A Comparison Study of the Reliability Coefficients from Five Approaches to Reliability Estimation

by

Wei Tang

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Measurement, Evaluation, and Cognition

Department of Educational Psychology
University of Alberta

Abstract

In order to estimate reliability by a single administration of one test form, various approaches and corresponding reliability coefficients have been proposed so far. Currently, the five most influential approaches are: internal consistency, lower bound, principal components analysis (PCA), exploratory factor analysis (EFA), and structural equation modeling (SEM). Facing various approaches and thus dozens of reliability coefficients derived for estimating reliability, practicing researchers are curious to know which reliability coefficient(s) performs best, and under what circumstances. However, a comprehensive comparison of the reliability coefficients from the aforementioned five approaches has not been conducted yet. Therefore, a Monte Carlo study was conducted to evaluate the performances of the reliability coefficients from the five approaches under the conditions that are known to have effect on reliability estimation. Monte Carlo design factors included twelve specific measurement models, two levels of item number, three levels of sample size, three levels of error correlation, and two levels of factor correlation. In total, 72 simulation conditions were created by the combination of all design factors, and each condition was replicated 1,000 times in R environment. The results were collected in two stages. In the first stage, the percentage relative bias, standard error and root mean square error of each reliability coefficient were calculated for each condition. The rounded percentages of estimation failure numbers for each SEM reliability coefficient under all the manipulated conditions were also obtained to identify the conditions with serious estimation issues for the second stage analysis. In the second stage of this study, the percentage relative bias, standard error and root mean square error of Bayesian SEM estimates of reliability for the selected conditions were calculated. Results showed that correctly specified SEM estimates of reliability were least biased and comparatively stable under most of the conditions across the twelve measurement models in this study. However, under the conditions of small item numbers and complicated models, correctly specified SEM estimates of reliability were least accurate and exceptionally unstable due to estimation problems. In addition, over-specified SEM estimates of reliability were examined under the conditions in Model 1 (the tau-equivalent model with independent errors), Model 4 (the congeneric model with independent errors), Model 7 (the correlated factor model with factor correlation at 0.2 and independent errors) and Model 10 (the correlated factor model with factor correlation at 0.6 and independent errors). Results indicated that over-specified SEM estimates of reliability were as accurate and stable as correctly specified SEM estimates of reliability unless estimation problems occurred. Results in the second stage showed that the Bayesian estimation method with non-informative priors could effectively solve estimation problems but fail to eradicate the biases in SEM estimates of reliability. In order to solve estimation problems as well as maintaining the accuracy of SEM estimates of reliability, more types of priors need be tested and compared when using Bayesian estimation methods in a future study.

Acknowledgements

Completing a Ph.D. program is one of my childhood dreams. Making this dream come true is significant for me. It strengthens my belief that if there is a will, there is a way. However, this dream would never come true without the support and help from many people. First of all, I would like to show my deepest gratitude to my academic advisor and dissertation supervisor, Dr. Ying Cui. Without her constant support, I can certainly not finish my Ph.D. program with so many joyful experiences. Dr. Ying Cui has gradually changed my thinking patterns towards research through her constructive suggestions on my research work. She has also provided me with the challenges and opportunities to cultivate my research independence and to develop my abilities as a quantitative researcher.

I would also like to thank the members of my doctoral committee for their insightful comments on my research: Dr. Mark J. Gierl, Dr. Mary Roduta Roberts, Dr. Damien Cormier, Dr. Rebecca Gokiert, and Dr. Bruno D. Zumbo. They all have taught me how to make the study to be more applicable in practice. I am heartily thankful to Dr. Mary Roduta Roberts for providing me with detailed feedback for each chapter in my dissertation.

I have been fortunate to study in the Centre for Research in Applied Measurement and Evaluation (CRAME). Dr. Jacqueline P. Leighton has been an important mentor during my program in CRAME. She has strictly supervised every step involved in conducting empirical research work, and trained me to work professionally. I felt fortunate that I have the opportunity of working with her as her research assistant. I have also made good friends with my peers in CRAME. I would like to express my sincere thankfulness to Qi Guo and Amin Mousavi. They have been very helpful and supportive. I really enjoyed learning all the mathematical and statistical courses with Qi, and working closely with him through my Ph.D. program.

Further, I wish to thank my mother for her encouragement, support and unconditional love over these years. Finally to my dear husband, Guangting, your love, support, and bravery has been truly crucial for me. Without you, I could not have changed so much and achieved so much. Thank you for everything.

