree earned, major at the time of enrollment, and enrollment status (Barnhart, 2011; Brown, 2007; Damon, 1996; McNeil, 1997; Nakajima, 2008; Napoli & Wortman, 1998; Nora, 1985; Pascarella et al, 1986; Pearl, 1993; Pietropaolo, 1994; White, 1998).

32% (n = 8) of studies included pre-college standardized assessments (e.g., placement exams, pre-tests, ACT scores, SAT scores) as student background characteristics (Hoffman; 1998; Jumpeter, 2005; Kraemer, 1993; Myers, 2001; Napoli & Wortman, 1998; Santos-George, 2012; Showalter, 2002; Stryker, 1997).

28% (n = 7) of studies included other external variables such as parental education, marital status, housing status, and socio-economic status as background variables (Barnhart, 2011; Brown, 2007; Napoli & Wortman, 1998; Nora, 1985; Pietropaolo, 1994; Santos-George, 2012).

Lastly, 12% of studies (n = 3) reported additional variables such as 1st generation status, prior on campus experiences, and high school social accomplishments as background characteristics (Aycock, 2011; Damon, 1996; Pascarella, 1986).

Institutional Characteristics

Institutional characteristics was the only construct to have missing correlations. Of the 25 studies, 8 did not report institutional characteristics. Of the 17 studies that did report institutional characteristics, almost 65% of the studies (n = 11) reported student satisfaction measures with campus characteristics such as campus size, reputation, campus climate, and academic/social satisfaction measures (Brown, 2007; Hackett, 2011; Jumpeter, 2005; Kraemer, 1993; Napoli & Wortman, 1998; Nippert, 2000; Nora, 1985; Pascarella et al, 1986; Pearl, 1993; Tovar, 2013; White, 1998).

Roughly 35% of the studies (n = 6) reported use and satisfaction with advising, counseling, academic and social services, and course availability as institutional characteristics (Aycock, 2011; Bengfort, 2012; Hillard, 1996; McNeil, 1997; Nakajima, 2008; Stryker, 1997).

Academic Integration

Academic Integration was defined in five major ways. 60% (n = 15 studies) included faculty-student interactions (both inside and outside of the classroom) in the definition of academic integration (Aycock, 2011; Barnhart, 2011; Bengfort, 2012; Brown, 2007; Hackett, 2011; Henningsen, 2003; Hillard, 1996; Hoffman, 1998;

Jumpeter, 2005; McNeil, 1997; Myers, 2001; Nakajima, 2008; Santos-George, 2012; Showalter, 2002; White, 1998).

48% of the studies (n = 12) included academic experiences and activities such as meeting with an advisor, participating in class, participating in study groups and learning communities, and joining an honor society (Barnhart, 2011; Bengfort, 2012; Brown, 2007; Damon, 1996; Kraemer, 1993; Napoli & Wortman, 1998; Nippert, 2000; Pascarella et al, 1986; Pearl, 1993; Stryker, 1997; Tovar, 2013). 48% of the studies (n = 12) included college GPA or grades as measures of academic integration (Henningsen, 2003; Hillard, 1996; Jumpeter, 2005; Kraemer, 1993; Najajima, 2008; Napoli & Wortman, 1998; Nippert, 2000; Nora, 1985; Pascarella et al, 1986; Pearl, 1993; Stryker, 1997; Tovar, 2013).

28% (n = 7) of the studies contained academic and intellectual development measures (e.g., study behaviors, intellectual growth questions) (Damon, 1996; Henningsen, 2003; Hoffman, 1998; Jumpeter, 2005; Nakajima, 2008; Napoli & Wortman, 1998; White, 1998). Lastly, 8% (n = 2) of the studies reported hours/credits enrolled as part of academic integration (Nora, 1985; Pietropaolo, 1994).

Social Integration

Social integration measures generally fell within three main categories. 72% (n = 18) of the studies reported participation and/or satisfaction in extracurriculars and social events on campus (e.g., clubs, sports, fine arts, fraternity/sorority) was a measure of social integration (Aycock, 2011; Barnhart, 2011; Bengfort, 2012; Brown, 2008; Damon, 1996; Hackett, 2011; Henningsen, 2003; Jumpeter, 2005; Kraemer, 1993; Myers, 22001;

Nakajima, 2008; Napoli & Wortman, 1998; Nippert, 2000; Pearl, 1993; Showalter, 2002; Stryker, 1997; Tovar, 2013; White, 1998).

56% of studies (n = 14) reported perceptions of and quality of the interactions with friends and peer groups (Brown, 2007; Damon, 1996; Hackett, 2011; Hillard, 1996; Hoffman, 1998; Jumpeter, 2005; Kraemer, 1993; McNeil, 1997; Napoli & Wortman, 1998); Nora, 1985; Pascarella et al, 1986; Pearl, 1993; Pietropaolo, 1994; Tovar, 2013).

32% of the studies (n = 8) reported informal faculty and staff interactions as social integration (Hillard, 1996; Kraemer, 1993; McNeil, 1997; Napoli & Wortman, 1998; Nora, 1985; Pascarella et al, 1986; Peal, 1993; Santos-George, 2012, White, 1998). Lastly, 8% of the studies (n = 2) reported sense of belonging and self-efficacy measures as social integration (Henningsen, 2003; Myers, 2001).

Data Analysis

After coding the studies, a full correlation matrix was obtained from each study. Then, descriptive statistics were calculated (sample size, mean, standard deviation, five-number summary, skewness and kurtosis) and QQ-plots and histograms were generated to observe the data for any non-normality. Univariate normality tests (e.g., Shapiro-Wilk test and Anderson-Darling test) as well as multivariate normality tests (Henze-Zirkler, Mardia's Skewness and Kurtosis, Royston, Doornik-Hansen, and E-Statistic) were conducted as well to check for any normality concerns as well.

Then, eight different approaches were used to conduct a MASEM; fixed-effect univariate *r* approach, random-effects univariate *r* approach, fixed-effect univariate *z* approach, random-effects univariate *z* approach, fixed-effect TSSEM approach, random-effects TSSEM approach, fixed-effect OSMASEM approach, and random-effects OSMASEM approach. See Chapter 2 for descriptions of each method.

In the first stage for the univariate *r*, univariate *z*, and TSSEM MASEM approaches, pooled correlations were calculated and a forest plot was constructed to visually inspect the relative strength and heterogeneity of each set of correlations. Heterogeneity indices (Q-statistic, I^2 index, estimated τ^2, and estimated τ) were calculated to estimate heterogeneity between studies. Then, a path model was fitted to the data. For the fixed-effect TSSEM approach, multi-group SEM was used to determine if the correlation matrices were the same across studies. As the random-effects OSMASEM is a single step, the estimated τ^2 was calculated. The fixed-effect OSMASEM also provided multi-group SEM statistics to determine homogeneity across studies. Differences between each method (fixed effect vs random effects and univariate vs multivariate) were calculated for comparisons.

In the second stage of the univariate *r*, univariate *z*, and TSSEM approaches as well as the OSMASEM approach, goodness-of-fit indices, parameter estimates, standard errors, and confidence intervals were compared by finding the difference between each of the estimates.

Lastly, publication bias was examined using funnel plots, test for asymmetry (Egger's Regression Test) and trim-and-fill analysis. A funnel plot is a graph that displays the relationship between an effect size and standard error. Studies with larger sample sizes and smaller standard errors are plotted towards the top of the graph, whereas studies with smaller sample sizes and larger standard errors are towards the bottom of the graph (Cooper et al, 2019). If there is no publication bias, studies will be scattered symmetrically around the mean effect; if there is asymmetry, this suggests that there may be publication bias. In the test for asymmetry (or Egger's Regression Test (Egger etg al, 1997), a regression line is fitted through the funnel plot. If the slope of the regression line is statistically significant, it can be inferred that there is asymmetry in the funnel plot. A third way to assess publication bias is using Duval and Tweedie (2000)'s trim-and-fill procedure. This test first "trims" studies that the procedure determines is causing the asymmetry, followed by "filling in" studies to achieve asymmetry.

Sensitivity analysis was also conducted using the leave-1-out method and by calculating three measures to determine possible influential points: studentized residual, difference in fits, and Cook's d. The studentized residual calculates the standardized difference in the observed effect size and predicated effect size had the influential point been removed. For the univariate case, influential points may have absolute standardized residuals greater than 3. The difference in fits value represents the number of standard deviations the predicated effect size changes for an individual observation after excluding the influential effect size. For the univariate case and for small to medium data sets, a difference in fits value greater than 1 would be identified as an influential point (Neter et al, 2005). Lastly, Cook's d values greater than $\chi^2_{p,0.50}$ could be identified as an influential point (Jensen & Ramirez, 1998). Since univariate analysis is conducted, $p = 1$ and $\chi^2_{1,0.50} = 0.45$.

All analyses were conducted in R and Microsoft Excel.

CHAPTER IV

RESULTS

Descriptive Statistics

For reference, "SS" refers to student success, "AI" refers to academic integration, "SI" refers to social integration, "SBC" refers to student background characteristics, and "IC" refers to institutional characteristics.

Table 5 presents the descriptive statistics across each bivariate relationship. Skewness for each correlation is close to 0 (with the largest value being 0.87) and kurtosis being between -2 and 2. This suggests that each univariate set of data is not severely non-normal (Kline, 2016). Visual inspection of the data using univariate QQ-plots (Figure 6) and histograms (Figure 7) show approximately normal distributions as well except for the correlations between SS and SBC. The Shapiro-Wilk test found the correlations between SS and SBC were not normal, however the *p*-value was 0.0464 and the significance level was set at 0.05 (See Table 6).

The results for multivariate normality tests suggest that the data is multivariate normal. See Table 7 for the multivariate normality test results.

Comparisons of Pooled Correlation Matrices ("Stage 1")

For the univariate and TSSEM approaches, a pooled correlation matrix was estimated before fitting the path model. Table 8 presents the respective parameter estimates, standard errors, and 95% confidences. The pooled correlation matrices for each method can be found in Appendix C. For the univariate *z* approach, all *z* values and standard errors have been converted back to *r'*s.

All but two relationships (between SBC/IC and SI/SBC) were statistically significant. However, relationships were generally weak (correlations less than 0.30) with the largest correlation being a moderate relationship between academic integration and social integration ($r = 0.264, SE = 0.0074, p < 0.0001$), and social integration and institutional characteristics ($r = 0.2575, SE = 0.0405, p < 0.0001$). The weakest correlations were between student background characteristics and institutional characteristics ($r = -0.0052, SE = 0.0097, p = 0.5952$).

Comparison Between Fixed-Effect and Random-Effects Models

The difference in parameter estimates and standard errors between fixed-effect and random-effects models can be found on Table 9. The differences between fixed-effect and random-effects models are small (at most to two decimal places) and mostly negative, indicating smaller fixed-effect estimates compared to random effects. The differences in summary correlations ranged from -0.083 to 0.0203. The most notable differences across the univariate and multivariate methods were found for the SI/IC relationship.

The standard error for random effects was larger compared to the fixed-effect models, although the differences are small. Although the difference in standard error between the fixed-effect and random-effects TSSEM is positive, the difference is to four decimal places. As the correlation between SBC and IC is close to 0, we would expected the standard errors to be very similar if not the same. The differences ranged from -0.041 to 0.0006.

When it comes to significant effects, fixed-effect methods only found the relationship between SBC/IC as insignificant, whereas random-effects methods found this summary effect as well as SI/SBC as insignificant. Overall, majority of the summary correlations were statistically significant.

Comparison Between Univariate and Multivariate Approaches

The difference in parameter estimates and standard errors across univariate and multivariate methods are provided in Table 10. Within the fixed-effect methods, the differences in summary correlations between univariate and multivariate methods ranged between -0.0054 and 0.0488. The differences are very small between estimates (at most two decimal places), however the univariate *r* approach appears to produce slightly larger estimates compared to the univariate *z* approach and TSSEM approach. This trend is reserved for random-effects models.

Within the random-effects methods, the differences in summary effects between univariate and multivariate methods ranged between -0.0115 and 0.0148. Notable differences were between univariate *r* and *z* approaches (with univariate r being smaller than univariate *z*) and the univariate *z* and TSSEM approach (with univariate z tending to be larger).

The trend in differences between standard errors is the same for the univariate *r*/univariate *z* approaches across model assumptions and reversed between fixed-effect and random-effects models for the univariate *r*/TSSEM approaches. Overall, the differences in standard errors across fixed-effect methods are virtually 0 (differences to four decimal places). The differences ranged from -0.0008 to 0.0005. The differences in standard error across random-effects models ranged from -0.0070 to 0.0096.

Comparison of Heterogeneity Measures

Forest plots are provided in Appendix G Figures 9-18. The forest plot of SI/IC showed the most heterogeneity between studies with individual study correlations ranging from -0.06 and 0.66. The forest plots of SBC/IC, SI/SBC, and SS/IC showed the least heterogeneity between studies with individual study correlations ranging from -0.08 to 0.013, -0.09 to 0.16, and -0.04 to 0.17 respectively.

Table 11 provides the heterogeneity measures (Q-statistic, I^2, estimated τ^2, and estimated τ) for the random-effects approaches. Tables 14-16 provide the multigroup comparisons for the fixed-effect TSSEM approach as well as LR statistic comparisons for the fixed-effect OSAMSEM approach.

For the univariate approaches, the test for homogeneity using the Q-statistic suggests that the null hypothesis that correlations are the same across studies is rejected (except for SBC/IC; the results of the hypothesis test fail to reject the null hypothesis). For the TSSEM approach, which uses multivariate meta-analysis in the first stage, also calculated a Q-statistic and p-value which rejected the null hypothesis. Using either multivariate or univariate methods appears to generally suggest that correlations are not the same across studies.

Next, the I^2 indices are also similar across univariate/multivariate methods, although there are larger discrepancies compared to the differences in correlation estimates and standard errors. See Table 12 for differences in I^2 values. The I^2 value can be interpreted as the percentage of variability between studies that is not due to sampling error. The univariate approaches seem to have larger I^2 values compared to the multivariate TSSEM method; in some cases, being almost 7-9% larger compared to the TSSEM I^2. Six out of the 10 variables had I^2 values higher than 89%. There were the relationships between: SS/AI, SS/SI, SS/IC, AI/SI, AI/IC, and SI/IC. Three out of the 10 variables had I^2 values that were between 37-61%. One variable, SBC/IC, had an I^2 value close to zero.

The differences in the univariate vs multivariate estimated τ^2 (and consequently estimated τ) are very small (almost 3 decimal places in most cases). The differences are smallest between the two multivariate approaches. See Table 14 for the differences in estimated τ^2 and estimated τ values.

The multi-group SEM from the TSSEM FE approach also suggests heterogeneity amongst the correlation matrices as the model fit is poor (see Table 15), suggesting the constraints are not equal across studies. Additionally, the OSMASEM FE multi-group SEM as well as the LR statistic comparing the saturated model (step 0) with the multigroup SEM (step 1) suggest that the correlation matrices are not homogeneous (see Tables 16 and 17, respectively).

Overall, the heterogeneity measures across model assumptions and across univariate/multivariate methods suggest that the correlation matrices are not homogeneous across studies. The estimated τ values suggest that some relationships have significant variation (SI/IC, AI/IC, AI/SI, SS/IC, SS/SI, SS/AI), as the values are great than 0.10.

Comparison of SEM ("Stage 2")

The goodness-of-fit indices are provided in Table 19, with the parameter estimates, standard errors, and confidence intervals provided on Table 21. Table 20 provides estimated variances and covariances.

The goodness-of-fit indices for univariate approaches are virtually the same between fixed- and random-effects models (with only a 0.001-unit difference between TLI values) and are similar between multivariate approaches. Each method suggests good fit of the model of the data.

Paths that were statistically significant across methods and model types were: SS/AI, AI/IC, and SI/IC. Within fixed-effect and random-effects models, there weren't many differences in the number of significant paths. However, the number of significant paths were very different between univariate and multivariate methods. The largest path estimates across all methods were SS/AI, AI/IC, and SI/IC. The smallest path estimate was SI/SBC.

Comparison Between Fixed- and Random-Effects Models

Table 17 provides the differences in parameter estimates and standard errors between fixed-effect and random-effects models. It seems that random-effect models are slightly larger, in both parameter estimates and standard errors, however the differences are very small (two decimal places), which is similar to stage 1.

Table 14 provides the differences in estimated covariances and variances. Again, the differences are small, and the fixed-effect models provide slightly larger estimates compared to the random-effects model.

When it comes to significant paths, results are similar. The fixed-effect and random-effects univariate r model produced the same number of significant paths. The remaining fixed-effect and random-effect comparisons produced at most 2 paths that were different.

Comparison Between Univariate and Multivariate Approaches

Table 22 and 23 provides the differences in parameter estimates and standard errors between univariate and multivariate methods. The differences in path estimates across univariate and multivariate methods is very small (to three decimal places for random-effects models and two decimal places for fixed-effect models). The path estimates within multivariate approaches are virtually identical (-0.001 was the only difference found between the TSSEM and OSMASEM approach). One most notable observation is that the univariate standard errors are slightly larger (by two decimal places) compared to the multivariate approaches. The standard errors are virtually identical for with univariate approaches and within multivariate approaches.

Because the sample size differs considerably between univariate and multivariate methods, the number of significant paths varied substantially between the univariate and multivariate approaches (see confidence intervals in Table 21). For univariate approaches, at most 3 out of the 8 paths are *significant* (SS/AI, AI/IC, and SI/IC). For the multivariate approaches, at most two paths are *nonsignificant* (SS/IC and SS/SBC). This is a significant difference when it comes to interpretation of the results.

Table 24 provides the differences in estimated covariances and variances. Again, the differences are small, however for estimated variances, univariate approaches seemed to be smaller compared to the multivariate approaches.

Results Pertaining to College Persistence

To interpret the results of the MASEM in the context of college persistence, the random-effects OSMASEM is used. There is a theoretical justification that the correlations do not originate from a single population as different studies included different measures for each construct (see Chapter 3). This is supported by the heterogeneity measures (large Q statistics with small p-values, moderate to high I^2 values, and non-zero τ^2 and τ estimates) and the forest plots. Additionally, a multivariate approach should be used because several correlations are being reported from each study; therefore, there is a dependency in those relationships that otherwise would not be found in independent relationships.

The relationship between student background characteristics on academic integration was significant ($r = 0.042, p = 0.0011$). The relationship between student background characteristics and social integration was not significant ($r = 0.013, p = 0.2534$). The relationship between student characteristics on student success was significant ($r = 0.035), p = 0.0019$). The relationship between institutional characteristics on academic integration was significant ($r = 0.116, p = 0.0013$). The relationship between institutional characteristics and social integration was significant

($r = 0.243, p < 0.0001$). The relationship between institutional characteristics and student success was not significant ($r = 0.054, p = 0.0938$). The relationship between academic integration on student success was significant ($r = 0.106, p = 0.0011$). The relationship between social integration and student success was significant ($r = 0.084, p = 0.0.0071$). See Figure 5 for fitted model with path estimates. An * next to the estimate means that it was statistically significant with significance level set to 0.05.