TABLE OF CONTENTS

LIST OF TABLES viii

LIST OF FIGURES xi

LIST OF ABBREVIATIONS xii

LIST OF TERMS xiii

CHAPTER 1 INTRODUCTION 1

CHAPTER 2 LITERATURE REVIEW 7

2.1 Reliability and Classical Test Theory 7

2.2 Five Approaches to Reliability Estimation 9

2.2.1. The Internal Consistency Approach 10

2.2.2. The Lower Bound Approach 16

2.2.3. The PCA Approach 21

2.2.4. The EFA Approach 23

2.2.5. The SEM Approach 25

2.3. Methods Used in Previous Reliability Coefficients Comparison Studies 37

2.4. Study Purpose and Research Questions 42

CHAPTER 3 METHODS 46

3.1. Study Design 46

3.2. Data Generation and Analysis 51

3.3. Evaluation Criteria 53

CHAPTER 4 RESULTS 55

4.1. Results in Stage One 55

4.1.1 Reliability Estimates for Tau-Equivalent Models 55

4.1.2 Reliability Estimates for Congeneric Models 67

4.1.3 Reliability Estimates for Correlated Models 78

4.1.4 Percentage of Estimation Failure Number 100

4.2. Results in Stage Two 105

CHAPTER 5 DISCUSSION AND CONCLUSIONS 110

5.1 Discussion on the Studied Reliability Coefficients 110

5.1.1 Coefficient alpha 110

5.1.2 Glb 112

5.1.3 Theta and Theta.g 114

5.1.4 Omega 116

5.1.5 SEM Estimates of Reliability Using ML and Bayesian Estimation 118

5.2 Summary of Answers to Research Questions 121

5.3 Limitations and Suggestions for Future Research 128

REFERENCES 130

LIST OF TABLES

Table

1. Model Design 49
2. Percentage Bias, SE and RMSE of Reliability Estimates of the Data Generated from the Tau-equivalent Model with Independent Errors (Model 1) 58
3. Percentage Bias, SE and RMSE of Reliability Estimates of the Data Generated from the Tau-equivalent Model with Error Correlation at 0.2 (Model 2) 61
4. Percentage Bias, SE and RMSE of Reliability Estimates of the Data Generated from the Tau-equivalent Model with Error Correlation at 0.5 (Model 3) 64
5. Percentage Bias, SE and RMSE of Reliability Estimates of the Data Generated from the Congeneric Model with Independent Errors (Model 4).................... 69
6. Percentage Bias, SE and RMSE of Reliability Estimates of the Data Generated from the Congeneric Model with Error Correlation at 0.2 (Model 5)............. 72
7. Percentage Bias, SE and RMSE of Reliability Estimates of the Data Generated from the Congeneric Model with Error Correlation at 0.5 (Model 6)............. 75
8. Percentage Bias, SE and RMSE of Reliability Estimates of the Data Generated from the Correlated Factor Model with Factor Correlation at 0.2 and Independent Errors (Model 7) 82
9. Percentage Bias, SE and RMSE of Reliability Estimates of the Data Generated from the Correlated Factor Model with Factor Correlation at 0.2 and Error

Correlation at 0.2 (Model 8) 86

10. Percentage Bias, SE and RMSE of Reliability Estimates of the Data Generated from the Correlated Factor Model with Factor Correlation at 0.2 and Error Correlation at 0.5 (Model 9) 89

11. Percentage Bias, SE and RMSE of Reliability Estimates of the Data Generated from the Correlated Factor Model with Factor Correlation at 0.6 and Independent Errors (Model 10) 92

12. Percentage Bias, SE and RMSE of Reliability Estimates of the Data Generated from the Correlated Factor Model with Factor Correlation at 0.6 and Error Correlation at 0.2 (Model 11) 94

13. Percentage Bias, SE and RMSE of Reliability Estimates of the Data Generated from the Correlated Factor Model with Factor Correlation at 0.6 and Error Correlation at 0.5 (Model 12) 97

14. The Rounded Percentage of Estimation Failures in SEM.tau Estimates of Reliability 102

15. The Rounded Percentage of Estimation Failures in SEM.cong Estimates of Reliability 103

16. The Rounded Percentage of Estimation Failures in SEM.CE Estimates of Reliability 103

17. The Rounded Percentage of Estimation Failures in SEM.CF Estimates of

Reliability 104

18. The Rounded Percentage of Estimation Failures in SEM.CFCE Estimates of Reliability 105

19. Percentage Bias, SE and RMSE of Bayesian SEM Reliability of the Data Generated from the Correlated Factor Model with Factor Correlation at 0.2 and Independent Errors (Model 7) 107

20. Percentage Bias, SE and RMSE of Bayesian SEM Reliability Estimates of the Data Generated from the Correlated Factor Model with Factor Correlation at 0.2 and Error Correlation at 0.2 (Model 8)... 107

21. Percentage Bias, SE and RMSE of Bayesian SEM Reliability Estimates of the Data Generated from the Correlated Factor Model with Factor Correlation at 0.2 and Error Correlation at 0.5 (Model 9)... 108

LIST OF FIGURES

Figure

1. Graham (2006)'s Structural Equation Model for Reliability Estimation .. 28

2. Raykov (1997)'s Structural Equation Model for Reliability Estimation... 30

3. A Two-layer Hierarchical Factor Model (the Bifactor Model).................. 33

4. A Correlated Factor Model in Raykov and Shrout (2002)'s...................... 34

5. Raykov and Shrout (2002)'s Structural Equation Model for Reliability Estimation.. 35

6. The Model with Moving Average Component of Order 1 in Green & Hershberger (2000)'s 37

7. The Autoregressive Model in Green & Hershberger (2000)'s 37

LIST OF ABBREVIATIONS

BSEM: Bayesian Structural Equation Modeling

CFA: Confirmatory Factor Analysis

CRCMM: Composite Reliability for Congeneric Measures Model

CSM: Correctly Specified Model

EFA: Exploratory Factor Analysis

FA: Factor Analysis

MCMC: Markov Chain Monte Carlo

ML: Maximum Likelihood

OLS: Ordinary Least Squares

OSM: Over-specified Model

PCA: Principal Components Analysis

SEM: Structural Equation Modeling

USM: Under-specified Model

LIST OF TERMS

BSEM.CE: BSEM estimates of reliability using the congeneric model with correlated errors as analysis model

BSEM.CF: BSEM estimates of reliability using the correlated factor model with independent errors as analysis model

BSEM.CFCE: BSEM estimates of reliability using the correlated factor model with correlated errors as analysis model

BSEM.cong: BSEM estimates of reliability using the congeneric model with independent errors as analysis model

BSEM.tau: BSEM estimates of reliability using the tau-equivalent model with independent errors as analysis model

glb: the greatest lower bound to reliability

SEM.CE: SEM estimates of reliability using the congeneric model with correlated errors as analysis model

SEM.CF: SEM estimates of reliability using the correlated factor model with independent errors as analysis model

SEM.CFCE: SEM estimates of reliability using the correlated factor model with correlated errors as analysis model

SEM.cong: SEM estimates of reliability using the congeneric model with independent errors as analysis model

SEM.tau: SEM estimates of reliability using the tau-equivalent model with independent errors as analysis model

theta.g: generalized theta

CHAPTER 1 INTRODUCTION

Reliability is generally interpreted as the precision of test scores or other measurements in the field of psychometrics (Haertel, 2006; McDonald, 1999). The concern of reliability is to quantify the consistency of results from a measurement procedure across replications. Spearman (1904) first proposed to use the index of reliability to correct a correlation coefficient for attenuation due to measurement errors. These errors, as pointed out by Spearman, attenuate a correlation coefficient in a manner that cannot be remedied by increasing the number of individuals. The importance of reliability is not limited to correcting an attenuated correlation coefficient. As stated by Cronbach (1951, p. 297), "Even those investigators who regard reliability as a pale shadow of the more vital matter of validity cannot avoid considering the reliability of their measures." Currently, reliability has become one of the most commonly reported psychometric properties when evaluating the quality of an educational or psychological measure. Researchers are recommended to "provide reliability coefficients of the scores for the data being analyzed even when the focus of their research is not psychometric" (Wilkinson & APA Task Force on Statistical Inference, 1999, p. 596). Therefore, studies on reliability estimation and its corresponding coefficients are of great importance to practicing researchers.

Statistically, reliability is defined as the square of the correlation between observed scores and true scores (Spearman, 1904). Under the assumptions of the true-score model

(Lord & Novick, 1968, see details in Chapter Two 2.1), it is equal to the ratio of true score variance to observed score variance, that is,

$$\rho_{XX} = \rho^2(X,\ T) = \frac{\sigma_{XT}^2}{\sigma_X^2 \sigma_T^2} = \frac{\sigma_T^2}{\sigma_X^2}. \quad (1)$$

Thus, reliability can be interpreted as the amount of the observed variance attributable to systematic differences among the individuals in the population. Since σ_T^2 (true score variance) is unknown, reliability has to be estimated instead of being calculated directly. Spearman initially proposed equating the reliability coefficient to parallel-form or test-retest correlation. However, it is difficult to construct parallel forms and evaluate whether two or more forms of a test are parallel. It is also unreasonable to expect that examinees will be the same from one time to another due to the individual change over time or due to practice effect. It is possible that examinees will remember from one administration to another unless the time interval is quite long. In response to the difficulties inherent in using parallel forms and test-retest methods, methods for estimating reliability by a single administration of one test form have been developed.

In order to estimate reliability by a single administration of one test form, various approaches and corresponding reliability coefficients have been proposed so far. Currently, the five most influential approaches are: internal consistency, lower bound, principal components analysis (PCA), exploratory factor analysis (EFA), and structural equation modeling (SEM). These five approaches and their corresponding reliability coefficients are reviewed in detail in Chapter Two. The SEM approach is used here as a general approach which subsumes confirmatory factor analysis (CFA) approach, because CFA is a special case of SEM (Bollen, 1989). Other approaches to reliability estimation, such as analysis of variance (ANOVA) approach (Cronbach et al., 1963, 1972; Hoyt, 1941; Shrout & Joseph, 1979), and hierarchical linear modeling (HLM) approach (Bryk & Raudenbush, 1992; Wang, 2002; Snijders & Bosker, 2012), are not considered in this study. The reason lies in the fact that external factors (e.g., the test administration methods or location) other than test items need to be considered in these approaches to obtain an optimal estimate of reliability. Otherwise, these two approaches generate reliability estimates at most as good as coefficient alpha (Hoyt, 1941; Shrout & Joseph, 1979).

Facing various approaches and thus dozens of reliability coefficients derived for estimating reliability, practicing researchers are curious to know which reliability coefficient(s) performs best, and under what circumstances, because overestimation of reliability may cause false confidence in the quality of the measure being evaluated, and underestimation of reliability may cause more time and effort to revise or even redevelop the measure. Although many researchers have conducted studies to compare these reliability coefficients, the results are not always consistent and the whole picture of how these reliability coefficients perform in comparison is not well understood due to the following reasons. First, there is a severe lack of the comprehensive studies which compare reliability coefficients from various approaches. To the author's knowledge, there is no study that has simultaneously evaluated the reliability coefficients from the aforementioned five approaches. So far, the two studies conducted by Osburn (2000), and Revelle and Zinbarg (2009) can be regarded as the most comprehensive studies, each comparing more than 10 reliability coefficients yet from no more than three approaches. Using hypothetical data, Osburn (2000) compared coefficient alpha and 10 other reliability coefficients from the internal consistency approach and the lower bound approach. Revelle and Zinbarg (2009) compared 13 reliability coefficients, including the internal consistency approach, the lower bound approach and the EFA approach, with both real and hypothetical data. However, the sampling distribution properties of these reliability coefficients were not provided in these two well-known studies.