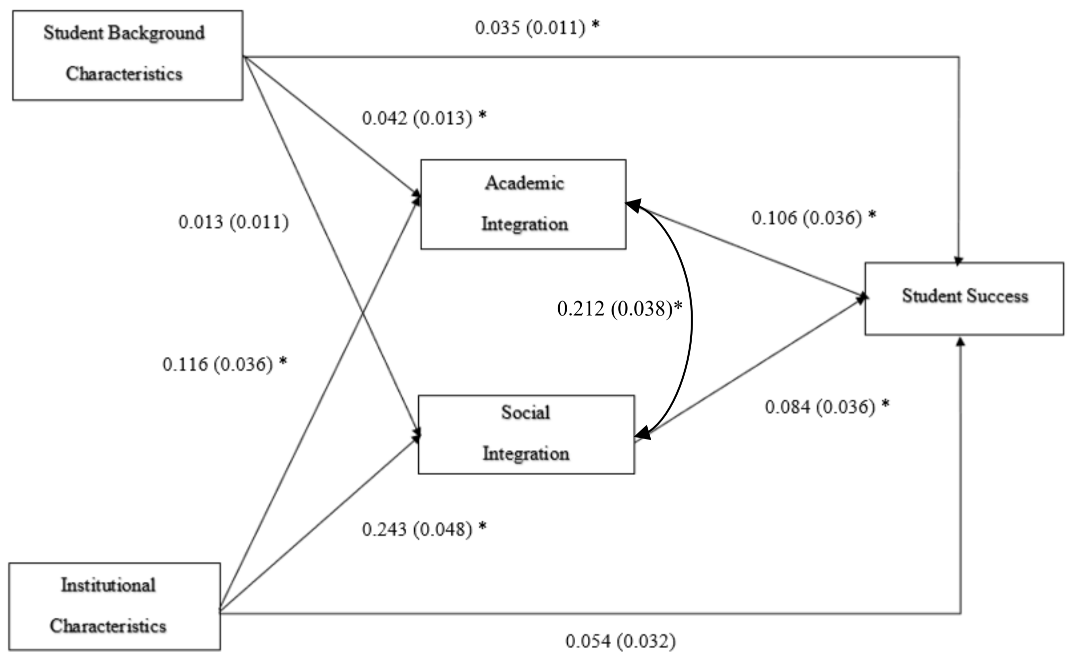

Figure 5: Fitted MASEM Model using OSMASEM RE

Publication Bias & Sensitivity Analysis

Three methods that were used to determine publication bias were funnel plots, test for asymmetry (Egger's Regression Test), and the Trim-and Fill method. For sensitivity analysis, multivariate chi-square QQ-plots were observed, as well as conducting the leave-1-out procedure and examining three measures of influence; the studentized residual (*rstudent*), difference in fits (*DFFITS*), and Cook's distance (*cook.d*).

As previously stated, funnel plots (found in Appendix H Figures 19-29) help visualize effect sizes and their corresponding standard errors. In terms of symmetry, all but one plot appears to be symmetrical; the AI/SI plot looks a bit asymmetrical. This is confirmed with the test for asymmetry – the only result that was significant was for AI/SI.

The trim-and-fill method identified additional missing studies in all but three analyses (SS/AI; AI/SI; SI/IC). For the correlation between SS/SI, there were an estimated 3 studies missing and the new estimated correlation is 0.1553 with standard error 0.0300 (originally 0.1250 and 0.0279, respectively). For the correlation between SS/SBC, there was an estimated 4 missing studies, with new correlation 0.0508 and standard error 0.0117 (originally 0.0407 and 0.0121, respectively). For the correlation between SS/IC, there was an estimated 3 missing studies with new correlation 0.1248 and standard error 0.0331 (originally 0.0884 and 0.0313, respectively). For the correlation between AI/SBC, there were an estimated 2 missing studies with new correlation 0.0518 and standard error 0.0148 (originally 0.0432 and 0.0143, respectively). The relationship between AI/IC had 6 missing studies (the largest) with new correlation 0.1936 and standard error 0.0402 (originally 0.1183 and 0.0391, respectively). For the correlation between SI/SBC, there were an estimated 2 missing studies with new correlation 0.0215 and standard error 0.0126 (originally 0.0134 and 0.0123, respectively). Lastly, there were 2 studies missing for SBC/IC with the new correlation 0.006 and standard error 0.0102 (originally -0.0052 and 0.0097, respectively).

Overall, it appears that there are minor concerns regarding publication bias. The new estimated correlations do not have large differences (other than the trim and fill procedure for AI/IC).

As for sensitivity analysis, the multivariate chi-square QQ-plot suggests that there may be 3 possible outliers (study 8, 12 and 17). The three measures of possible influence identified that possibly study 6 contains influential points. The correlation estimates without study 6 for SS/IC is 0.0696 (SE=0.03720), for AI/IC is 0.0888 (SE = 0.0281), and for SI/IC is 0.2190 (SE = 0.0453).

CHAPTER V

DISCUSSION & CONCLUSION

Summary of the Results & Discussion

The overall findings across methods showed significant summary correlations in "stage 1" for all relationships except SBC and IC (for fixed-effect models) in addition to SI and SBC (for random-effects models). All other relationships were statistically significant with $p < 0.05$. In "stage 2", each set of model fit indices indicated good model fit.

Regarding research question 1, the differences in parameter estimates, standard errors, and goodness-of-fit indices were small. Regarding heterogeneity, each approach suggests that the correlations across studies are not homogeneous. Although there were some notable differences in a few I^2 values, the measures of heterogeneity across studies were consistent with each other.

Although the differences in estimates were very small (to two or three decimal places), the significance of estimated correlations in stage 1 and path estimates in stage 2 varied significantly between univariate and multivariate approaches. This is mostly likely due to sample size, as the sample size for the univariate approaches was 349 and for multivariate approaches was 15274. As multivariate approaches use the sum of the samples, there will rarely be a chosen sample size for univariate approaches that will be larger than this, especially in situations with missing correlations. The harmonic mean was used based on the literature, however future research may want to explore simulation studies using different possibilities for sample size (Cheung & Chan, 2005; Sheng et al, 2016).

The findings in this dissertation are consistent with findings in other studies. The TSSEM RE and OSMASEM RE produced nearly identical results similar to Jak and Cheung (2020). The differences in fixed-effect/random-effect models and univariate/multivariate models being small has also been found through simulation studies (Cai & Fan, 2020; Zhang, 2011; Cho, 2015).

Regarding research question 2, it appears that academic integration ($r = 0.106, p < 0.05$) and social integration ($r = 0.084, p < 0.05$), on average, have a significant small mediating impact on student success. The effects are similar as well. Student background characteristics have a small but significant impact on student success ($r = 0.035, p < 0.05$) and academic integration ($r = 0.042, p < 0.05$); however, no significant impact on social integration ($r = 0.013, p > 0.05$). Institutional characteristics cannot predict student success alone ($r = 0.054, p > 0.05$), however, has significant effects on academic ($r = 0.116, p < 0.05$) and social integration ($r = 0.243, p < 0.05$). Lastly, there was a significance covariance between academic and social integration (cov =0.212, $p < 0.05$).

Limitations & Future Research

There are four limitations worth noting. First, dissertations comprised 22 out of the 25 studies. During the literature search, there were 398 articles that were screened and only 47 were included for full text review because a correlation matrix was included. However, studies that do not report a full correlation matrix is a common issue when attempting to conduct a MASEM (Sheng et al, 2016). This could be due to number of reasons (e.g., page limit for manuscripts), however because there is generally no requirement or protocol to produce a correlation matrix, this may remain an issue in collecting study information.

Another limitation involved the usefulness of an average effect. Although understanding general trends can be useful, when developing programs or interventions, it is more useful to know the present trend. The most recent study in the sample was published in 2013. Since almost 10 years have passed since this study, it is not unreasonable to assume that the relationships between academic and social integration and student success look different since major events have occurred (e.g., the rise of social media, pandemic). For example, there is a body of research that has suggests that social integration is less important compared to academic integration for two-year students due to having existing external commitments such as family commitments or work commitments (Crisp & Mena, 2012). However, with the rise in social mediums and the increased use of virtual services after the pandemic, traditional ways of integration socially (like clubs, extracurriculars) may be less of a barrier than they were before. Thus, the relationship between academic and social integration may not be the same in the new decade. Although this dissertation provides an overall effect and measures of heterogeneity, it may not be as useful as conducting another MASEM with subgroups by time/decade and conducting a mixed-effect MASEM or a multi-level MASEM.

Thirdly, only the author coded the studies. It is usually recommended that several coders go through the articles to enhance reliability and validity to prevent mono-method bias (Cooper et al, 2019).

Lastly, the fourth limitation is the precision in differences in estimates. The output of results for each method varied; some methods, like the TSSEM approach, reported estimates to three decimal places, whereas the univariate approaches reported up to 4. As the differences in estimates were generally very small, it is difficult to determine exact differences in approaches. Therefore, there should be particular attention to formatting as this can produce different results.

Conclusions

Borenstein et al (2021) stated the ethical imperative of comprehensively evaluating bodies of evidence instead of reviewing primary studies in isolation. With the current trend of studies using SEM, MASEM will also become more popular and thus a need to investigate its properties and protocols.

This dissertation found small difference across model assumptions and types of approaches, which may be somewhat of a relief to researchers. If an incorrect method is chosen, it appears that when within study sample sizes are large enough and correlations range from weak to moderate, the differences in these methods are generally small. However, for univariate approaches, choosing an appropriate sample size appears to be a nontrivial task. The largest difference between the univariate and multivariate approaches were significant pooled correlations and path estimates.

For two-year institutions, the results of this MASEM are generally encouraging. Although institutional characteristics alone could not predict student retention or persistence, it has had a meaningful impact student experience both academically and socially, which do have an impact on student retention and persistence. Additionally, students should feel assured that background characteristics have a small role to play in student success.

REFERENCES

Aycock, G. R. (2011). *A model of achievement and persistence for community college students: Identifying the impact of counseling* (Doctoral dissertation, The Claremont Graduate University).*

Astin, A. W., & Oseguera, L. (2012). Pre-college and institutional influences on degree attainment. *College student retention: Formula for student success*, 119-146.

Barbera, S. A., Berkshire, S. D., Boronat, C. B., & Kennedy, M. H. (2020). Review of undergraduate student retention and graduation since 2010: Patterns, predictions, and recommendations for 2020. *Journal of College Student Retention: Research, Theory & Practice*, *22*(2), 227-250.

Barnhart, D. (2011). *The relationship of academic and social integration to veterans' educational persistence*. The George Washington University.*

Baldwin, C., Bensimon, E. M., Dowd, A. C., & Kleiman, L. (2011). Measuring Student Success. *New Directions for Community Colleges*, *153*, 75-88.

Bean, J. P., & Metzner, B. S. (1985). A conceptual model of nontraditional undergraduate student attrition. *Review of educational Research*, *55*(4), 485-540.

Becker, B. J., Aloe, A. M., & Cheung, M. W. L. (2020). Meta-analysis of correlations, correlation matrices, and their functions. In *Handbook of meta-analysis* (pp. 347-370). Chapman and Hall/CRC.

Bengfort, R. R. (2012). *Exploring sense of community and persistence in the community college*. Morgan State University.*

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological bulletin*, *107*(2), 238.

Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). The handbook of research synthesis and meta-analysis. *LV Cooper, JC Hedges, & JC Valentine (2nd (Eds.), Effect sizes for continuous data. New York, NY: Russell Sage Foundation*.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis*. John Wiley & Sons.

Braunstein, A., & McGrath, M. (1997). The retention of freshman students: An examination of the assumptions, beliefs, and perceptions... *College Student Journal*, *31*(2), 188-200.

Braxton, J. M., Hirschy, A. S., & McClendon, S. A. (2004). Understanding and Reducing College Student Departure. *ASHE ERIC HIGHER EDUCATION REPORT*, *3*.

Brown, C. M. (2007). *An empirical test of the nontraditional undergraduate student attrition model using structural equation modeling* (Doctoral dissertation, Ohio 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*(1), 63-77.

Browne, M. W. (1982). Covariance structures. *Topics in applied multivariate analysis*, 72-141.

Browne M.W., & Cudeck, R. (1993) Alternative ways of assessing model fit in *Testing structural equation models* (ed. Bollen KA and Long JS). Sage Publications, Inc., Newbury Park, CA, pp.136-162

Cai, Z., & Fan, X. (2020). A comparison of fixed-effects and random-effects models for multivariate meta-analysis using an SEM approach. *Multivariate Behavioral Research*, *55*(6), 839-854.

Chen, L., Duchan, C., Durante, A., Kreiss, K., Merry, E. A., Robles, B. J., ... & Zabek, M. (2020). Report on the Economic Well-Being of US Households in 2018. *Reports and Studies*, 1-64.

Cheung, M. W. L., & Chan, W. (2004). Testing dependent correlation coefficients via structural equation modeling. *Organizational Research Methods*, *7*(2), 206-223.

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

Cheung, M. W. L. (2015). *Meta-analysis: A structural equation modeling approach*. John Wiley & Sons.

Cho, K. (2015). *Meta-analysis of factor analyses: Comparison of univariate and multivariate approaches using correlation matrices and factor loadings* (Doctoral dissertation, The Florida State University).

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics*, *10*(1), 101-129.

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2019). *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation.

Crisp, G., & Mina, L. (2012). The community college. *College student retention: Formula for student success*, *147*.

Damon, K. (1997). *Factors influencing community college student attrition: an application of Tinto's model at a public community college in Hawaii*. University of Southern California.*

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled clinical trials*, *7*(3), 177-188.

Dolan, A. A. (2019). *Synthesizing Undergraduate College Student Persistence: A Meta-analytic Structural Equation Model* (Doctoral dissertation, Kent State University).

Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455-463.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj*, *315*(7109), 629-634.

Fung, T. Y. H. (2010). *Analysis of graduation rates for four-year colleges: A model of institutional performance using IPEDS*. University of North Texas.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational researcher*, *5*(10), 3-8.

Hackett, C. D. (2011). *Predictors of commitment to graduation: An application toward African American male student athletes within the California community college system*. California Lutheran University.*

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

Henningsen, J. D. (2003). *Assessing the fit of Tinto's longitudinal model of institutional departure at a community college*. University of Central Florida.*

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in medicine*, *21*(11), 1539-1558.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *Bmj*, *327*(7414), 557-560.

Hillard, M. K. (1996). *An assessment of persistence and mattering among nontraditional, community college students*. University of Northern Colorado.*

Hoffman, J. L. (1998). *Application of Tinto's theoretical model of college withdrawal to developmental reading students at a two-year residential college*. Texas A&M University.*

Hom, P. W., Caranikas-Walker, F., Prussia, G. E., & Griffeth, R. W. (1992). A meta-analytical structural equations analysis of a model of employee turnover. *Journal of applied psychology*, *77*(6), 890.

Hussar, B., Zhang, J., Hein, S., Wang, K., Roberts, A., Cui, J., ... & Dilig, R. (2020). The Condition of Education 2020. NCES 2020-144. *National Center for Education Statistics*.

Jackson, D., White, I. R., & Riley, R. D. (2020). Multivariate meta-analysis. In *Handbook of meta-analysis* (pp. 163-186). Chapman and Hall/CRC.

Jak, S., & Cheung, M. W. L. (2018). Testing moderator hypotheses in meta-analytic structural equation modeling using subgroup analysis. *Behavior Research Methods*, *50*(4), 1359-1373.

Jak, S., & Cheung, M. W. L. (2020). Meta-analytic structural equation modeling with moderating effects on SEM parameters. *Psychological methods*, *25*(4), 430.

Jensen, D. R., & Ramirez, D. E. (1998). Detecting outliers with Cook's DI statistic. *Computing Science and Statistics*, *29*(1), 581-586.

Jumpeter, H. D. (2005). *The influence of student academic behaviors on the academic integration and the persistence of students at a two-year campus of a multi-campus university: An exploratory study*. The Pennsylvania State University.*

Kline, R. B. (2016). *Principles and practice of structural equation modeling*. Guilford publications.

Kraemer, B. A. (1995). Factors affecting Hispanic student transfer behavior. *Research in Higher Education*, *36*(3), 303-322.*

Ma, J., Pender, M., & Welch, M. (2019). Education pays 2019. *College Board*.

McNeill, K. M. (1997). *Determinants of different types of retention: Goal completion, institutional persistence, transfer and dropout at the community college*. University of Oregon.*

Mertes, S. J., & Hoover, R. E. (2014). Predictors of first-year retention in a community college. *Community College Journal of Research and Practice*, *38*(7), 651-660.

Morrison, L., & Silverman, L. (2012). Retention theories, models, and concepts. *College student retention: Formula for student success*, *61*.

Mortenson, T. G. (2012). Measurements of persistence. *College student retention: Formula for student success*, 35-60.

Mullin, C. M. (2012). Student success: Institutional and individual perspectives. *Community College Review*, *40*(2), 126-144.

Myers, R. L. (2001). *Persistence of technical degree seekers*. The University of Memphis.*

Nakajima, M. A. (2008). *What factors influence student persistence in the community college setting?*. University of Southern California.*

Napoli, A. R., & Wortman, P. M. (1998). Psychosocial factors related to retention and early departure of two-year community college students. *Research in higher education*, *39*(4), 419-455.*

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). Applied linear statistical models.

Nippert, K. (2000). Influences on the educational degree attainment of two-year college students. *Journal of College Student Retention: Research, Theory & Practice*, *2*(1), 29-40.*

Nora, A. (1987). Determinants of retention among Chicano college students: A structural model. *Research in higher education*, *26*(1), 31-59.*

Oort, F. J., & Jak, S. (2016). Maximum likelihood estimation in meta-analytic structural equation modeling. *Research synthesis methods*, *7*(2), 156-167.

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., … McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ (Clinical research ed.)*, *372*, n160. https://doi.org/10.1136/bmj.n160

Pascarella, E. T. (1986). Long-Term Persistence of Two-Year College Students. ASHE 1986 Annual Meeting Paper.*

Paule, Robert C., and John Mandel. "Consensus values and weighting factors." *Journal of research of the National Bureau of Standards* 87, no. 5 (1982): 377.

Pearl, A. B. (1993). *Extended freshman orientation participation and persistence behavior of students at a rural commuter community college*. State University of New York at Albany.*

Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *BMJ* 1904; 3:1243-1246

Pietropaolo, P. A. (1994). *The effects of an academic intervention on persistence of science/technology majors at a community college*. Syracuse University.*

Premack, S. L., & Hunter, J. E. (1988). Individual unionization decisions. *Psychological Bulletin*, *103*(2), 223.

Santos-George, A. A. (2012). *An empirical test of Tinto's integration framework for community colleges using structural equation modeling*. Illinois State University.*

Schneider, M. (2010). Finishing the First Lap: The Cost of First Year Student Attrition in America's Four Year Colleges and Universities. *American Institutes for Research*.

Seidman, A. (2005). Where we go from here. *College student retention: Formula for student success*, *295*.

Showalter, H. R. (2002). *Persistence of associate of arts degree students*. The University of Memphis.*

Spady, W. G. (1970). Dropouts from higher education: An interdisciplinary review and synthesis. *Interchange*, *1*(1), 64-85.

Stijnen, T., White, I. R., & Schmid, C. H. (2020). Analysis of univariate study-level summary data using normal models. In *Handbook of Meta-Analysis* (pp. 41-64). Chapman and Hall/CRC.

Stryker, S. (1997). *Predicting institutional persistence in first-year community college students*. Boston College.*

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of educational research*, *45*(1), 89-125.

Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition*. University of Chicago Press.

Tinto, V. (2012a). *Completing college: Rethinking institutional action*. University of Chicago Press.

Tinto, V. (2012b). Moving from theory to action: A model of institutional action for student success. *College student retention: Formula for student success*, *2*, 251-266.

Tovar, E. (2013). *A conceptual model on the impact of mattering, sense of belonging, Engagement/Involvement, and socio-academic integrative experiences on community college students' intent to persist*. The Claremont Graduate University.*

U.S. Department of Education, National Center for Education Statistics. (2019). *Digest of Education Statistics, 2018* (NCES 2020-009), Chapter 3.

Wagner, R. K. (1988). Causal relations between the development of phonological processing abilities and the acquisition of reading skills: A meta-analysis. *Merrill-Palmer Quarterly (1982-)*, 261-279.