In my dissertation, reliability coefficients developed from each of the five approaches were compared. Specifically, performances of the most popular reliability coefficient (coefficient alpha) and recommended reliability coefficients from the other four approaches were evaluated under simulated conditions in a Monte Carlo study. The reliability coefficients compared in this study are respectively, coefficient alpha (the internal consistency approach), the most popular reliability coefficient in practice and literature; the glb (the lower bound approach), recommended by Sijtsma (2009), Ten Berge and Sočan (2004); generalized theta (PCA approach), recommended by Şimşek and Noyan (2013); omega (EFA approach), recommended by Revelle and Zinbarg (2009), and SEM estimates of reliability (SEM approach), recommended by Green and Yang

(2009). Prior to the Monte Carlo/simulation study for comparing these reliability coefficients, a review of reliability and classical test theory, the five approaches to reliability estimation, and the previous methodologies for reliability coefficient comparison are provided to justify the design of the current study.

In addition to a comparison of reliability coefficients, methods were proposed and examined in order to improve the quality of SEM estimates of reliability when (1) the analysis model is misspecified and (2) sample size is small. The previous simulation study (Yang & Green, 2010) on SEM estimates of reliability suggested that the SEM estimates tended to be poorer if the model was misspecified by examining the conditions of both under-specified and over-specified models. To deal with issues of misspecification, using a general model as the analysis model was proposed and tested although it is an over-specified model. To overcome the estimation problems for small sample size data, Bayesian estimation was applied in the second stage of this study. The rationale for proposing these methods are provided in Chapter 2 Section 2.4.

My dissertation consists of five chapters. The first chapter introduces the concept of reliability, the general formula for reliability estimation, and then major approaches to reliability estimation; finally briefly states the intention of the study. The second chapter reviews the relationship between reliability and the true-score model, the five approaches to reliability estimation and their corresponding reliability coefficients in detail; then reviews the methodologies in previous reliability coefficient comparison studies. In addition, the purposes of the study and research questions are specifically proposed at the end of Chapter Two. The third chapter introduces the methodology of this simulation study. The design factors for generating and analyzing data are addressed first, followed by the procedures for data generation and analysis, as well as the programs for computing the reliability coefficients in this study. Evaluation criteria for assessing the quality of these reliability coefficients are also provided. The results of the simulation study are presented in Chapter Four. The relative biases, standard errors and root mean square errors of reliability coefficients are summarized separately under each measurement model. For SEM estimates of reliability, both the results from the correctly specified and misspecified analysis models are reported. In the last chapter, each reliability coefficient examined in this study is discussed, conclusions are summarized to address the research questions, and directions for future research on reliability estimates are provided.

CHAPTER 2 LITERATURE REVIEW

In this Chapter, the relationship between reliability and the Classical Test Model, the five approaches to reliability estimation and their corresponding reliability coefficients, and methods for comparing reliability coefficients are reviewed in detail. Based on the gaps identified in the literature, the purposes of this study and research questions are explicitly stated at the end of this chapter.

2.1 Reliability and Classical Test Theory

Classical test theory may be regarded as roughly synonymous with true score theory (Lord & Novick, 1968; McDonald, 1999) since the measurement model in classical test theory is named as the true-score model. With the contribution by Guttman (1945), Lord and Novick (1968), and Novick and Lewis (1967), classical test theory is under a rigorously statistical treatment, in which the measurement model is expressed as

$$X = T + E, \tag{2}$$

where X is the random variable defined over a population of persons and taking values of the observed scores obtained on different persons, T and E are respectively the true-score and error-score random variables taking values of unobserved true scores and error scores of these different persons. In the classical measurement model, the expectation of the error-score variable is defined as zero and thus the expectation of the true score variable is equal to the expectation of the observed score, that is,

$$\mathrm{E}[T] = \mathrm{E}[X - E] = \mathrm{E}[X] - \mathrm{E}[E] = \mathrm{E}[X]. \tag{3}$$

Further, the error-score variable is independent from the true-score variable, that is, the correlation between T and E is zero:

$$\rho_{TE} = 0. \tag{4}$$

Therefore, the covariance between X and T (σ_{XT}) is equal to the true-score variance (σ_T^2).

$$\sigma_{XT} = \mathrm{E}[XT] - \mathrm{E}[X]\mathrm{E}[T] = \sigma_T^2 + \mathrm{E}[XE] - \mathrm{E}[E]\mathrm{E}[T] = \sigma_T^2. \tag{5}$$

Based on equation (5) , reliability, which is defined as the square of the correlation between observed scores and true scores, is equal to the ratio of true score variance to observed score variance, that is,

$$\rho_{XX} = \rho^2(X,\ T) = \frac{\sigma_{XT}^2}{\sigma_X^2\sigma_T^2} = \frac{(\sigma_T^2)^2}{\sigma_X^2\sigma_T^2} = \frac{\sigma_T^2}{\sigma_X^2}. \tag{6}$$

From equation (2), we have

$$\sigma_X^2 = \sigma_{(T+E)}^2 = \sigma_T^2 + \sigma_E^2 + 2\sigma_{TE}. \tag{7}$$

Since

$$\sigma_{TE} = \rho_{TE}\sigma_T\sigma_E = 0, \tag{8}$$

Equation (7) is simplified to

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2, \tag{9}$$

and thus

$$\rho_{XX} = \frac{\sigma_T^2}{\sigma_X^2} = 1 - \frac{\sigma_E^2}{\sigma_X^2}. \tag{10}$$

Therefore, reliability can be obtained by either estimating the true-score variance or error-score variance.

2.2 Five Approaches to Reliability Estimation

Among the five approaches to reliability estimation introduced in Chapter One, internal consistency and lower bound approach are the earlier attempts for estimating reliability by a single administration of one test form. Among all the reliability coefficients derived from the two approaches, Coefficient alpha (Cronbach, 1951), or named as Cronbach's alpha, is the most often used in practice and in literature. Coefficient alpha, derived from the internal consistency approach, is mathematically equivalent to Guttman's λ_3 (Guttman, 1945), derived from the lower bound approach. Since the 1970s, more reliability coefficients based on factor analytic approaches, for example, theta (Armor, 1974) and omega (McDonald, 1978; 1999), have been developed with the thriving of the factor analytic techniques. The reliability of a test can be estimated by the principal components analysis (PCA) approach (Armor, 1974; Şimşek & Noyan, 2013), exploratory factor analysis (EFA) (Revelle and Zinbarg, 2009; Şimşek & Noyan, 2013) and structural equation modeling (SEM) techniques (Brunner & Sub, 2005; Graham, 2006; Green & Hershberger, 2000; Green and Yang, 2009; Komaroff, 1997; Miller, 1995; Raykov, 1997a, 1997b, 1998, 2000, 2001; Raykov & Shrout, 2002). Particularly, examples include Raykov (1997a)'s composite reliability for congeneric measures model (CRCMM), Raykov and Shrout (2002)'s composite reliability for underlying correlated factor model and Green and Hershberger (2000)'s reliability coefficient for correlated error models. Although the five approaches differ in the models or methods for estimating the true score variance or the error score variance, they all comply with the definitional formula of reliability: $\rho_{XX} = \frac{\sigma_T^2}{\sigma_X^2} = 1 - \frac{\sigma_E^2}{\sigma_X^2}$.

2.2.1. The Internal Consistency Approach

The core of the internal consistency approach involves dividing a test into two or more constituent parts and estimating reliability based on the assumption of a certain level of consistency (e.g., tau-equivalence) of the performance across these test parts. The internal consistency refers to the interrelatedness of constituent parts of a test (Green, Lissitz & Mulaik, 1977; McDonald, 1981; Miller, 1995; Schmitt, 1996), and these test parts may be two halves of a test, or sets of items pertaining to one same reading passage or the same context, or individual items. In short, the reliability coefficients based on the internal consistency approach require an assumption of a certain degree of internal consistency and the tenability of the assumption generally affects the accuracy of these coefficients as reliability estimates.

Reliability Coefficients Requiring Parallelism of Test Parts

The problem of estimating test score reliability from a single administration of a single test form was first taken up by Spearman (1910) and Brown (1910). They independently arrived at the solution later named as Spearman-Brown procedure. There are two ways of using Spearman-Brown procedure: 1) split a test into two parallel halves and use the formula

$$\rho_{XX'} = \frac{2\rho_{X_1X_2}}{1+\rho_{X_1X_2}} \quad (11)$$

to estimate reliability; 2) divide a test into as many parts as they are parallel and use the formula

$$\rho_{XX'} = \frac{k\rho_{X_1X_2}}{1+(k-1)\rho_{X_1X_2}} \quad (12)$$

to estimate reliability (k represents the number of parallel parts).

Derivation of the Spearman-Brown formula requires the assumptions of tau equivalence and equal error variance (Novick & Lewis, 1967). Tau equivalence means that each individual in the population has identical true scores across the parallel test parts and thus the true scores in the population under consideration have the same distribution across all the parallel test parts. Besides, the population error variances of the test parts are the same. Suppose a test is divided into two parallel halves, then we have $T_1 \equiv T_2$ and $Var(E_1) = Var(E_2)$. T_1 and T_2 are the true score variables of the two split halves, and $T_1 \equiv T_2$ represents the true score variables that are identical across the two parallel halves. E_1 and E_2 are the random error variables of the split halves 1 and 2. Directly testing the assumptions of parallelism is not possible and the only available information is the distributions of the observed scores of the divided test parts. The assumptions of parallel test parts are redefined by using available information only. That is, parallel test parts assume identical observed-score distributions (i.e. equal mean in the first moment and equal variance in the second moment) and equal covariances among all the divided test parts (it is not required if a test is merely split into halves).

Assessing these assumptions are tedious (Gullikson, 1950) and splitting a test into two or more parallel parts was difficult to achieve. Thus, alternative estimation procedures of reliability which require a less stringent assumption than parallelism were developed.

Reliability Coefficients Requiring Tau-equivalence of Test Parts

Flanagan and Guttman-Rulon (Guttman, 1945; Rulon, 1939) split half coefficients were derived to estimate test reliability under a weaker assumption that two halves are (essentially) tau equivalent, dropping the requirement of equal error score variances in the parallelism required in Spearman-Brown procedure. They derived several equivalent formulas for the split-half reliability estimate by assuming essential tau-equivalence for X_1 and X_2.

Spearman-Brown formula is mathematically equal to Flanagan-Rulon (Rulon, 1939)'s formula $\rho_{XX'} = 4\sigma_{X_1X_2}/\sigma_X^2$ under the assumption of essential tau-equivalence, since

$$\sigma_T^2 = \sigma_{T_1+T_2}^2 = \sigma_{T\,1}^2 + \sigma_{T\,2}^2 + 2\sigma_{T_1T_2}, \quad (13)$$

$$\sigma_{X_1X_2} = \sigma_{(T_1+E_1)(T_2+E_2)} = \sigma_{T_1T_2} + \sigma_{T_1E_2} + \sigma_{T_2E_1} + \sigma_{T_1E_2}, \quad (14)$$

$$\sigma_{T_1T_2} = \sigma_{T_1}\sigma_{T_2}, \quad (15)$$

and

$$\sigma_{T\,1}^2 = \sigma_{T\,2}^2, \quad (16)$$

$$\sigma_T^2 = 4\sigma_{T_1}^2 = 4\sigma_{T_2}^2 = 4\sigma_{T_1T_2} = 4\sigma_{X_1X_2}. \quad (17)$$

(Note: equation $\sigma_T^2 = 4\sigma_{X_1X_2}$ is derived from the assumption of essential tau-equivalence.)