White, E. A. (1998). *An analysis of the influence of the fit between learning styles and educational experiences on retention of selected community college students*. New York University.*

Yu, H. (2015). Student retention at two-year community colleges: A structural equation modeling approach. *International Journal of Continuing Education & Lifelong Learning*, *8*(1).

Yuan, K. H. (2016). Meta analytical structural equation modeling: comments on issues with current methods and viable alternatives. *Research synthesis methods*, *7*(2), 215-231.

Ziskin, M., Lucido, J., Torres, V., Chung, E., & Hossler, D. (2012). Securing the Future Retention Models in Community Colleges.

Zhang, Y. (2011). *Meta-analytic structural equation modeling (MASEM): comparison of the multivariate methods*. The Florida State University.

APPENDIX A

KEYWORDS

Table 2:First Set of Key Terms Used in Literature Search

Population	Retention Theory	Method	Student Success
Two-year college	Tinto	Regression	Persistence
Community college	Spady	Mediation	Retention
Two-year institution	Kamens	SEM	Student Success
Two year institution	Bean & Metzner	Path	Graduation
Two-year college		Correlation	
Two year college			
Junior college			

Table 3: Second Set of Key Terms Used in Literature Search

1	2	3	4	5	6	7
Population	**Student Background**	**Institutional Characteristics**	**Social Integration**	**Academic Integration**	**Student Success**	**method**
Two-year college	High school GPA	Campus size	social integration	academic integration	Persistence	Regression
Community college	ACT scores	Campus finances	social engagement	academic engagement	Retention	Mediation
Two-year institution	SAT score	Student services	sense of belonging	faculty interaction	Success	SEM
Two year institution	Gender	College satisfaction	social satisfaction	academic experience	Graduation	Structural model
Junior college	race	Percent adjunct	peer interaction	academic satisfaction	Dropout	Path model
Two year college	ethnicity	Faculty student ratio	extracurricular	advising	Attrition	correlation
	Pre-college characteristics	Financial aid	clubs	academic achievement	Completion	
	High school grades	expectations	student organizations	college GPA		
		selectivity	informal interactions	mentoring		
		Academic support	friends			
		Financial support				
		Social support				

APPENDIX B

CODING SHEET

Study ID: study identification number

Title: Title of the study

Author: Last names of the authors

Year: Year the study was published

Peer-reviewed? 1 for yes, 0 for no

Source Type: (e.g, journal article, dissertation)

Type of Model/Analysis: (e.g., path model, mediation analysis,

Instrument(s) used: list the surveys/instruments used in the study

Student Characteristics: list measures of student characteristics

Institutional Characteristics: list measures of institutional characteristics

Academic Integration: list measures of academic integration

Social Integration: list measures of social integration

Student Success: list measures of student success

Sample size: sample size reported in the study

Pg Number of Matrix: Page to find the correlation matrix

Full correlation matrix: Print and attach to coding sheet from study

APPENDIX C

POOLED CORRELATION MATRICES

Table 4: Pooled Correlation Matrices Across Methods

	SS	AI	SI	SBC	IC
Fixed-Effect - Univariate r					
	SS	AI	SI	SBC	IC
SS	1.0000				
AI	0.1444	1.0000			
SI	0.1196	0.2640	1.0000		
SBC	0.0182	0.0359	0.0236	1.0000	
IC	0.0694	0.1054	0.2229	-0.0052	1.0000
Random-Effects – Univariate r					
	SS	AI	SI	SBC	IC
SS	1.0000				
AI	0.1354	1.0000			
SI	0.1250	0.2437	1.0000		
SBC	0.0407	0.0432	0.0134	1.0000	
IC	0.0884	0.1183	0.2460	-0.0052	1.0000
Fixed-Effect - Univariate z					
	SS	AI	SI	SBC	IC
SS	1.0000				
AI	0.1290	1.0000			
SI	0.1131	0.2411	1.0000		
SBC	0.0438	0.0351	0.0216	1.0000	
IC	0.0644	0.0880	0.1746	-0.0052	1.0000
Random-Effects - Univariate z					
	SS	AI	SI	SBC	IC
SS	1.0000				
AI	0.1385	1.0000			
SI	0.1263	0.2512	1.0000		
SBC	0.0405	0.0427	0.0209	1.0000	
IC	0.0889	0.1215	0.2575	-0.0052	1.0000
Fixed-Effect - TSSEM					
	SS	AI	SI	SBC	IC
SS	1.0000				
AI	0.1297	1.0000			
SI	0.1146	0.2407	1.0000		
SBC	0.0439	0.0354	0.0174	1.0000	
IC	0.0698	0.0930	0.1741	-0.0021	1.0000
Random-Effects - TSSEM					
	SS	AI	SI	SBC	IC
SS	1				
AI	0.1342	1.0000			
SI	0.1237	0.2413	1.0000		
SBC	0.0404	0.0417	0.0113	1.0000	
IC	0.0863	0.1159	0.2427	-0.0045	1.0000

APPENDIX D

TABLES AND FIGURES

Table 5: Descriptive Statistics

Variables	n	Mean	St. Dev	Min	Q1	Median	Q3	Max	Skewness	Kurtosis
SS-AI	25	0.13	0.17	-0.22	0.01	0.11	0.27	0.48	0.09	-0.77
SS-SI	25	0.12	0.14	-0.12	0.04	0.12	0.23	0.36	0.03	-1.01
SS-SBC	25	0.04	0.06	-0.04	-0.01	0.02	0.07	0.17	0.71	0.46
SS-IC	17	0.09	0.13	-0.10	0.03	0.06	0.14	0.37	0.68	-0.16
AI-SI	25	0.24	0.20	-0.22	0.13	0.26	0.36	0.63	-0.29	-0.02
AI-SBC	25	0.05	0.07	-0.08	-0.01	0.03	0.07	0.20	0.72	-0.51
AI-IC	17	0.12	0.16	-0.19	0.03	0.11	0.15	0.56	0.87	1.62
SI-SBC	25	0.01	0.06	-0.09	-0.04	0.01	0.05	0.16	0.25	-0.34
SI-IC	17	0.25	0.20	-0.06	0.07	0.27	0.34	0.66	0.41	-0.91
SBC-IC	17	0.00	0.06	-0.09	-0.04	-0.01	0.01	0.13	0.71	-0.14

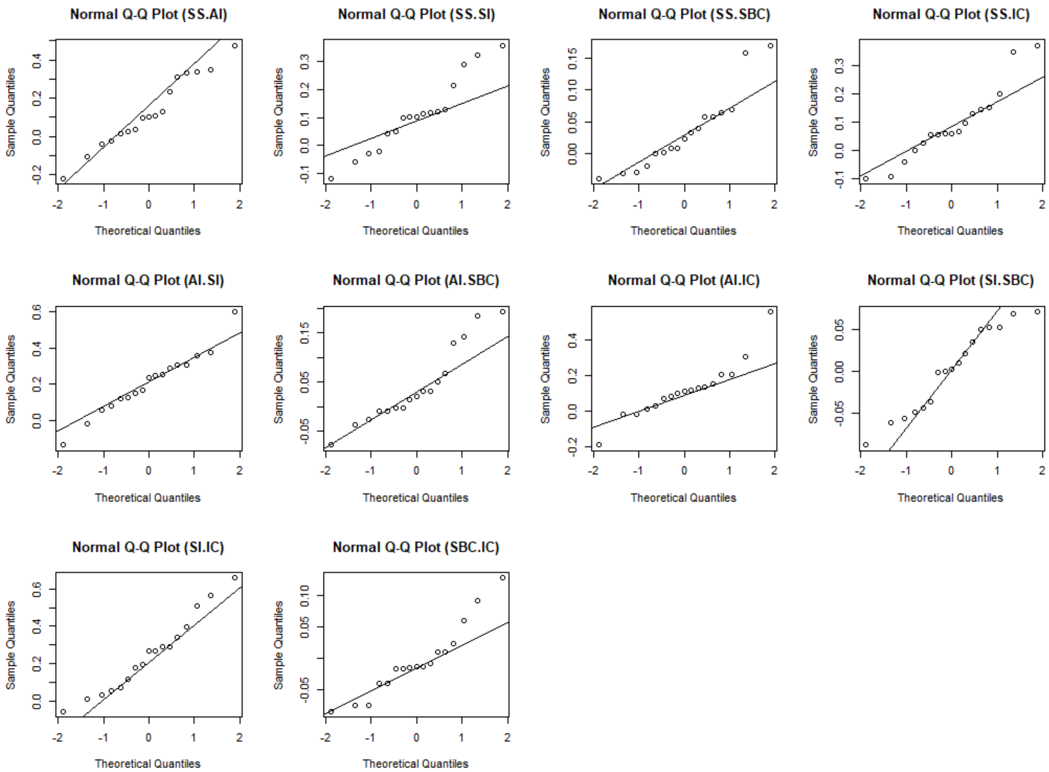

Figure 6: Univariate QQ-Plots

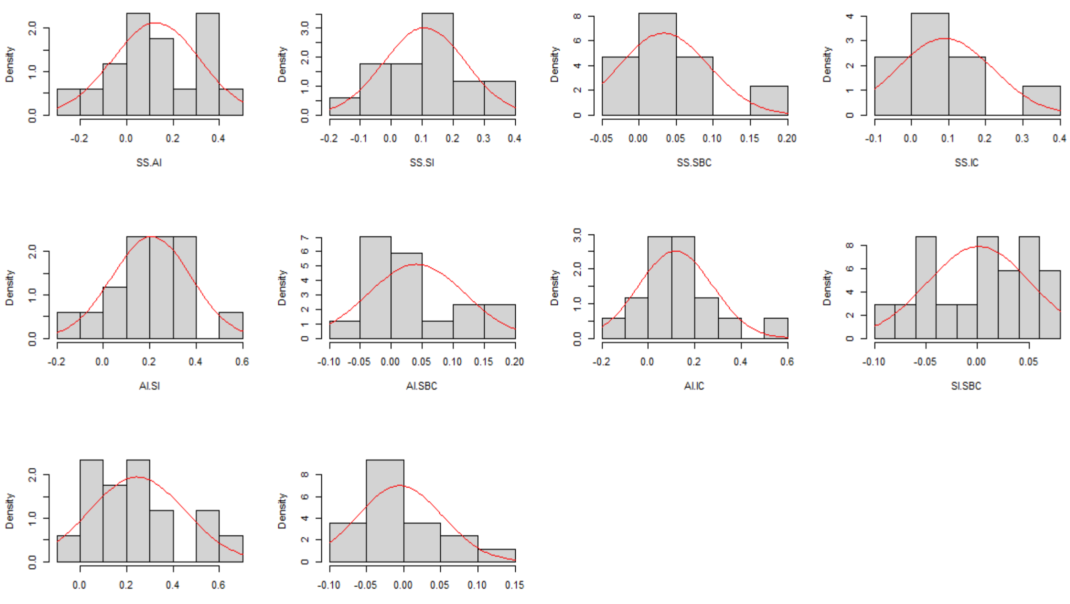

Figure 7: Univariate Histograms and Curves

Table 6: Univariate Normality Tests

Shapiro Wilk Test for Normality			
Variables	Statistic	p-value	Normality?
SS-AI	0.9688	0.7963	YES
SS-SI	0.9502	0.4605	YES
SS-SBC	0.8900	0.0464	No
SS-IC	0.9235	0.1694	YES
AI-SI	0.9780	0.9371	YES
AI-SBC	0.9129	0.1119	YES
AI-IC	0.9063	0.0868	YES
SI-SBC	0.9413	0.3344	YES
SI-IC	0.9600	0.6324	YES
SBC-IC	0.9225	0.1629	YES
Anderson Darling Test for Normality			
Variables	Statistic	p-value	Normality?
SS-AI	0.3015	0.5389	YES
SS-SI	0.4351	0.2645	YES
SS-SBC	0.6218	0.0878	YES
SS-IC	0.5171	0.1624	YES
AI-SI	0.2167	0.8131	YES
AI-SBC	0.6654	0.0677	YES
AI-IC	0.6483	0.0750	YES
SI-SBC	0.3744	0.3748	YES
SI-IC	0.2583	06720	YES
SBC-IC	0.5600	0.1247	YES

Table 7: Multivariate Tests of Normality

Test	Test Statistic	P-value	Normality?
Henze-Zirkler	0.9535	0.37805	YES
Mardia Skewness	249.1640	0.08617	YES
Mardia Kurtosis	-1.0010	0.3168	YES
Royston	16.9023	0.0785	YES
Doornik-Hansen	16.4561 (df=20)	0.6879	YES
E-statistic	1.5952	0.2830	YES

Table 8: Stage 1 Parameter Estimates, Standard Errors, and 95% Confidence Intervals

	Variables	Univariate r FE	Univariate r RE	Univariate z FE (converted to r)	Univariate z RE (converted to r)	TSSEM FE	TSSEM RE
Parameter Estimates	SS-AI	0.1444	0.1354	0.1291	0.1385	0.1297	0.1342
	SS-SI	0.1196	0.1250	0.1131	0.1263	0.1146	0.1238
	SS-SBC	0.0443	0.0407	0.0439	0.0405	0.0439	0.0406
	SS-IC	0.0694	0.0884	0.0644	0.0889	0.0698	0.0863
	AI-SI	0.264	0.2437	0.2411	0.2512	0.2407	0.2414
	AI-SBC	0.0359	0.0432	0.0351	0.0427	0.0354	0.0420
	AI-IC	0.1054	0.1183	0.0880	0.1215	0.0930	0.1159
	SI-SBC	0.0182	0.0134	0.0173	0.0126	0.0174	0.0124
	SI-IC	0.2229	0.2460	0.1746	0.2575	0.1741	0.2427
	SBC-IC	-0.0052	-0.0052	-0.0052	-0.0052	-0.0021	-0.0044
Standard Errors	SS-AI	0.0077	0.0347	0.0076	0.0304	0.0081	0.0330
	SS-SI	0.0078	0.0279	0.0076	0.0249	0.0081	0.0265
	SS-SBC	0.0081	0.0121	0.0078	0.0114	0.0081	0.0113
	SS-IC	0.0095	0.0313	0.0093	0.0278	0.0097	0.0292
	AI-SI	0.0074	0.0387	0.0069	0.0321	0.0077	0.0371
	AI-SBC	0.0081	0.0143	0.0078	0.0134	0.0081	0.0130
	AI-IC	0.0094	0.0391	0.0092	0.0347	0.0097	0.0362
	SI-SBC	0.0081	0.0123	0.0081	0.0120	0.0081	0.0113
	SI-IC	0.0089	0.0501	0.0088	0.0405	0.0096	0.0475
	SBC-IC	0.0097	0.0097	0.0093	0.0094	0.0097	0.0091
95% Confidence Intervals	SS-AI	(0.1293, 0.1594)	(0.0674, 0.2034)	(0.1142, 0.1439)	(0.0789, 0.1982)	(0.1138, 0.1455)	(0.0695,0.1989)
	SS-SI	(0.1042, 0.1350)	(0.0704, 0.1796)	(0.0981, 0.1281)	(0.0776, 0.1750)	(0.0988, 0.1304)	(0.0717,0.1758)
	SS-SBC	(0.0285, 0.0601)	(0.0171, 0.0644)	(0.0286, 0.0592)	(0.0181, 0.0629)	(0.0281, 0.0598)	(0.0184,0.0627)
	SS-IC	(0.0507, 0.0881)	(0.0271, 0.1496)	(0.0463, 0.0826)	(0.0345, 0.1433)	(0.0508, 0.0888)	(0.0291,0.1435)
	AI-SI	(0.2495, 0.2784)	(0.1679, 0.3196)	(0.2276, 0.2547)	(0.1883, 0.3141)	(0.2256, 0.2559)	(0.1685,0.3141)
	AI-SBC	(0.0201, 0.0517)	(0.0152, 0.0712)	(0.0197, 0.0504)	(0.0164, 0.0689)	(0.0195, 0.0512)	(0.0162,0.0672)
	AI-IC	(0.0870, 0.1238)	(0.0417, 0.1948)	(0.0701, 0.1061)	(0.0535, 0.1895)	(0.0740, 0.1121)	(0.0450,0.1868)
	SI-SBC	(0.0023, 0.0341)	(-0.0106, 0.0375)	(0.0014, 0.0332)	(-0.0109, 0.0361)	(0.0015, 0.0332)	(-0.0098, 0.0346)
	SI-IC	(0.2055, 0.2403)	(0.1478, 0.3441)	(0.1574, 0.1918)	(0.1781, 0.3370)	(0.1553, 0.1930)	(0.1495,0.3358)
	SBC-IC	(-0.0242, 0.0139)	(-0.0242, 0.0139)	(-0.0235, 0.0131)	(-0.02372, 0.0132)	(-0.0211, 0.0169)	(-0.0222,0.0134)

Table 9: Differences Between Fixed-Effect and Random-Effects Models

	Variables	Univariate r (FE vs RE)	Univariate z (FE vs RE)	TSSEM (FE vs RE)
Parameter Estimates	SS-AI	0.009	-0.0094	-0.0045
	SS-SI	-0.0054	-0.0132	-0.0092
	SS-SBC	0.0036	0.0034	0.0033
	SS-IC	-0.019	-0.0245	-0.0165
	AI-SI	0.0203	-0.0101	-0.0007
	AI-SBC	-0.0073	-0.0076	-0.0066
	AI-IC	-0.0129	-0.0335	-0.0229
	SI-SBC	0.0048	0.0007	0.005
	SI-IC	-0.0231	-0.0829	-0.0686
	SBC-IC	<0.0001	<0.0001	0.0023
Standard Errors	SS-AI	-0.027	-0.0228	-0.0249
	SS-SI	-0.0201	-0.0173	-0.0184
	SS-SBC	-0.004	-0.0036	-0.0032
	SS-IC	-0.0218	-0.0185	-0.0195
	AI-SI	-0.0313	-0.0252	-0.0294
	AI-SBC	-0.0062	-0.0056	-0.0049
	AI-IC	-0.0297	-0.0255	-0.0265
	SI-SBC	-0.0042	-0.0084	-0.0032
	SI-IC	-0.0412	-0.0317	-0.0379
	SBC-IC	<0.0001	-0.0001	0.0006

Table 10: Differences Across Univariate/Multivariate Methods

		Fixed-Effect Models			Random-Effects Model		
	Variables	Univariate r/Univariate z	Univariate r/TSSEM	Univariate z/TSSEM	Univariate r/Univariate z	Univariate r/TSSEM	Univariate z/TSSEM
Estimated Correlations	SS-AI	0.0153	0.0147	-0.0006	-0.0031	0.0012	0.0043
	SS-SI	0.0065	0.005	-0.0015	-0.0013	0.0012	0.0025
	SS-SBC	0.0004	0.0004	<0.0001	0.0002	0.0001	-0.0001
	SS-IC	0.005	-0.0004	-0.0054	-0.0005	0.0021	0.0026
	AI-SI	0.0229	0.0233	0.0004	-0.0075	0.0023	0.0098
	AI-SBC	0.0008	0.0005	-0.0003	0.0005	0.0012	0.0007
	AI-IC	0.0174	0.0124	-0.005	-0.0032	0.0024	0.0056
	SI-SBC	-0.0034	0.0008	0.0042	-0.0075	0.001	0.0085
	SI-IC	0.0483	0.0488	0.0005	-0.0115	0.0033	0.0148
	SBC-IC	<0.0001	-0.0031	-0.0031	<0.0001	-0.0008	-0.0008
Standard Errors	SS-AI	0.0001	-0.0004	-0.0005	0.0043	0.0017	-0.0026
	SS-SI	0.0002	-0.0003	-0.0005	0.003	0.0014	-0.0016
	SS-SBC	0.0003	<0.0001	-0.0003	0.0007	0.0008	0.0001
	SS-IC	0.0002	-0.0002	-0.0004	0.0035	0.0021	-0.0014
	AI-SI	0.0005	-0.0003	-0.0008	0.0066	0.0016	-0.0050
	AI-SBC	0.0003	<0.0001	-0.0003	0.0009	0.0013	0.0004
	AI-IC	0.0002	-0.0003	-0.0005	0.0044	0.0029	-0.0015
	SI-SBC	0.0003	<0.0001	-0.0003	-0.0039	0.001	0.0049
	SI-IC	0.0001	-0.0007	-0.0008	0.0096	0.0026	-0.0070
	SBC-IC	0.0004	<0.0001	-0.0004	0.0003	0.0006	0.0003