Coefficient alpha (Cronbach, 1951), or named as Cronbach alpha, was derived to estimate reliability under the assumption that all the items in a test are (essentially) tau equivalent. Coefficient alpha works for both dichotomously and polytomously scored items, which is equivalent to Kuder and Richardson (1937)'s KR 20 when items are dichotomously scored. Coefficient alpha is also equivalent to Guttman's λ_3 (one of a series of six lower bounds denoted from λ_1 to λ_6) (Guttman, 1945). Later, Ten Berge and Zegers (1978) proved that the first coefficient μ_0 of their series of lower bounds is equal to Guttman's λ_3, and thus equal to coefficient alpha.

$$\alpha = \lambda_3 = \mu_0 = \frac{n}{n-1}\left(1 - \frac{\sum \sigma_i^2}{\sigma_X^2}\right). \quad (18)$$

Coefficient alpha has become the most popular reliability coefficient because it requires no split of a test for estimating test reliability. However, people may forget that the assumption that all items in a test are (essentially) tau equivalent is more difficult to hold in practice compared with the assumption that the two split halves are (essentially) tau equivalent. Although warnings have been given by researchers that Cronbach alpha is an inaccurate estimate of reliability when the assumption of (essential) tau equivalence is violated (Graham, 2006; Green, Lissitz & Mulaik, 1977; Novick & Lewis, 1967; Osburn, 2000; Zimmerman, Zumbo & Lalond, 1993), statistical test of the assumption was seldom conducted before applying alpha or other internal consistency reliability coefficients.

Reliability Coefficients Requiring Congeneric Test Parts

Since the assumption of (essential) tau equivalence may not hold in practice, estimates of reliability under a more relaxed assumption of congeneric forms (Jöreskog, 1971) were developed. Congeneric forms require neither (essential) tau equivalence nor equal error score variances. Trues score in congeneric forms are linearly related. If we have two split halves with congeneric forms, then $\mathrm{T}_1 = \varphi(\mathrm{T}_2)$, where φ represents a linear function. Several coefficients, e.g., Kristof's coefficient (Kristof, 1974), Angoff-Feldt coefficient (Feldt & Brennan, 1989), Raju's coefficient (Raju, 1977), Feldt-Gilmer coefficient (Gilmer & Feldt, 1983), Feldt's coefficient (Feldt & Brennan, 1989) were derived under the assumption of congeneric forms (see Osburn, 2000, for a detailed summary of these coefficients).

For three congeneric parts of unknown length, the model is just identified, meaning that there are exactly as many observed variances and covariances as the number of parameters to be estimated (Lord & Novick, 1968; Kristof, 1974). Kristof (1974) reports that his three-part division appears to give quite stable results across alternative partitions of a test into two parts. For a test divided into more than three congeneric parts, more parameters have to be estimated and thus the estimation process becomes more complicated. Jöreskog (1971) proposed the congeneric model

$$X_i = U_i + \beta_i T + E_i,\ \ (i = 1, 2, \dots, p), \qquad (19)$$

and a maximum likelihood solution for parameter estimation. The reliability for the *ith* part of a test is estimated by

$$\hat{\rho}_i = \frac{\hat{\beta}_i^2}{\hat{\beta}_i^2 + \hat{\theta}_i^2} , \tag{20}$$

where $\hat{\theta}_i^2$ is the estimated variance of E_i. The reliability for the test composed by these congeneric parts is then equal to

$$\rho = \frac{\boldsymbol{\alpha}'\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{\alpha}}{\boldsymbol{\alpha}'\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{\alpha} + \boldsymbol{\alpha}'\boldsymbol{\Theta}_\epsilon{}'\boldsymbol{\alpha}} , \tag{21}$$

where $\boldsymbol{\alpha}' = (\alpha_1, \alpha_2, \ldots, \alpha_m)$ is the relative weight vector and it is equal to $(1,1, \ldots ,1)$ when these congeneric parts are unweighted; $\boldsymbol{\beta}$ is the vector containing the β_is and $\boldsymbol{\Theta}_\epsilon$ is the measurement error variance covariance matrix.

Internal consistency reliability estimates often require the division of a test into more than two separate parts. The sampling variance of the estimated reliability coefficient is related to the number of units into which the test is divided. That is, greater precision is obtained if the test can be divided into a larger number of separate parts (Kristof, 1963). However, division into more parts is only better if model assumptions are satisfied (Brennan, 2001a). Therefore, the common practice of basing internal consistency reliability estimation on division of tests into the smallest possible units, the individual items, is not problematic only when model assumptions are satisfied. Internal consistency approach to reliability estimation is within the framework of classical test theory, and therefore all the models are under the umbrella of unidimensionality, meaning only one latent trait (true score variable) is assumed to exist in a test. The problem of violating the unidimensionality assumption vanishes in the following four approaches (i.e., the lower bound, PCA, EFA and SEM) to reliability estimation because they require no assumption of unidimensionality.