Table 11: Heterogeneity Measures

	Variables	Univariate r	Univariate z	TSSEM	OSMASEM
Q-statistic, p-value	SS-AI	$Q\ (24) = 614.6497, p < .0001$	$Q\ (24) = 534.7200, p < .0001$	$Q\ (208) = 1681.802, p < 0.0001$	NA
	SS-SI	$Q\ (24) = 293.8798, p < .0001$	$Q\ (24) = 275.6871, p < .0001$		
	SS-SBC	$Q\ (24) = 42.3377, p = 0.0118$	$Q\ (24) = 41.6013, p = 0.0143$		
	SS-IC	$Q\ (16) = 154.7665, p < .0001$	$Q\ (16) = 141.7254, p < .0001$		
	AI-SI	$Q\ (24) = 460.2918, p < .0001$	$Q\ (24) = 371.7710, p < .0001$		
	AI-SBC	$Q\ (24) = 56.1263, p = 0.0002$	$Q\ (24) = 54.4628, p = 0.0004$		
	AI-IC	$Q\ (16) = 290.1130, p < .0001$	$Q\ (16) = 219.6540, p < .0001$		
	SI-SBC	$Q\ (24) = 44.3958, p = 0.0069$	$Q\ (24) = 42.2111, p = 0.0122$		
	SI-IC	$Q\ (16) = 599.6174, p < .0001$	$Q\ (16) = 427.0553, p < .0001$		
	SBC-IC	$Q\ (16) = 19.3840, p = 0.2493$	$Q\ (16) = 19.0783, p = 0.2646$		
I^2	SS-AI	94.72%	94.56%	93.67%	NA
	SS-SI	91.29%	91.11%	89.96%	
	SS-SBC	45.67%	44.62%	37.52%	
	SS-IC	89.94%	89.81%	87.69%	
	AI-SI	96.14%	96.14%	95.44%	
	AI-SBC	61.63%	60.35%	52.39%	
	AI-IC	93.86%	94.16%	92.12%	
	SI-SBC	47.57%	44.55%	38.24%	
	SI-IC	96.72%	96.93%	95.57%	
	SBC-IC	0.02%	0.18%	00.00%	
Estimated τ^2	SS-AI	0.0275	0.0300	0.0246	0.0245
	SS-SI	0.0168	0.0177	0.0150	0.0149
	SS-SBC	0.0014	0.0014	0.0010	0.0010
	SS-IC	0.0144	0.0148	0.0122	0.0121
	AI-SI	0.0351	0.0429	0.0320	0.0318
	AI-SBC	0.0027	0.0026	0.0019	0.0018
	AI-IC	0.0238	0.0271	0.0199	0.0200
	SI-SBC	0.0015	0.0014	0.0011	0.0011
	SI-IC	0.0407	0.0049	0.0360	0.0362
	SBC-IC	0.0000	0.0000	0.0000	0.0000

Table 11: Heterogeneity Measures (cont'd)

Estimated τ	SS-AI	0.1658	0.1731	0.1568	0.1565
	SS-SI	0.1297	0.1329	0.1223	0.1221
	SS-SBC	0.0378	0.0373	0.0322	0.0316
	SS-IC	0.1200	0.1217	0.1103	0.1100
	AI-SI	0.1874	0.2071	0.1788	0.1783
	AI-SBC	0.0522	0.0512	0.0435	0.0424
	AI-IC	0.1541	0.1647	0.1410	0.1414
	SI-SBC	0.0392	0.0372	0.0332	0.0033
	SI-IC	0.2018	0.0699	0.1899	0.1903
	SBC-IC	0.0005	0.0017	0.0000	0.0000

Table 12: Difference in I^2 Values

Variables	uni r/uni z	uni r/TSSEM	uni z/TSSEM
SS-AI	0.16%	1.05%	0.89%
SS-SI	0.18%	1.33%	1.15%
SS-SBC	1.05%	8.15%	7.10%
SS-IC	0.13%	2.25%	2.12%
AI-SI	0.00%	0.70%	0.70%
AI-SBC	1.28%	9.24%	7.96%
AI-IC	-0.30%	1.74%	2.04%
SI-SBC	3.02%	9.33%	6.31%
SI-IC	-0.21%	1.15%	1.36%
SBC-IC	-0.16%	0.02%	0.18%

Table 13: Differences in τ^2 and τ

		Uni r vs Uni z	Uni r vs TSSEM	Uni r vs OSMASEM	Uni z vs TSSEM	Uni z vs OSMASEM	TSSEM vs OSMASEM
Estimated τ^2	SS-AI	-0.0025	0.0029	0.003	0.0054	0.0055	0.0001
	SS-SI	-0.0009	0.0018	0.0019	0.0027	0.0028	0.0001
	SS-SBC	<0.0001	0.0004	0.0004	0.0004	0.0004	<0.0001
	SS-IC	-0.0004	0.0022	0.0023	0.0026	0.0027	0.0001
	AI-SI	-0.0078	0.0031	0.0033	0.0109	0.0111	0.0002
	AI-SBC	0.0001	0.0008	0.0009	0.0007	0.0008	0.0001
	AI-IC	-0.0033	0.0039	0.0038	0.0072	0.0071	-0.0001
	SI-SBC	0.0001	0.0004	0.0004	0.0003	0.0003	<0.0001
	SI-IC	0.0358	0.0047	0.0045	-0.0311	-0.0313	-0.0002
	SBC-IC	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Estimated τ	SS-AI	-0.0073	0.009	0.0093	0.0163	0.0166	0.0003
	SS-SI	-0.0032	0.0074	0.0076	0.0106	0.0108	0.0002
	SS-SBC	0.0005	0.0056	0.0062	0.0051	0.0057	0.0006
	SS-IC	-0.0017	0.0097	0.01	0.0114	0.0117	0.0003
	AI-SI	-0.0197	0.0086	0.0091	0.0283	0.0288	0.0005
	AI-SBC	0.001	0.0087	0.0098	0.0077	0.0088	0.0011
	AI-IC	-0.0106	0.0131	0.0127	0.0237	0.0233	-0.0004
	SI-SBC	0.002	0.006	0.0359	0.004	0.0339	0.0299
	SI-IC	-0.0285	0.0119	0.0115	0.0404	-0.1204	-0.0004
	SBC-IC	-0.0012	0.0005	0.0005	0.0017	0.0017	<0.0001

Table 14: TSSEM FE Stage 1: Multigroup SEM Results

Goodness-of-Fit Indices	
n	15274
$\chi^2(df), p$ value	$\chi^2(208) = 1769.2878, p < 0.001$
CFI	0.5128
TLI	0.4894
RMSEA	0.1124
SRMR	0.0993

Table 15: OSMASEM FE - Step 1: Multigroup SEM Results

Goodness-of-Fit Indices	
n	15274
$\chi^2(df), p$ value	$\chi^2(208) = 1769.452, p < 0.001$
CFI	-
TLI	-
RMSEA	0.2217
SRMR	-

Table 16: OSMASEM FE Model Comparison

	-2LL	df	DiffLL	Diff DF	p-value
Model 0	68146.08	0			
Model 1	69915.53	208			
Model 2	69915.62	329			
Model 0 vs Model 1			1769.534	208	$p < 0.001$
Model 1 vs Model 2			0.0818	121	1

Table 17: Differences in Parameter Estimates Between Fixed-Effect and Random-Effects Models

	Variables	Uni-r FE vs RE	Uni-z FE vs RE	TSSEM FE vs RE	OSMASEM FE vs RE
Parameter Estimates	SS-AI	0.011	-0.004	-0.001	-0.001
	SS-SI	-0.006	-0.004	-0.003	-0.003
	SS-SBC	0.004	0.005	0.005	0.004
	SS-IC	-0.015	-0.013	-0.008	-0.008
	AI-SBC	-0.008	-0.007	-0.007	-0.006
	AI-IC	-0.012	-0.034	-0.023	-0.023
	SI-SBC	0.004	0.001	0.005	0.005
	SI-IC	-0.023	-0.083	-0.069	-0.069
Standard Errors	SS-AI	0.001	-0.001	-0.028	-0.028
	SS-SI	<0.0001	-0.001	-0.022	-0.023
	SS-SBC	<0.0001	<0.0001	-0.003	-0.003
	SS-IC	<0.0001	-0.001	-0.024	-0.022
	AI-SBC	<0.0001	<0.0001	-0.005	-0.005
	AI-IC	<0.0001	<0.0001	-0.026	-0.026
	SI-SBC	<0.0001	0.001	-0.003	-0.003
	SI-IC	<0.0001	0.001	-0.038	-0.038

Table 18: Differences in Estimated Covariance and Variances Between Fixed-Effect and Random-Effects Models

	Variables	Uni r FE vs RE	Uni z FE vs RE	TSSEM FE vs RE	OSMASEM FE vs RE
Cov	AI-SI	0.026	0.006	0.012	0.012
Var	SS	<0.001	0.005	<0.001	0.003
	AI	0.004	0.007	0.01	0.005
	SI	0.011	0.035	0.03	0.028

Table 19: "Stage 2" Goodness-of-Fit Indices

	Univariate R FE	Univariate R RE	Univariate z FE	Univariate z RE	TSSEM FE	TSSEM RE	OSMASEM FE	OSMASEM RE
n	349	349	349	349	15274	15274	15274	15274
$\chi^2(df), p$ value	$\chi^2(3) = 0.012, p = 1.000$	$\chi^2(3) = 0.012, p = 1.000$	$\chi^2(3) = 0.012, p = 1.000$	$\chi^2(3) = 0.012, p = 1.000$	$\chi^2(1) = 0.0818, p = 1.000$	$\chi^2(1) = 0.2377, p = 0.6259$	$\chi^2(121) = 0.0818, p = 1.000$	$\chi^2(1) = 0.140, p = 0.6766$
CFI	1.00	1.000	1.000	1.000	1.000	1.00	1.084	1.009
TLI	1.220	1.199	1.306	1.199	1.0062	1.0572	-	1.0907
RMSEA	0.000	0.000	0.000	0.000	0.000	0.000	0	0
SRMR	0.002	0.002	0.002	0.002	0.0009	0.0014	-	0.001

Table 20: Estimated Covariance and Variances

	Variables	Uni r FE	Uni r RE	Uni z FE	Uni z RE	TSSEM FE	TSSEM RE	OSMASEM FE	OSMASEM RE
Cov	AI - SI	0.239	0.213	0.224	0.218	0.224	0.212	0.224	0.212
	SS	0.966	0.966	0.970	0.965	0.97	0.97	0.972	0.969
Var	AI	0.985	0.981	0.988	0.981	0.99	0.98	0.990	0.985
	SI	0.947	0.936	0.966	0.931	0.97	0.94	0.969	0.941

Table 21: "Stage 2" Parameter Estimates, Standard Errors, and Confidence Intervals

	Variables	Univariate R FE	Univariate R RE	Univariate z FE	Univariate z RE	TSSEM FE	TSSEM RE	OSMASEM FE	OSMASEM RE
Parameter Estimates	SS-AI	0.118	0.107	0.105	0.109	0.105	0.106	0.105	0.106
	SS-SI	0.079	0.085	0.080	0.084	0.081	0.084	0.081	0.084
	SS-SBC	0.039	0.035	0.039	0.034	0.039	0.034	0.039	0.035
	SS-IC	0.040	0.055	0.041	0.054	0.046	0.054	0.046	0.054
	AI-SBC	0.036	0.044	0.036	0.043	0.035	0.042	0.036	0.042
	AI-IC	0.106	0.118	0.088	0.122	0.093	0.116	0.093	0.116
	SI-SBC	0.019	0.015	0.023	0.022	0.018	0.013	0.018	0.013
	SI-IC	0.223	0.246	0.175	0.258	0.174	0.243	0.174	0.243
Standard Errors	SS-AI	0.055	0.054	0.054	0.055	0.008	0.036	0.008	0.036
	SS-SI	0.056	0.056	0.055	0.056	0.010	0.032	0.008	0.031
	SS-SBC	0.053	0.053	0.053	0.053	0.008	0.011	0.008	0.011
	SS-IC	0.054	0.054	0.054	0.055	0.008	0.032	0.010	0.032
	AI-SBC	0.053	0.053	0.053	0.053	0.008	0.013	0.008	0.013
	AI-IC	0.053	0.053	0.053	0.053	0.010	0.036	0.010	0.036
	SI-SBC	0.052	0.052	0.053	0.052	0.008	0.011	0.008	0.011
	SI-IC	0.052	0.052	0.053	0.052	0.010	0.048	0.010	0.048
Confidence Intervals	SS-AI	(0.0102, 0.2258)	(0.0012, 0.2128)	(-0.0008,0.2108)	(0.0012,0.2168)	(0.0883, 0.1210)	(0.0359, 0.1764)	(0.0893,0.1207)	(0.0354,0.1766)
	SS-SI	(-0.0308, 0.1888)	(-0.0248, 0.1948)	(-0.0278,0.1878)	(-0.0258,0.1938)	(0.0640, 0.0970)	(0.0227, 0.1462)	(0.0653,0.0967)	(0.0232,0.1448)
	SS-SBC	(-0.0659, 0.1419)	(-0.0689, 0.1389)	(-0.0649,0.1429)	(-0.070,0.1379)	(0.0229, 0.0542)	(0.0120, 0.0569)	(0.0233,0.0547)	(0.0134,0.0566)
	SS-IC	(-0.0658, 0.1458)	(-0.0508, 0.1608)	(-0.0648,0.1468)	(-0.0538,0.1618)	(0.0270, 0.0653)	(-0.0094, 0.1168)	(0.0264,0.0656)	(-0.0087,0.1167)
	AI-SBC	(-0.0679, 0.1399)	(-0.0599, 0.1479)	(-0.0679,0.1399)	(-0.0609,0.1469)	(0.0197, 0.0513)	(0.0164, 0.0674)	(0.0203,0.0517)	(0.0165,0.0675)
	AI-IC	(0.0021, 0.2099)	(0.0141, 0.2219)	(-0.0159,0.191)	(0.0181,0.2259)	(0.0741, 0.1121)	(0.0452, 0.1869)	(0.0734,0.1126)	(0.0454,0.1866)
	SI-SBC	(-0.0769, 0.1269)	(-0.0869, 0.1169)	(-0.0809,0.1269)	(-0.0799,0.1239)	(0.0023, 0.0337)	(-0.0094, 0.0395)	(0.0020,0.1169)	(-0.0086,0.0334)
	SI-IC	(0.1211, 0.3249)	(0.1441, 0.3479)	(0.0711,0.2789)	(0.1561,0.3599)	(0.1554, 0.1936)	(0.1496, 0.3359)	(0.1544,0.1936)	(0.1489,0.3371)

Table 22: Differences of Path Estimates and Standard Errors Across Univariate and Multivariate Methods

		Fixed Effect					
	Variables	Uni r vs Uni z	Uni r vs TSSEM	Uni r vs OSMASEM	Uni z vs TSSEM	Uni z vs OSMASEM	TSSEM vs OSMASEM
Path Estimates	SS-AI	0.013	0.013	0.013	<0.0001	<0.0001	<0.0001
	SS-SI	-0.001	-0.002	-0.002	-0.001	-0.001	<0.0001
	SS-SBC	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	SS-IC	-0.001	-0.006	-0.006	-0.005	-0.005	<0.0001
	AI-SBC	<0.0001	0.001	<0.0001	0.001	<0.0001	0.001
	AI-IC	0.018	0.013	0.013	-0.005	-0.005	<0.0001
	SI-SBC	-0.004	0.001	0.001	0.005	0.005	<0.0001
	SI-IC	0.048	0.049	0.049	0.001	0.001	<0.0001
Standard Errors	SS-AI	0.001	0.047	0.047	0.046	0.046	<0.0001
	SS-SI	0.001	0.046	0.048	0.045	0.047	-0.002
	SS-SBC	<0.0001	0.045	0.045	0.045	0.045	<0.0001
	SS-IC	<0.0001	0.046	0.044	0.046	0.044	0.002
	AI-SBC	<0.0001	0.045	0.045	0.045	0.045	<0.0001
	AI-IC	<0.0001	0.043	0.043	0.043	0.043	<0.0001
	SI-SBC	-0.001	0.044	0.044	0.045	0.045	<0.0001
	SI-IC	-0.001	0.042	0.042	0.043	0.043	<0.0001

Table 23: Differences of Path Estimates and Standard Errors Across Univariate and Multivariate Methods (Random Effects)

		Random Effects					
	Variables	Uni r vs Uni z	Uni r vs TSSEM	Uni r vs OSMASEM	Uni z vs TSSEM	Uni z vs OSMASEM	TSSEM vs OSMASEM
Path Estimates	SS-AI	-0.002	0.001	0.001	0.003	0.003	<0.0001
	SS-SI	0.001	0.001	0.001	<0.0001	<0.0001	<0.0001
	SS-SBC	0.001	0.001	<0.0001	<0.0001	-0.001	-0.001
	SS-IC	0.001	0.001	0.001	<0.0001	<0.0001	<0.0001
	AI-SBC	0.001	0.002	0.002	0.001	0.001	<0.0001
	AI-IC	-0.004	0.002	0.002	0.006	0.006	<0.0001
	SI-SBC	-0.007	0.002	0.002	0.009	0.009	<0.0001
	SI-IC	-0.012	0.003	0.003	0.015	0.015	<0.0001
Standard Errors	SS-AI	-0.001	0.018	0.018	0.019	0.019	<0.0001
	SS-SI	<0.0001	0.024	0.025	0.024	0.025	0.001
	SS-SBC	<0.0001	0.042	0.042	0.042	0.042	<0.0001
	SS-IC	-0.001	0.022	0.022	0.023	0.023	<0.0001
	AI-SBC	<0.0001	0.04	0.04	0.04	0.04	<0.0001
	AI-IC	<0.0001	0.017	0.017	0.017	0.017	<0.0001
	SI-SBC	<0.0001	0.041	0.041	0.041	0.041	<0.0001
	SI-IC	<0.0001	0.004	0.004	0.004	0.004	<0.0001

Table 24: Differences in Estimated Covariance and Variances Between Univariate and Multivariate Approaches

Fixed Effect							
	Variables	Uni r vs Uni z	Uni r vs TSSEM	Uni r vs OSMASEM	Uni z vs TSSEM	Uni z vs OSMASEM	TSSEM vs OSMASEM
Cov	AI-SI	0.015	0.015	0.015	<0.0001	<0.0001	<0.0001
Variances	SS	-0.004	-0.004	-0.006	<0.0001	-0.002	0.001
	AI	-0.003	-0.005	-0.005	-0.002	-0.002	0.005
	SI	-0.019	-0.023	-0.022	-0.004	-0.003	0.029
Random Effects							
	Variables	Uni r vs Uni z	Uni r vs TSSEM	Uni r vs OSMASEM	Uni z vs TSSEM	Uni z vs OSMASEM	TSSEM vs OSMASEM
Cov	AI-SI	-0.005	0.001	0.001	0.006	0.006	<0.0001
Variances	SS	0.001	-0.004	-0.003	-0.005	-0.004	0.001
	AI	<0.0001	0.001	-0.004	0.001	-0.004	-0.005
	SI	0.005	-0.004	-0.005	-0.009	-0.01	-0.001