2.2.2. The Lower Bound Approach

The formula $\rho = 1 - \frac{\text{MaxTr}\sum_E}{\mathbf{1}'\sum_X \mathbf{1}}$ can be regarded as a general equation for the lower bound estimation to reliability, where $\sum_E$ is the error score covariance matrix and $\sum_X$ is the observed score covariance matrix, and $\text{MaxTr}\sum_E$ represents the maximum trace of the error score covariance matrix. By changing the constraints for estimating $\text{MaxTr}\sum_E$, different lower bound estimates of reliability can be derived. Jackson and Agunwamba (1977) identified the specific constraint for each of Guttman's six lower bounds (1945) and provided more types of constraints to derive other algebraic lower bounds.

Guttman's Lower Bounds to Reliability

Guttman (1945) named his six coefficients of reliability as lower bounds and developed the concept of lower bounds by proposing the idea of bounding the estimation of the true test score variance or error test score variance to derive the estimates of reliability. Bounding here means using specific constraints/inequalities to minimize the true test score variance or maximize error test score variance, thus the estimate of reliability can always be lower than or at most equal to the true reliability.

Guttman's lambda1 adopts the constraint $0 \le \theta_i \le \sigma_{ii}$ (θ_i is the trace element of $\sum_E$) to estimate $\text{MaxTr}\sum_E$, thus the maximum trace of $\sum_E$ is the trace of $\sum_X$. The equation of lambda1 is therefore given as

$$\lambda_1 = 1 - \frac{\mathrm{Tr}\Sigma_X}{\mathbf{1}'\Sigma_X \mathbf{1}}. \tag{22}$$

The constraint for deriving lambda2 is that the determinant of the true score covariance matrix must be non-negative (Jackson & Agunwamba, 1977), which leads to

$$t_i t_j \geq \sigma_{ij}^2 \ (i \neq j), \tag{23}$$

where t_i and t_j are the trace elements, i.e., the elements along the principal diagonal, of Σ_T. Based on the inequality $t_i t_j \geq \sigma_{ij}^2 \ (i \neq j)$, $\mathrm{MaxTr}\Sigma_\mathrm{E}$ is estimated by the following inequality,

$$\sum_{j=1}^{n} \theta_j \leq \sum_{j=1}^{n} \sigma_j^2 - \sqrt{\frac{n}{n-1}\left(\sum_{j=1}^{n}\sum_{i=1\ (i\neq j)}^{n} \sigma_{ij}^2\right)} \ , \tag{24}$$

where θ_j represents the jth trace element of Σ_E, and n is the number of items. Then from the general equation, lambda2 can be obtained as

$$\lambda_2 = 1 - \frac{\sum_{j=1}^{n} \sigma_j^2 - \sqrt{\frac{n}{n-1}\left(\sum_{j=1}^{n}\sum_{i=1\ (i\neq j)}^{n} \sigma_{ij}^2\right)}}{\mathbf{1}'\Sigma_X\mathbf{1}}. \tag{25}$$

Guttman's lambda3 (mathematically equivalent to coefficient alpha) then applies the constraint $\mathrm{t_i} + \mathrm{t_j} \geq 2\sigma_{\mathrm{ij}}(\mathrm{i} \neq \mathrm{j})$ to estimate the minimum trace of Σ_T. By summing the inequality $\mathrm{t_i} + \mathrm{t_j} \geq 2\sigma_{\mathrm{ij}} \ (\mathrm{i} \neq \mathrm{j})$ over $\frac{\mathrm{n(n-1)}}{2}$ pairs of items, we can obtain

$$\sum_1^n \mathrm{t_i} \geq \sum_1^n\sum_1^n \sigma_{\mathrm{ij}} \geq (\mathrm{n}-1)^{-1}\sum_1^n\sum_1^n \sigma_{\mathrm{ij}} \ (\mathrm{i} \neq \mathrm{j}). \tag{26}$$

Thus,

$$\mathbf{1}'\Sigma_T\mathbf{1} = \sum_1^n \mathrm{t_i} + \sum_1^n\sum_1^n \sigma_{\mathrm{ij}} \geq (\mathrm{n}-1)^{-1}\sum_1^n\sum_1^n \sigma_{\mathrm{ij}} + \sum_1^n\sum_1^n \sigma_{\mathrm{ij}} = \frac{\mathrm{n}}{\mathrm{n}-1}\sum_1^n\sum_1^n \sigma_{\mathrm{ij}}, \tag{27}$$

And from the general equation, we have

$$\lambda_3 = \frac{\mathrm{n}}{\mathrm{n}-1}\left(1 - \frac{\sum_1^n \sigma_{\mathrm{ii}}}{\mathbf{1}'\Sigma_X\mathbf{1}}\right). \tag{28}$$

To make use of the constraint that $\sum_T$ is positive semi-definite, that is, $\mathbf{u}'\sum_T\mathbf{u} \geq 0$, ($\mathbf{u}$ refers to any n-vector with k elements either $+1$ or -1) and the remaining elements zeros), let k be equal to n and there are 2^{n-1} such vectors as $\mathbf{u}$ yield different $\mathbf{u}'\sum_T\mathbf{u}$. Since $\sum_X = \sum_T + \sum_E$, then

$$\mathbf{u}'\Sigma_X\mathbf{u} = \mathbf{u}'\Sigma_T\mathbf{u} + \mathbf{u}'\Sigma_E\mathbf{u}, \tag{29}$$

and thus

$$\mathbf{u}'\sum_X\mathbf{u} \geq \mathbf{u}'\sum_E\mathbf{u} = \sum_{j=1}^{n}\theta_j. \tag{30}$$

Replace the $\text{MaxTr}\sum_E$ with $\mathbf{u}'\sum_X\mathbf{u}$ in the general formula $\rho = 1 - \text{MaxTr}\sum_E/\sum_X$, we have

$$\lambda_4 = 1 - \frac{\mathbf{u}'\Sigma_X\mathbf{u}}{\mathbf{1}'\Sigma_X\mathbf{1}}. \tag{31}$$

Guttman's lambda4 has therefore 2^{n-1}possible values, including any split-half reliability (Note here each half does not necessarily have the n/2 number of items since $+1$ or -1 are randomly assigned to the k elements in the vector $\mathbf{u}$). Among all the lambda4s, only the maximized lambda4 is of interest (Callender & Osburn, 1979; Jackson & Agunwamba, 1977).