Table 25: Publication Bias Measures

Variables	Egger's Regression Test	Trim-and-Fill
SS-AI	$z = -0.7761,$ $p = 0.4377$	0 missing studies
SS-SI	$z = -0.4005,$ $p = 0.6888$	3 missing studies, $0.1553(0.0300)$ $p < 0.0001$
SS-SBC	$z = -0.5118, p$ $= 0.6088$	4 missing studies $0.0508\ (0.0117)$ $p < 0.0001$
SS-IC	$z = 0.7653, p$ $= 0.4441$	3 missing studies $0.1248\ (0.0331)$ $p = 0.0002$
AI-SI	$z = -2.2662, p$ $= 0.0234$	0 missing studies
AI-SBC	$z = 0.7183,$ $p = 0.4726$	2 missing studies $0.0432(0.0143)$ $p < 0.05$
AI-IC	$z = -0.1731,$ $p = 0.8636$	6 missing studies $0.1936\ (0.0402)$ $p < 0.0001$
SI-SBC	$z = -0.2067, p$ $= 0.8362$	2 missing studies $0.0215(0.0126)$ $p = 0.0881$
SI-IC	$z = 0.3862, p$ $= 0.6994$	0 missing studies
SBC-IC	$z = 0.0158, p$ $= 0.9874$	2 missing studies $0.0006(0.0102)$ $p = 0.9506$

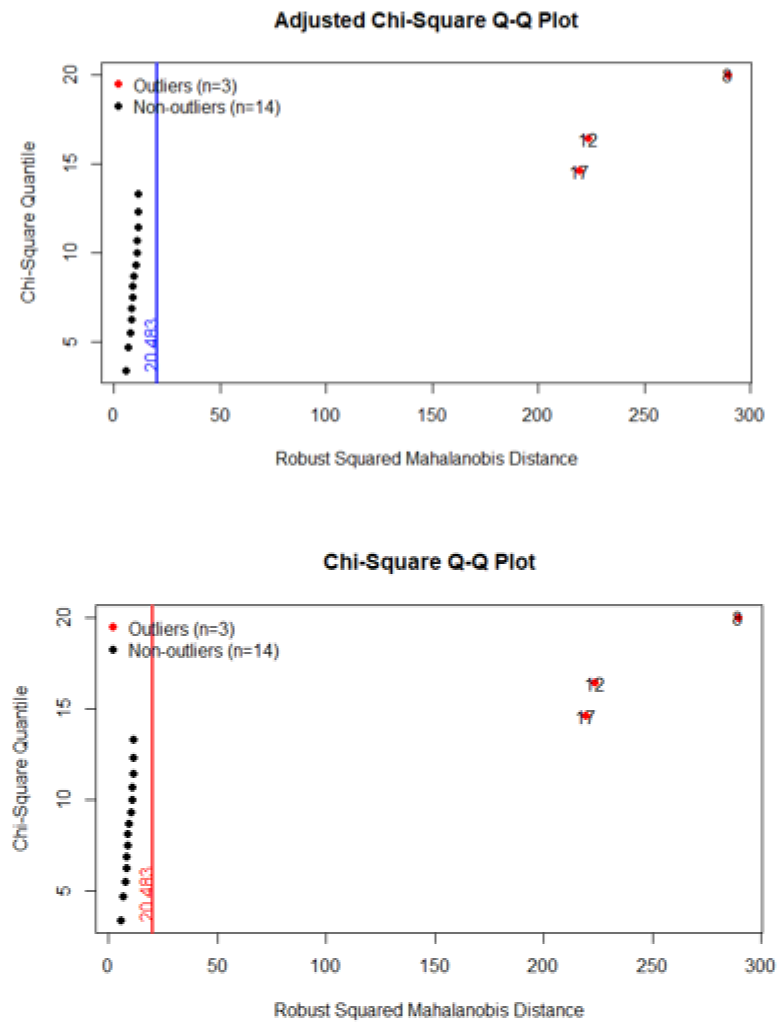

Figure 8: Chi-square and Adjusted Chi-Square Multivariate QQ-Plot

Table 26: Measures of Influence

Variables	$rstud > \|3\|$?	$diffits > 1$?	$cook.d > 0.45$?
SS-AI	No	No	No
SS-SI	No	No	No
SS-SBC	No	No	No
SS-IC	No	No	No
AI-SI	No	No	No
AI-SBC	No	No	No
AI-IC	Yes; Study 6 (Rstud = 4.2344)	Yes; study 6 Dffits = 1.1062	Yes; study 6 Cook.d = 0.5685
SI-SBC	No	No	No
SI-IC	No	Yes; Study 6 Dffits = 0.6156	No
SBC-IC	No	No	No

APPENDIX G

UNIVARIATE FOREST PLOTS

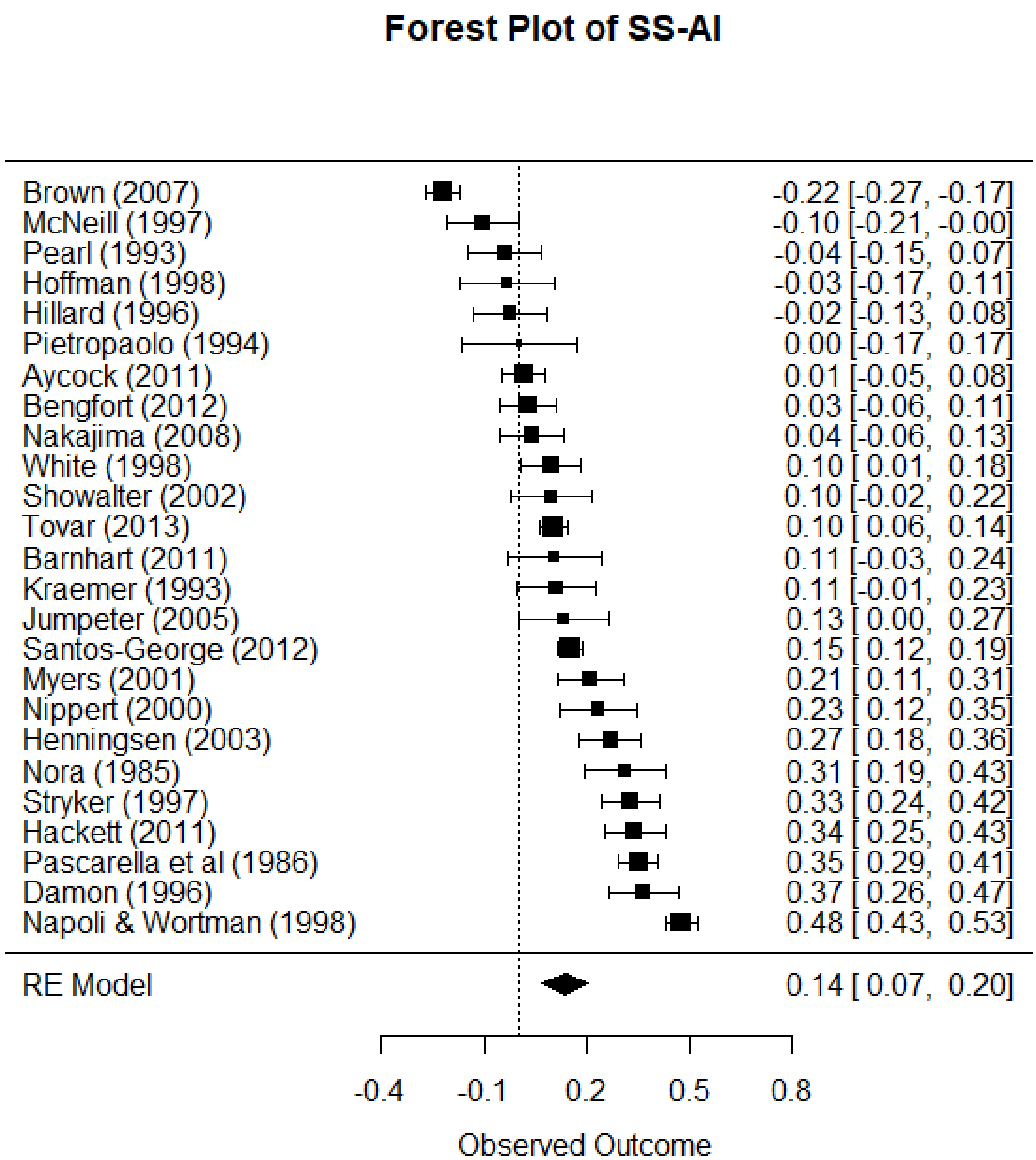

Figure 9: Forest Plot of SS/AI

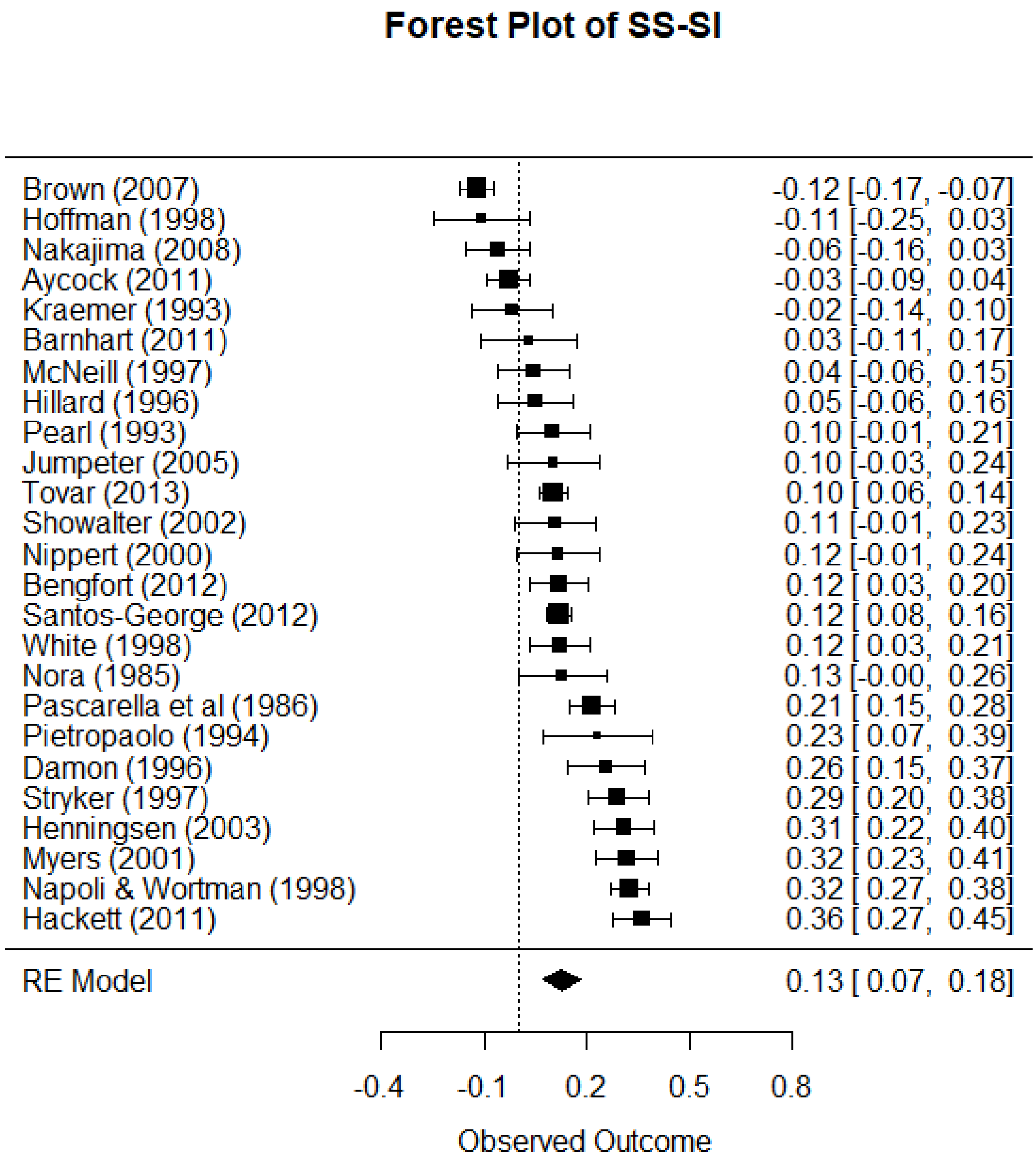

Figure 10:Forest Plot of SS/SI

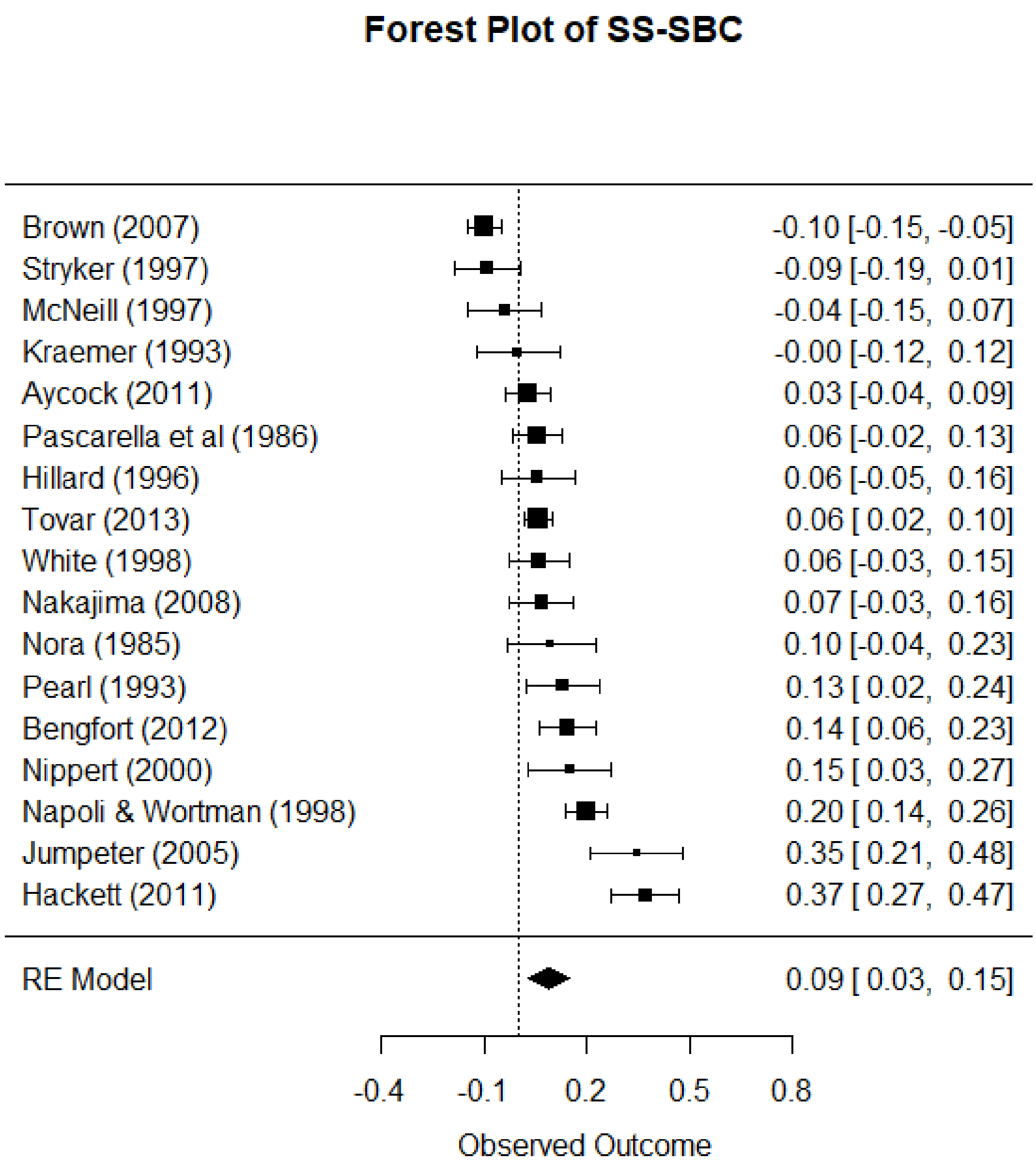

Figure 11:Forest Plot of SS/SBC

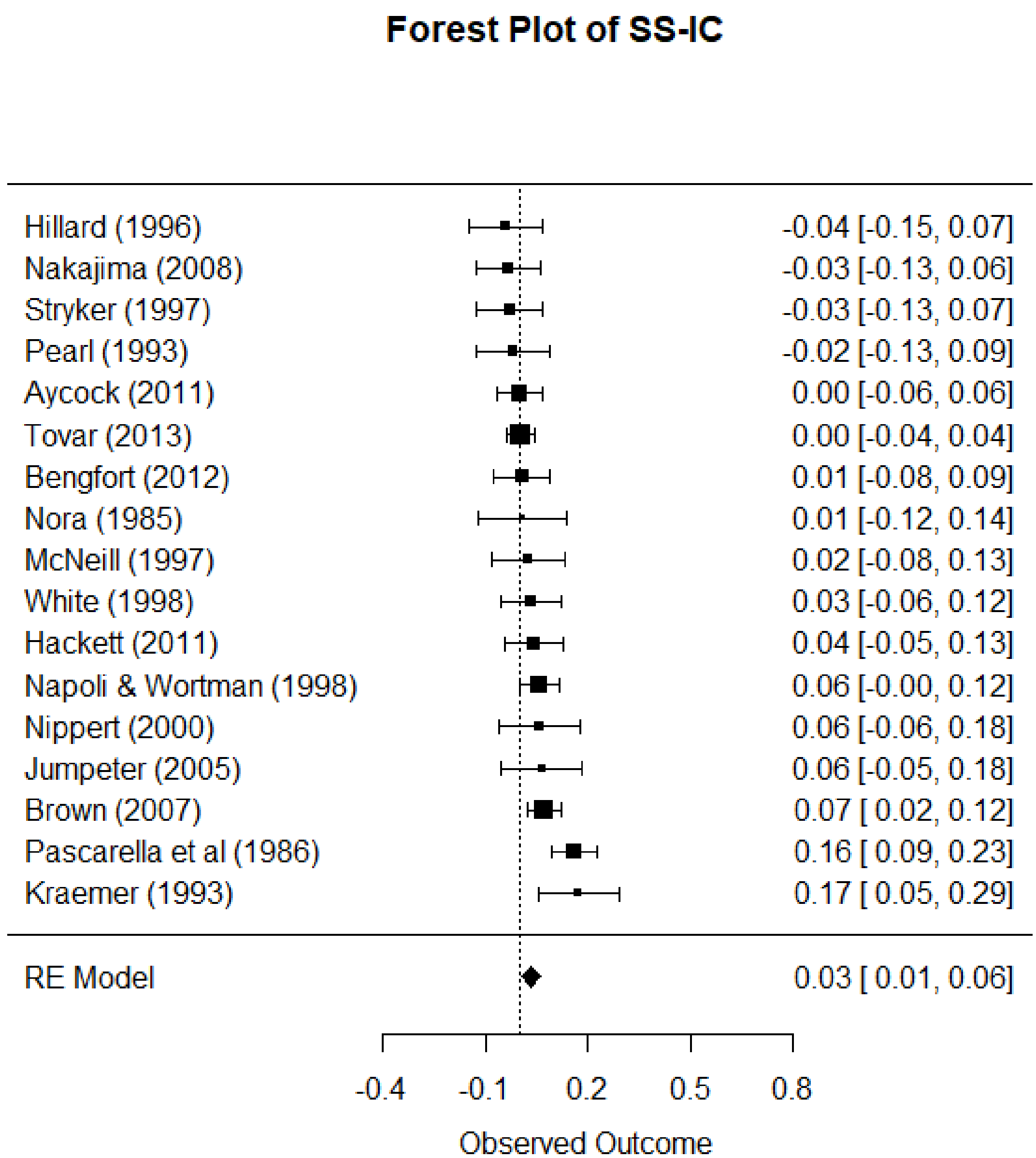

Figure 12:Forest Plot of SS/IC

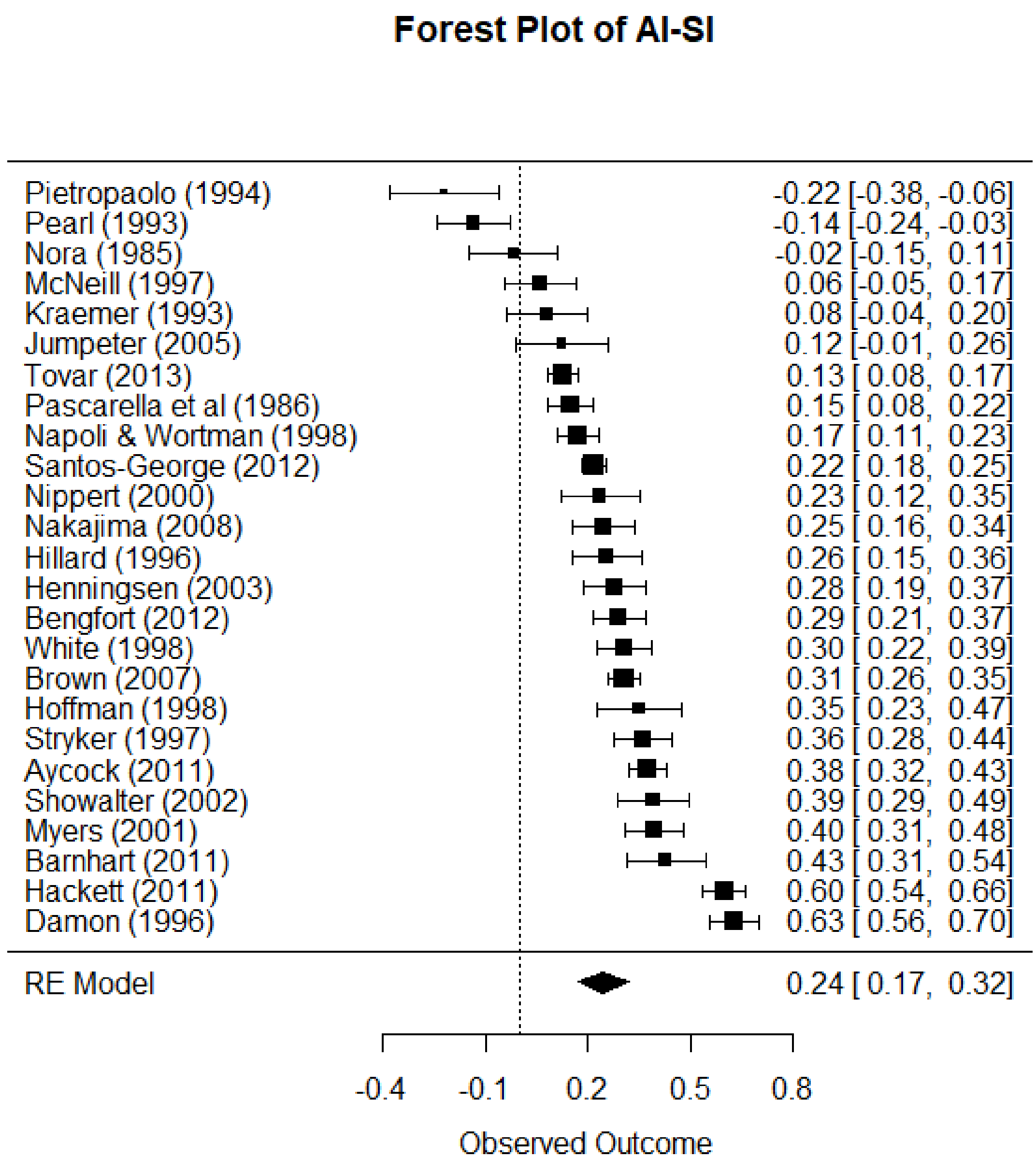

Figure 13:Forest Plot of AI/SI

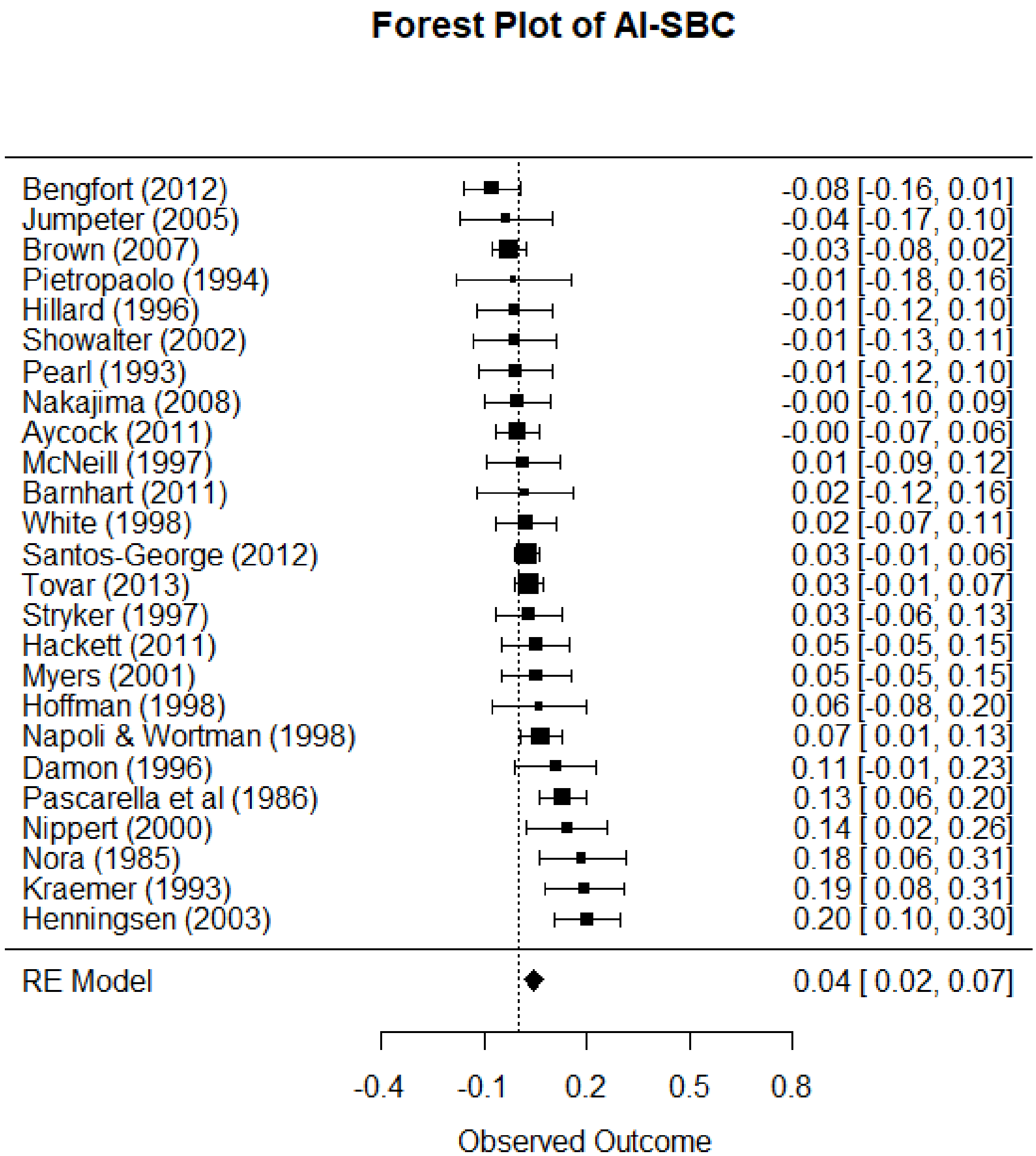

Figure 14:Forest Plot of AI/SBC

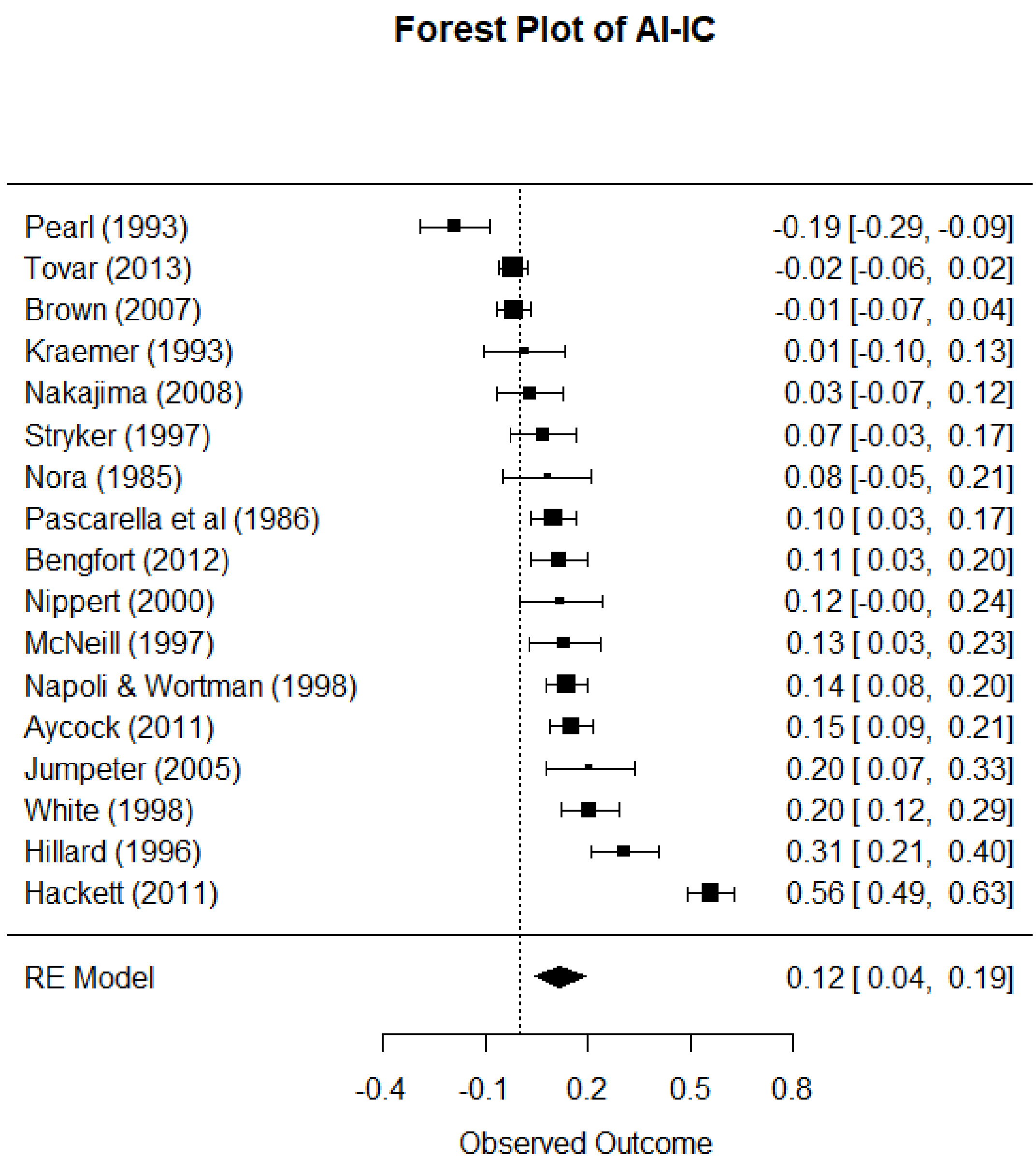

Figure 15:Forest Plot of AI/IC

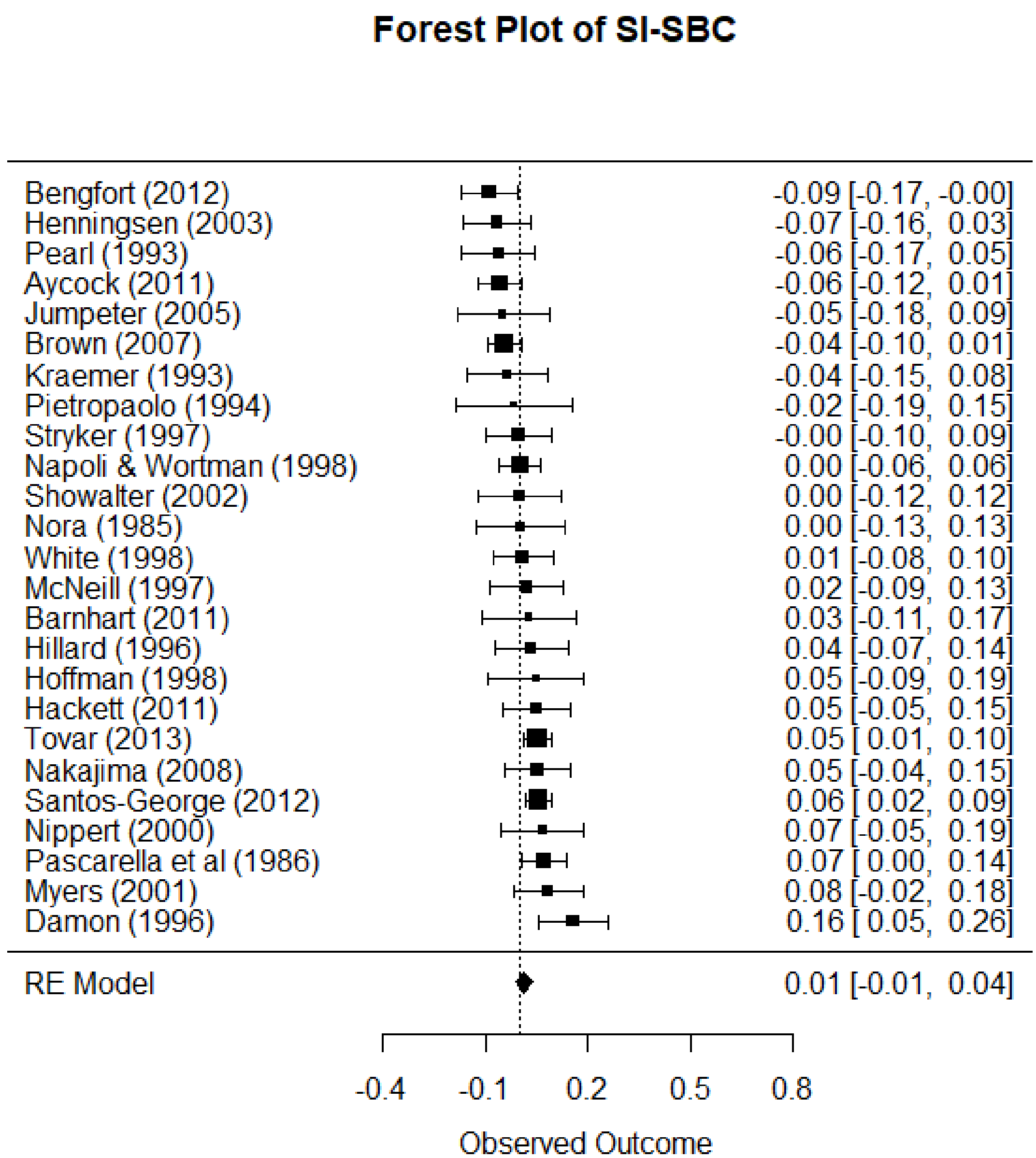

Figure 16:Forest Plot of SI/SBC

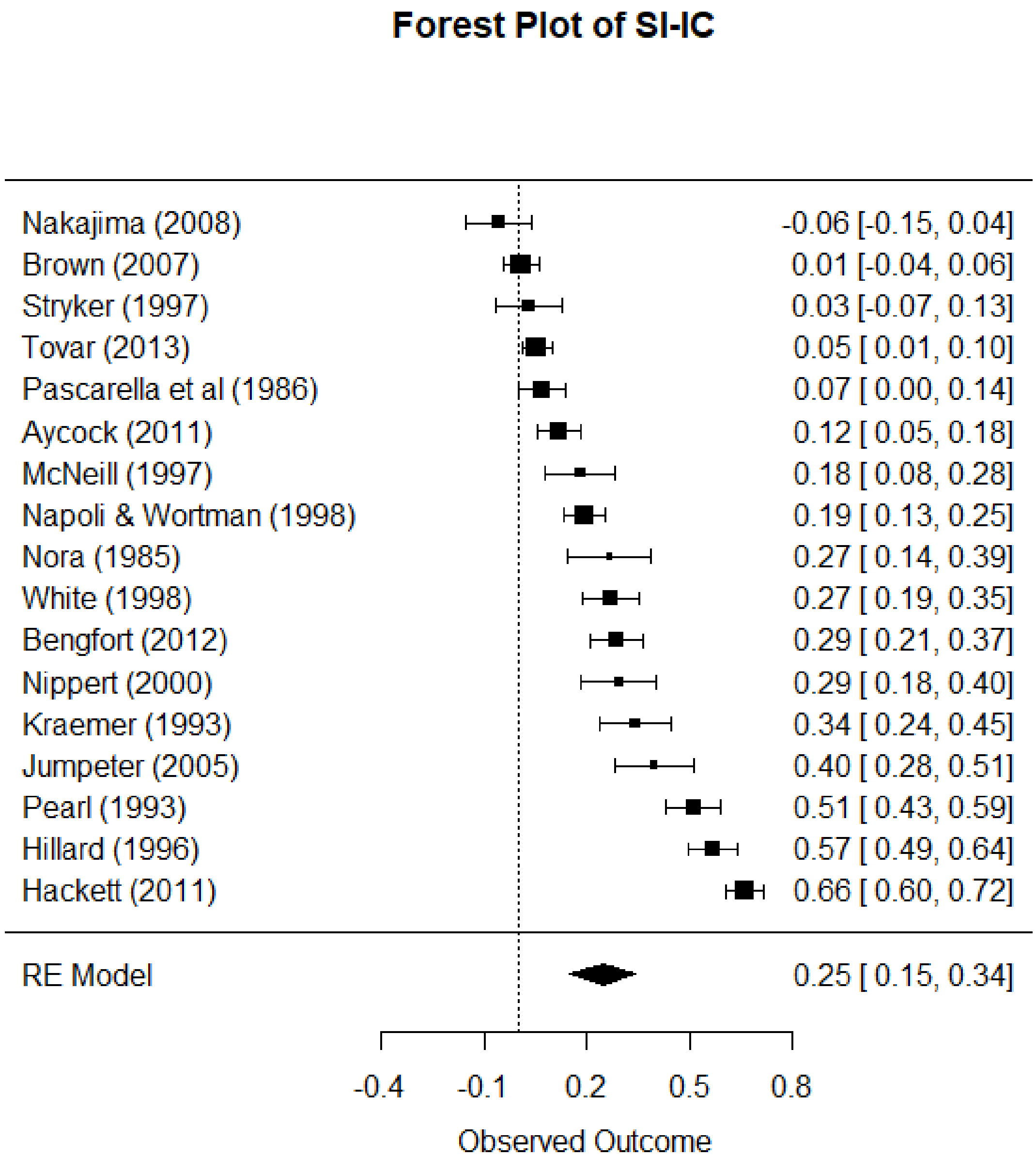

Figure 17:Forest Plot of SI/IC

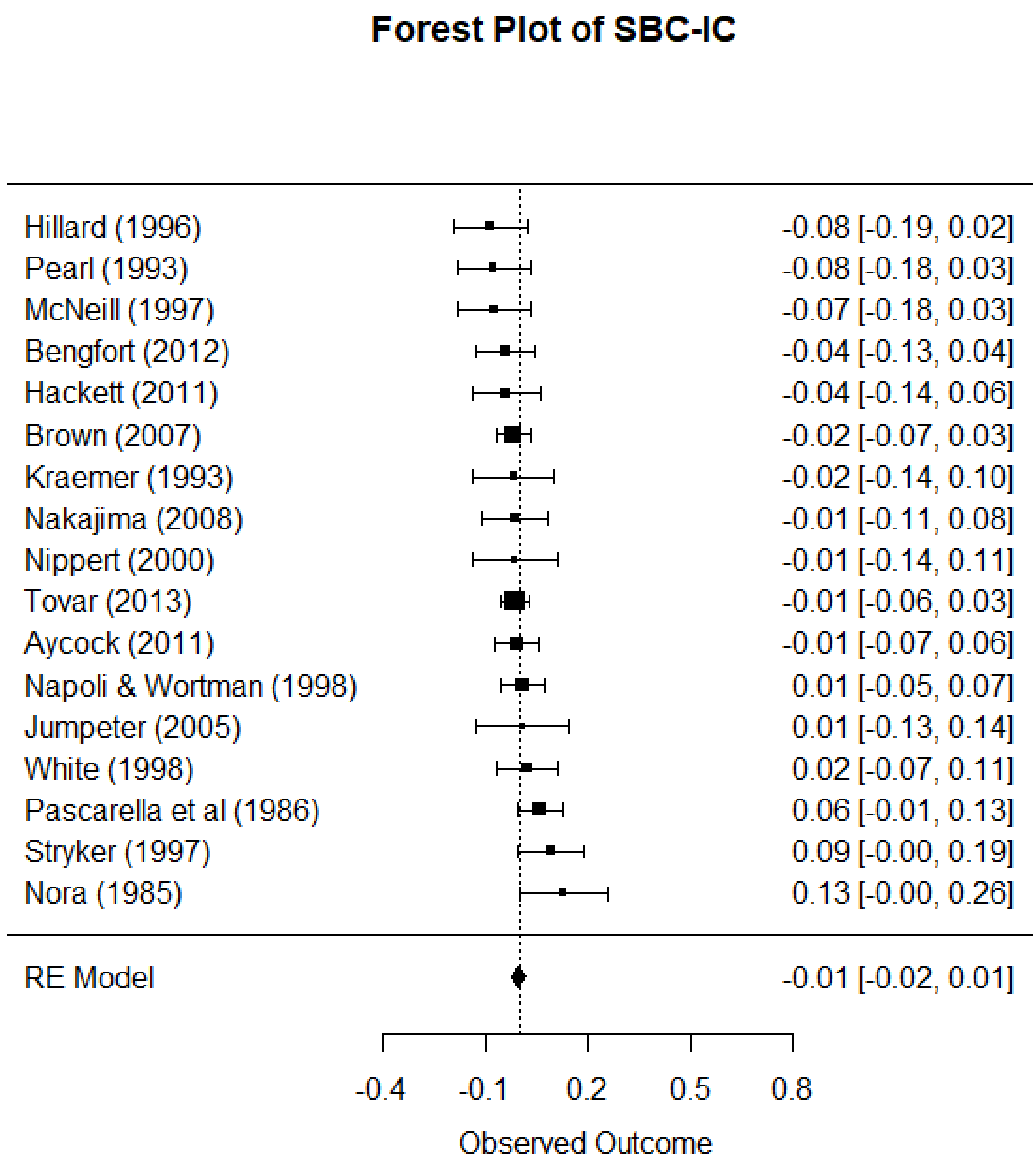

Figure 18:Forest Plot of SBC/IC

APPENDIX H

UNIVARIATE FUNNEL PLOTS

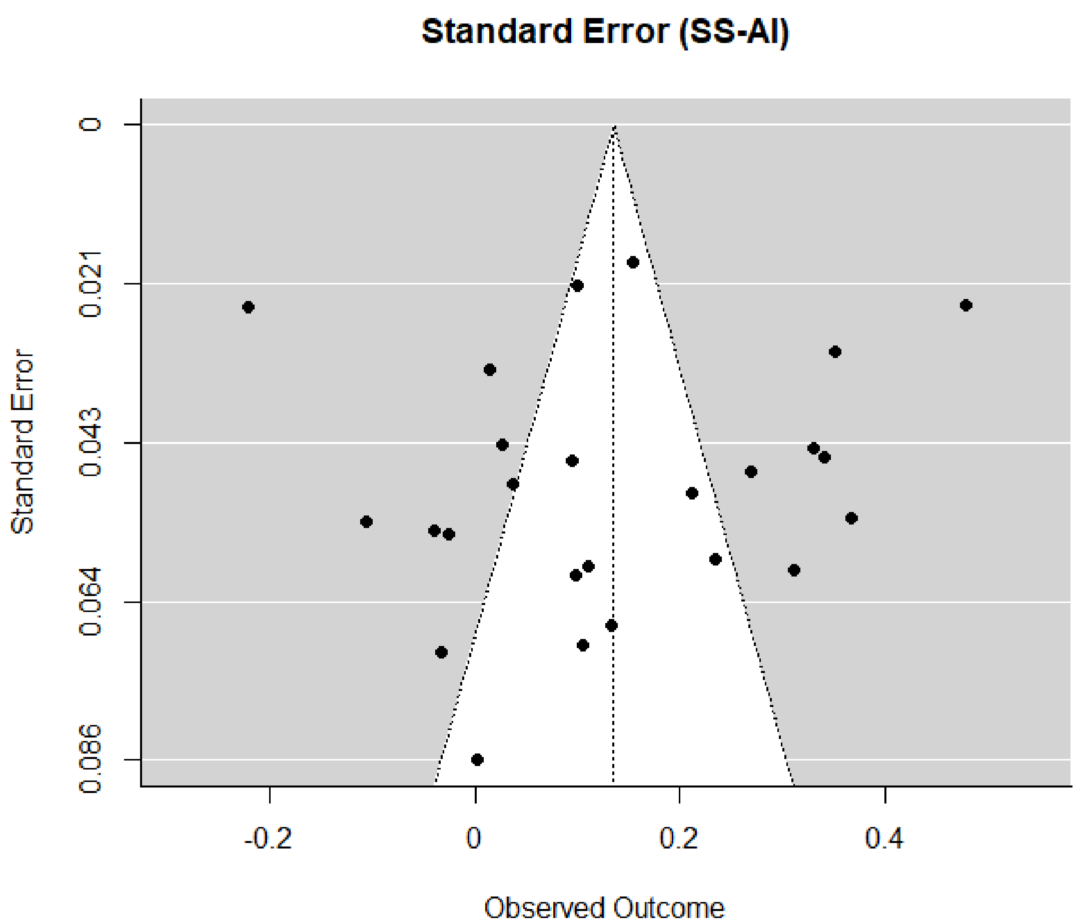

Figure 19: Trim-and-Filled Funnel Plot for SS/AI

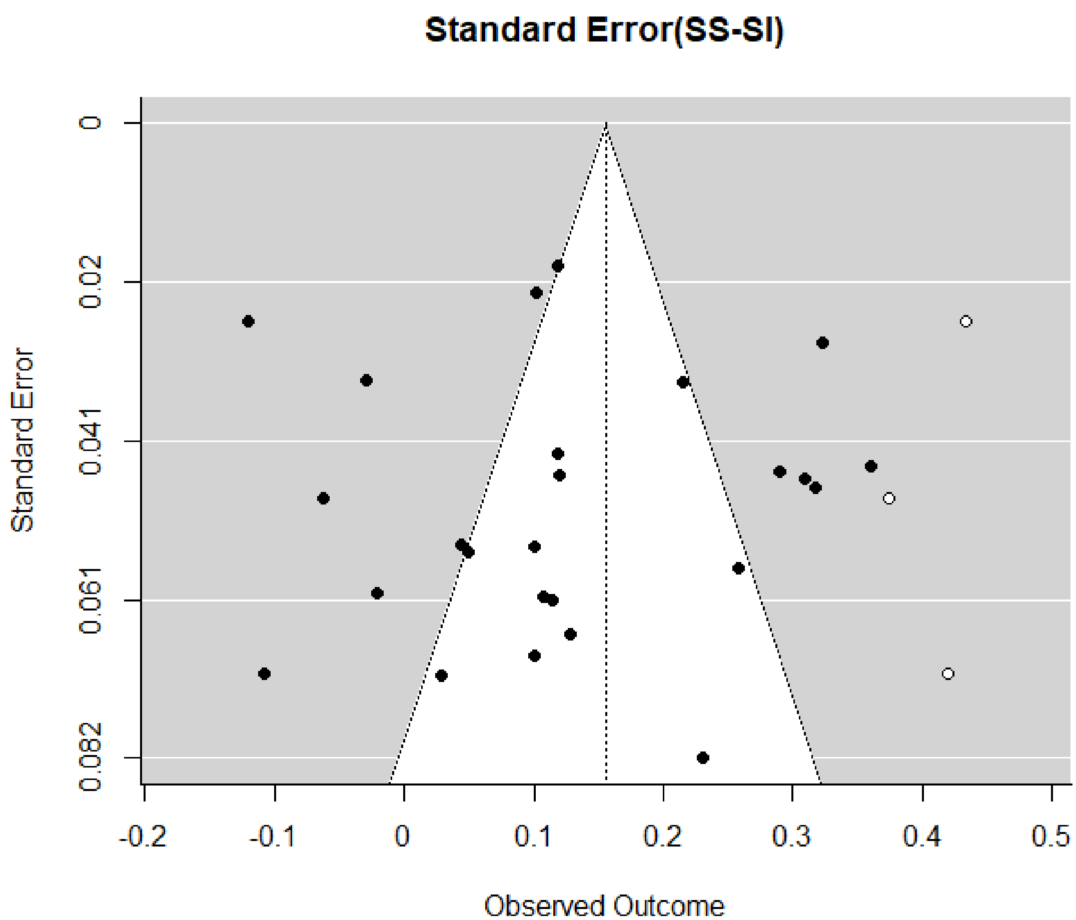

Figure 20: Trim-and-Filled Funnel Plot for SS/SI

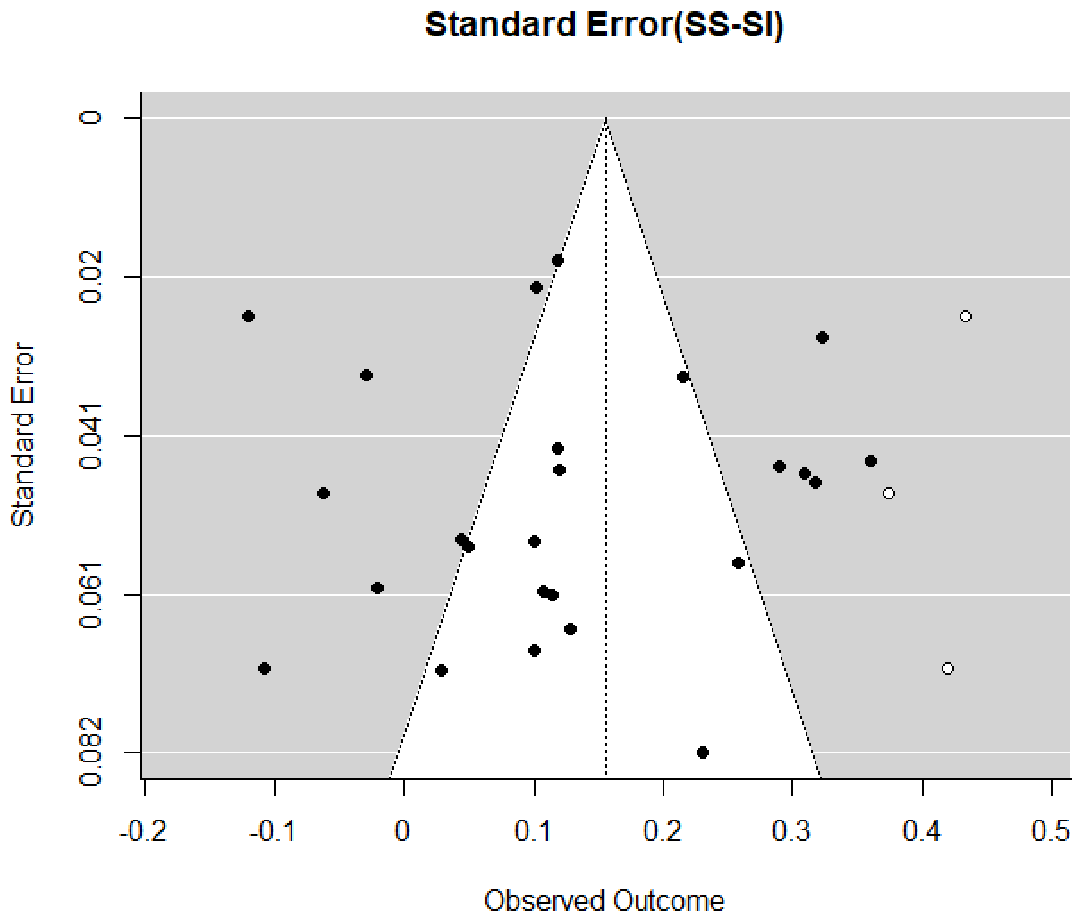

Figure 21: Trim-and-Filled Funnel Plot for SS/SI

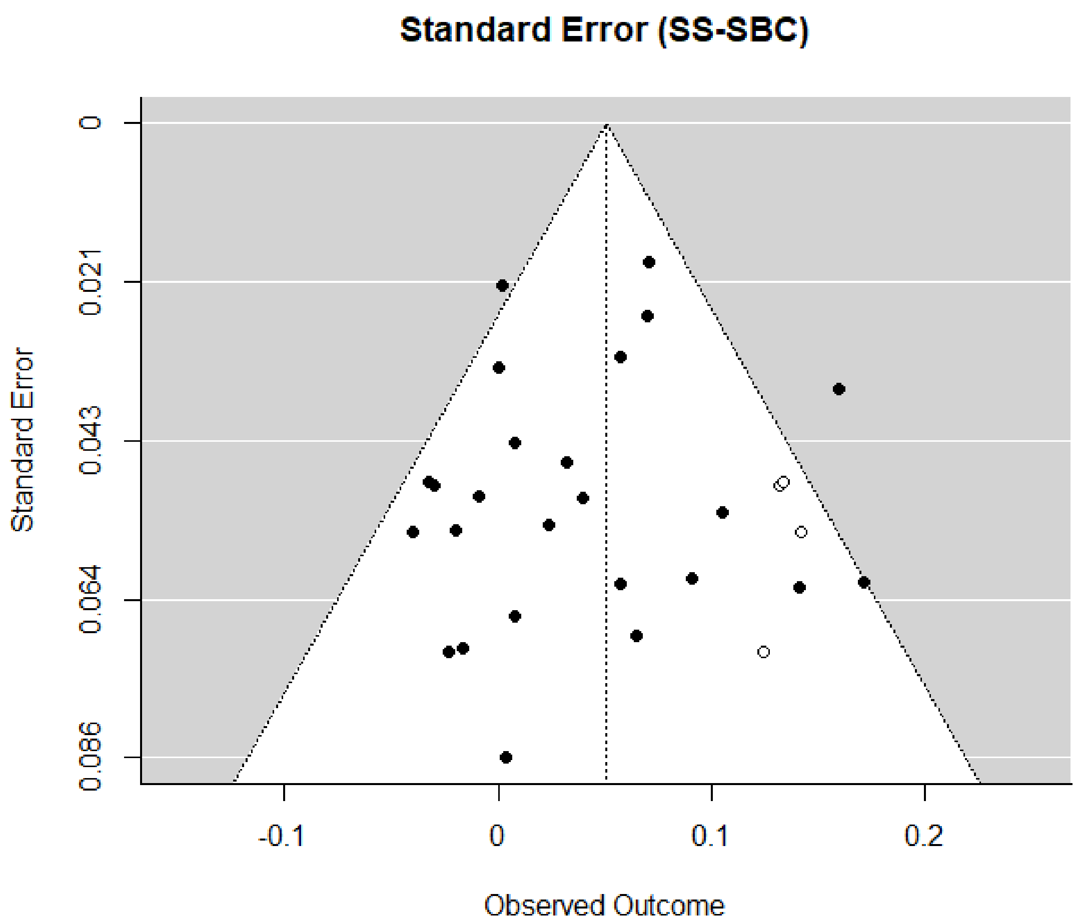

Figure 22: Trim-and-Filled Funnel Plot for SS/SBC

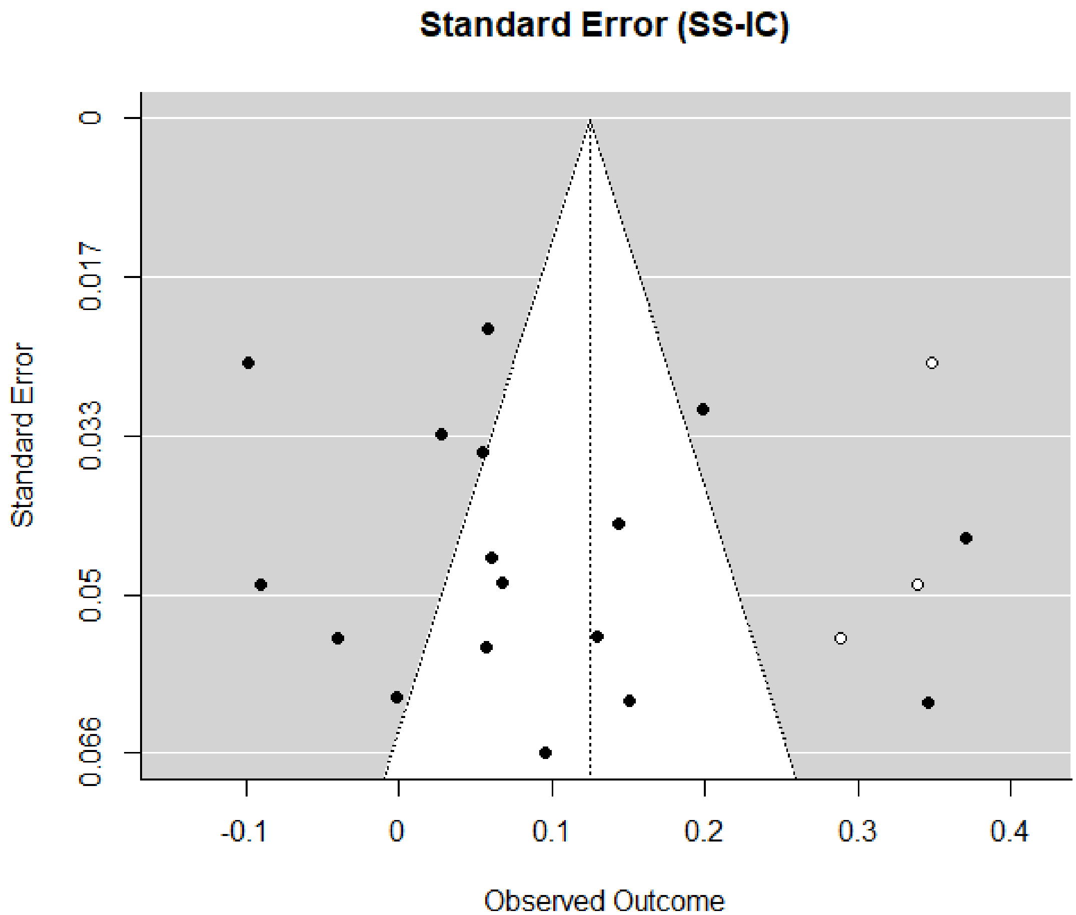

Figure 23: Trim-and-Filled Funnel Plot for SS/IC

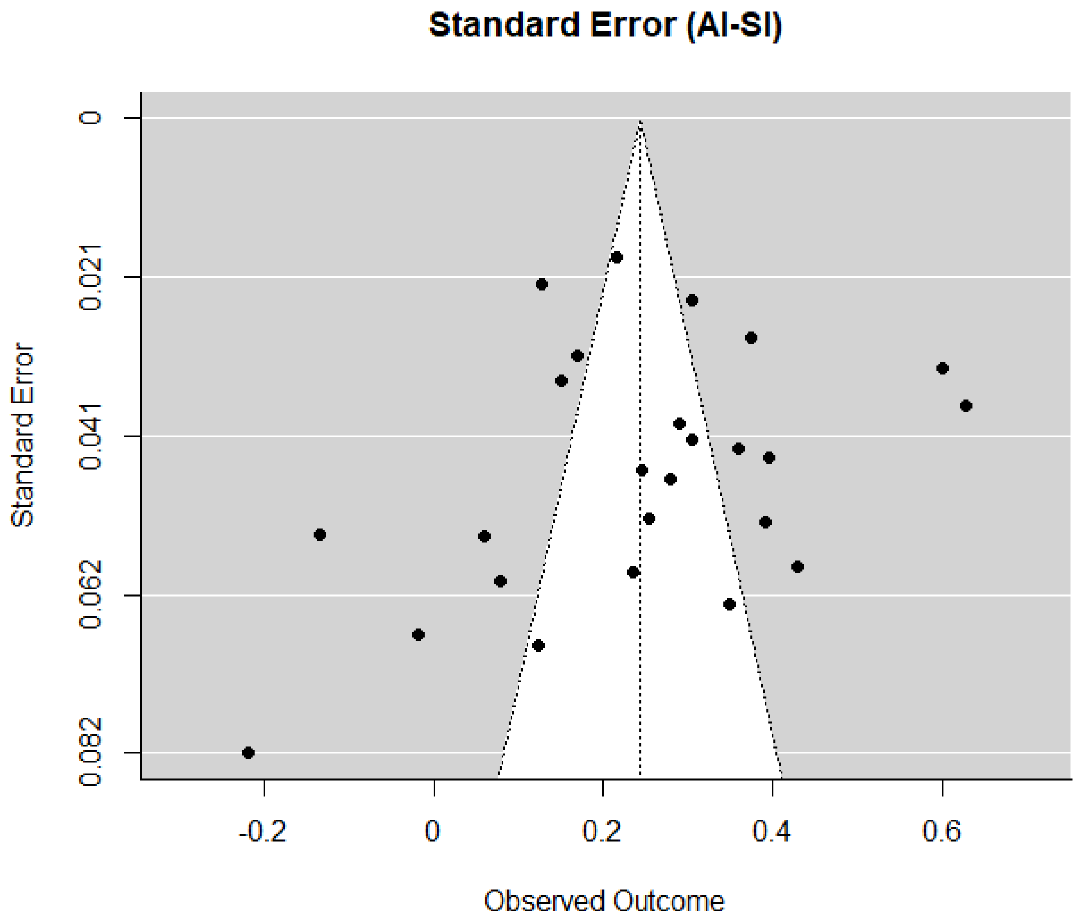

Figure 24: Trim-and-Filled Funnel Plot for AI/SI

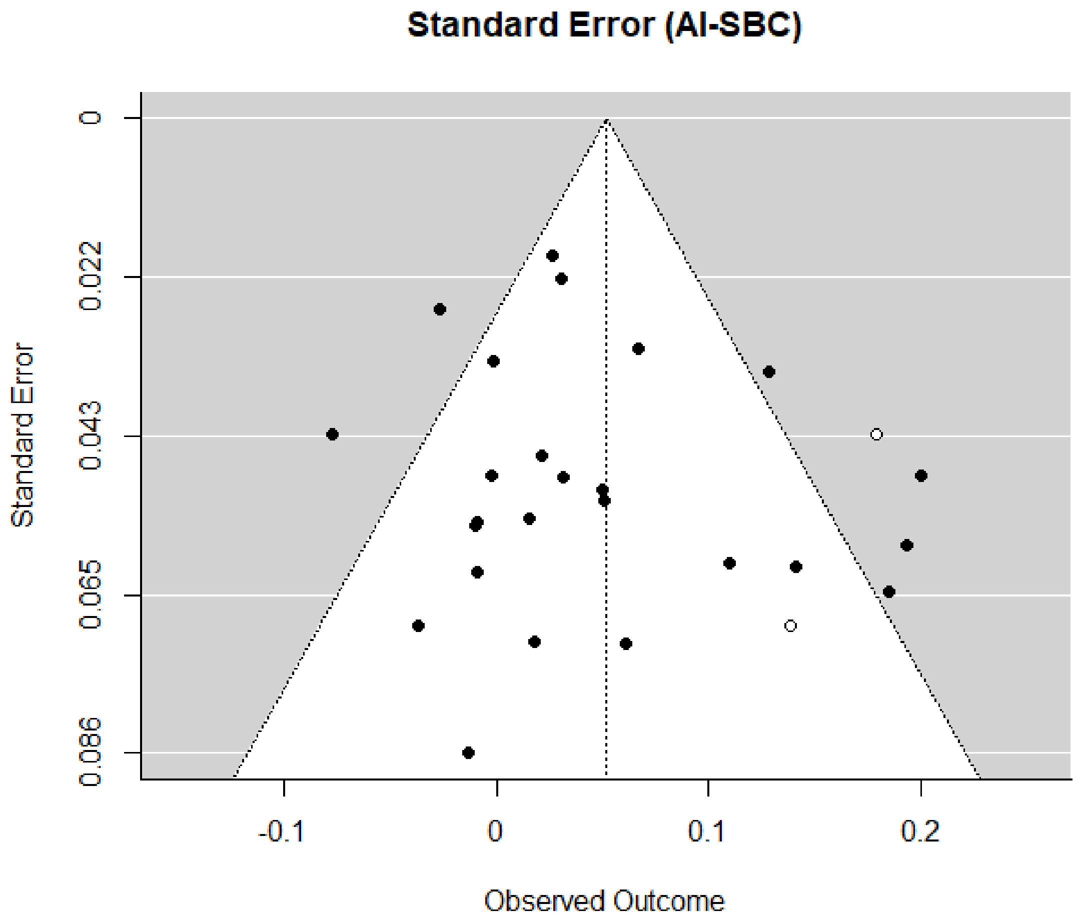

Figure 25: Trim-and-Filled Funnel Plot for AI/SBC

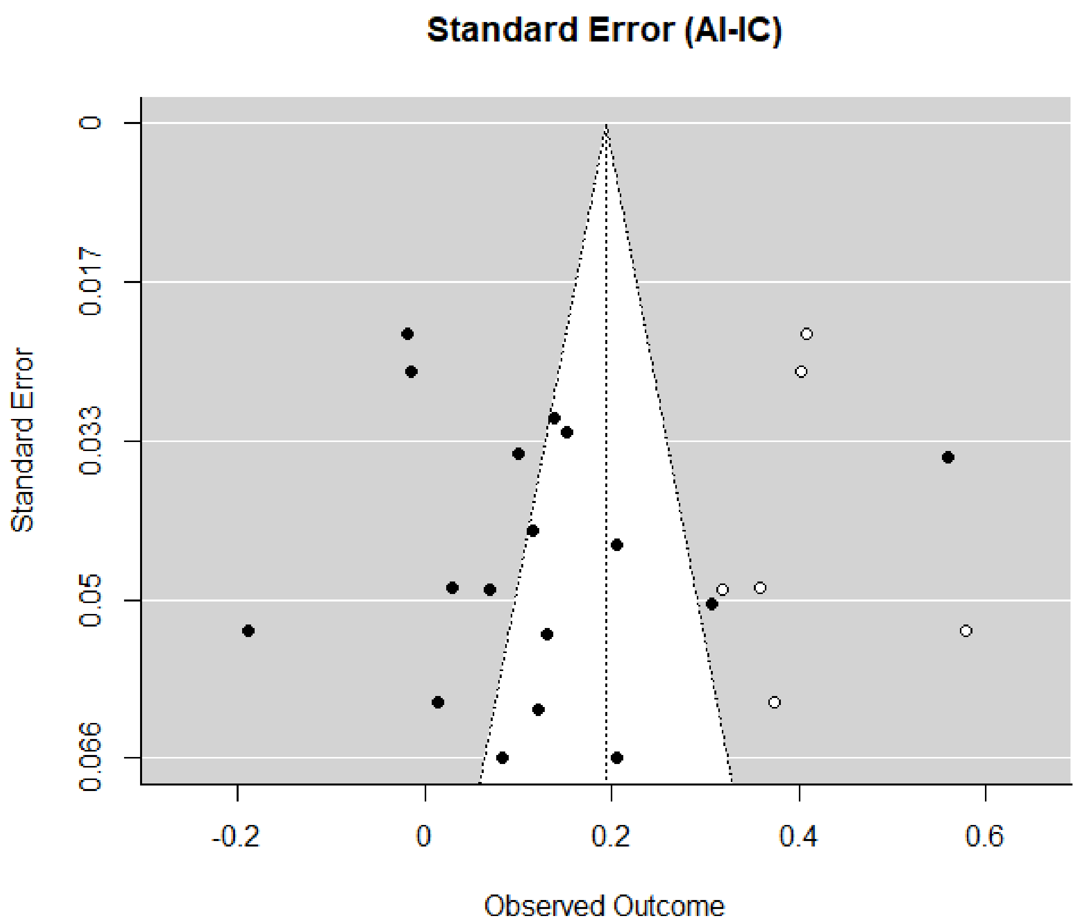

Figure 26: Trim-and-Filled Funnel Plot for AI/IC

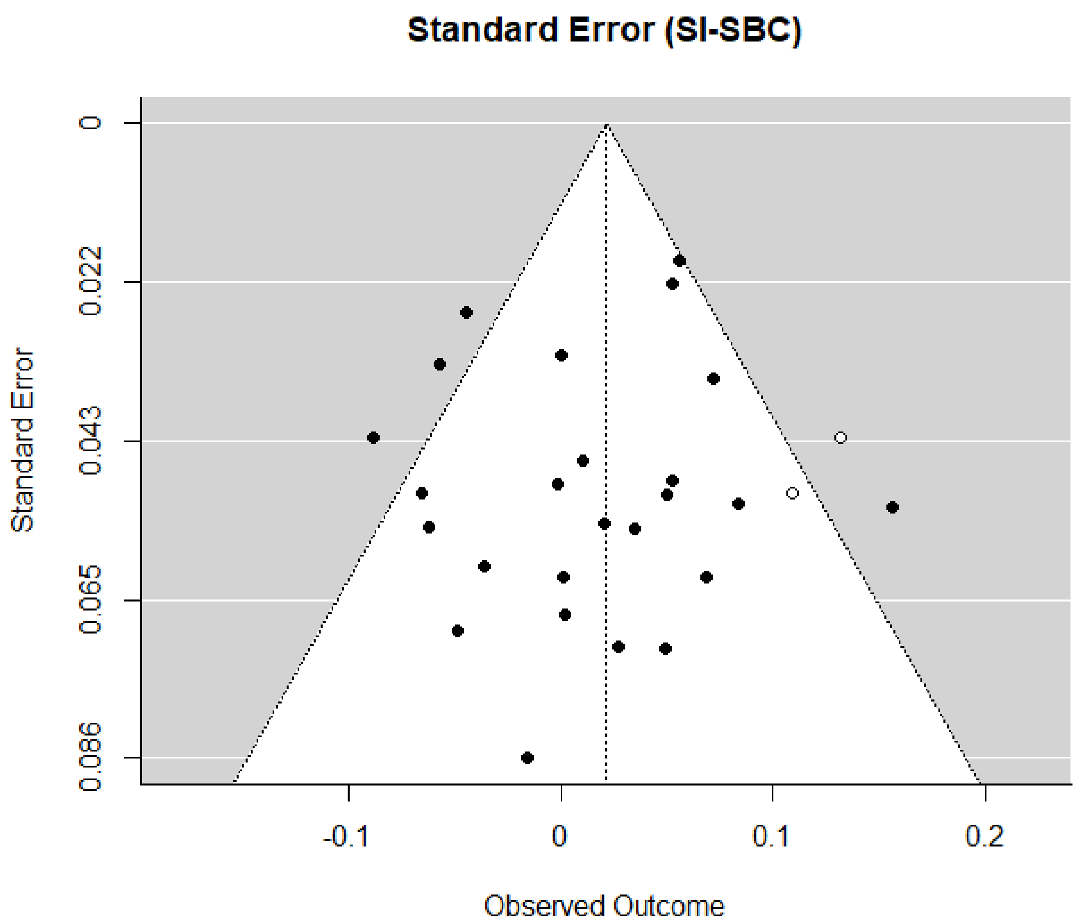

Figure 27: Trim-and-Filled Funnel Plot for SI/SBC

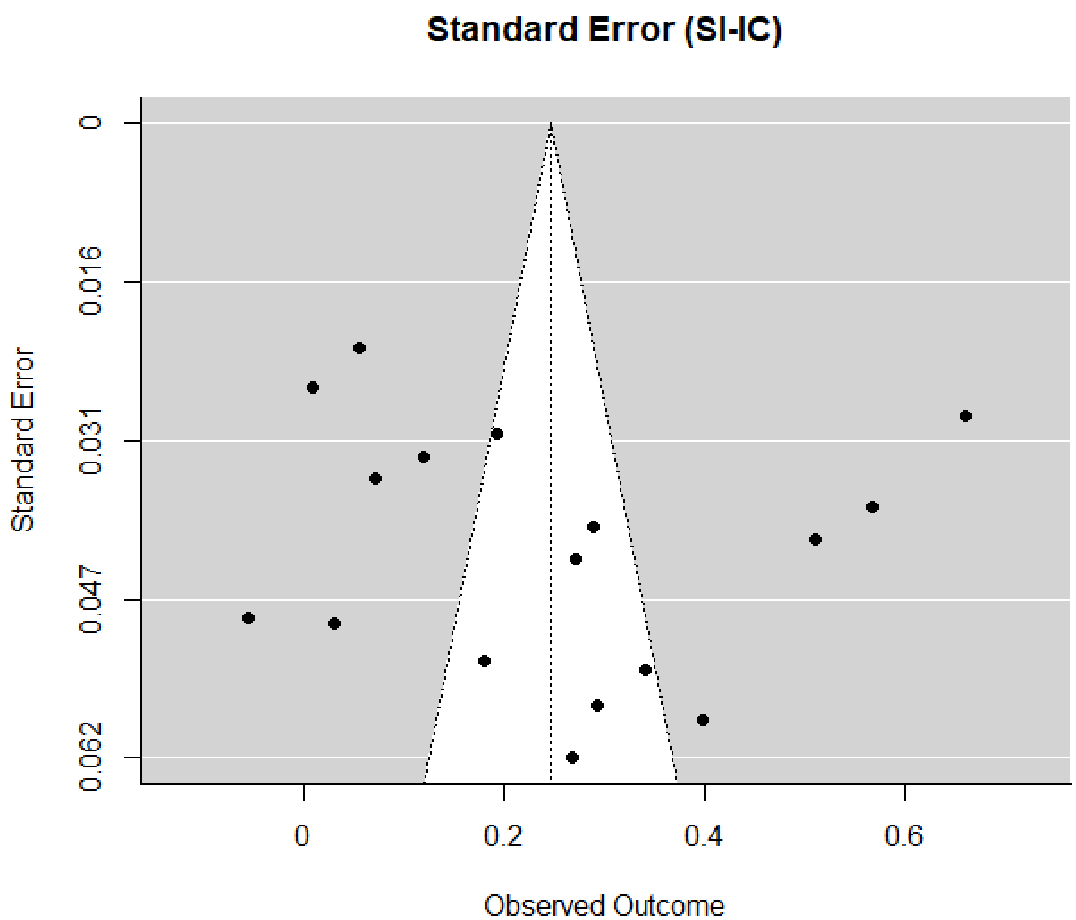

Figure 28: Trim-and-Filled Funnel Plot for SI/IC

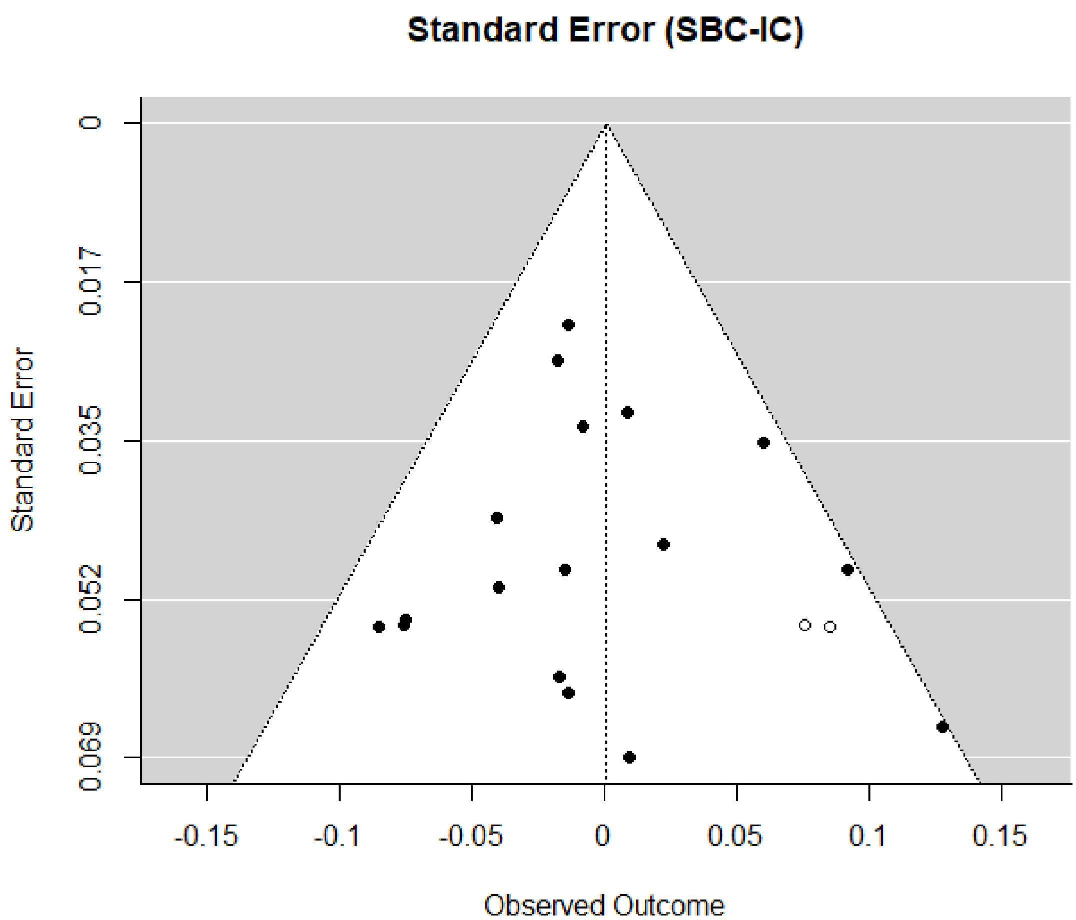

Figure 29: Trim-and-Filled Funnel Plot for SBC/IC