Guttman's lambda5 replaces the diagonal values of $\sum_T$ with twice the square root of the sum of squared interitem covariances from the column which has the maximized sum of squared interitem covariances, and we have

$$\lambda_5 = 1 - \frac{\sum_1^n \sigma_{ii} - 2\left(\sum \sigma_{ij\max}^2\right)^{\frac{1}{2}}}{1'\Sigma_X 1}, \tag{32}$$

where j_{max} means the position of the column having the maximized sum of squared interitem covariances. The constraint for deriving such a lower bound to the sum of the diagonal values of $\sum_T$ (and hence a lower bound to reliability) is the inequality that the arithmetic mean of two positive numbers is no less than their geometric mean.

Guttman's lambda6 sets the residual variance in the regression of the component X_i on the remaining components scores as the upper bound of the error variance θ_i, therefore we obtain

$$\lambda_6 = 1 - \frac{\Sigma(\Sigma_X{}^{-1})_{ii}{}^{-1}}{\mathbf{1}'\Sigma_X\mathbf{1}}. \tag{33}$$

Although Guttman initially recommended the use of lambda3 (i.e., coefficient alpha) and lambda4 among the six proposed measures, this recommendation was mainly due to the consideration of the relative ease of computation of the two coefficients. As proved by Guttman (1945), lambda2 was always equal to or greater than coefficient alpha. Furthermore, lambda5 and lambda6 were generally lower than lambda2 except for some limited conditions. In addition, Ten Berge and Zegers (1978) demonstrated an infinite series of successive improvement to Guttman's lower bounds where coefficient alpha and lambda2 were the first two in the series. As concluded by the authors, the series did not improve much after lambda2.

The Greatest Lower Bound

Based on the lower bound concept, Jackson and Agunwamba (1977) proposed more types of constraints to derive other algebraic lower bounds, including the well-known greatest lower bound (glb). According to Jackson and Agunwamba (1977), the glb can be expressed as

$$\rho_{glb} = \frac{\sum_{j=1}^{n}\sum_{i=1\,(i\neq j)}^{n}\sigma_{ij} + \mathrm{MinTr}\Sigma_{\mathrm{T}}}{\mathbf{1}'\Sigma_X\,\mathbf{1}} = 1 - \frac{\mathrm{MaxTr}\Sigma_{\mathrm{E}}}{\mathbf{1}'\Sigma_X\,\mathbf{1}}, \tag{34}$$

where $\sum_{\mathrm{T}}$, $\sum_{\mathrm{E}}$ and $\sum_{\mathrm{X}}$ are the true score, error score and observed score variance-covariance matrices, respectively, σ_{ij} is the covariance between item i and j, n is the number of items in the test, $\mathrm{MinTr}\sum_{\mathrm{T}}$ is the minimal trace of the true score variance, $\mathrm{MaxTr}\sum_{\mathrm{E}}$ is the maximal trace of the true score variance, and $\mathbf{1}'\sum_{\mathrm{X}}\mathbf{1}$ is the total score variance. The glb for the reliability of the total score on a test is derived by maximizing the trace of error covariance matrix $\sum_E$, which is equivalent to minimizing the trace of true score covariance matrix $\sum_T$, subject to the conditions that both $\sum_T$ and $\sum_E$ are non-negative definite matrix.

Different from the traditional reliability coefficients calculated from a single formula, the greatest lower bound is derived by an iterative procedure, which was initially proposed by Bentler (1972), and then Woodhouse and Jackson (1977), later improved by Bentler and Woodward (1980), and Ten Berge, Snijders and Zegers (1981). The computational algorithm described by Ten Berge and Kiers (1991) was programmed in a computer program named MRFA2 to obtain the glb (Ten Berge & Kiers, 2003). Since the estimation of the glb is to minimize the linear function (1) subject to the constraint that an affine combination of symmetric matrices ($\sum_T$ and $\sum_E$) is positive semidefinite, the glb estimation could be obtained by semidefinite programming (Vandenberghe & Boyd, 1996), and the R package “RcSEp” (Bravo, 2013) provided in the R interface utilizes the semidefinite programming. Specifically, the glb can be calculated by formula (34) and the optimized $\text{MinTr}\sum_T$ can be obtained using the function cSEp in the “RcSEp”. Although the calculation of the glb is more complicated because of the algorithm involved, it is worth studying as it is theoretically the optimal lower bound estimate of reliability (Jackson & Agunwamba, 1977; Sijtsma, 2009; Ten Berge & Sočan, 2004; Woodhouse & Jackson, 1977).

2.2.3. The PCA Approach

Armor (1974) introduced an approach to reliability estimation based on principal component factor analysis (PCA) and proposed the corresponding estimate of reliability named as *theta*. He claimed that principal components analysis offered the most straightforward and precise connection between reliability and factor scaling, and thus *theta* could assess optimal reliability.

The basic hypothesis of component analysis is that, given a set of p items, the score of a subject on each item can be decomposed into p number of independent components or factors. The lack of correlation among principal components is a useful property as it means that the components are measuring different "dimensions" in the data. Among these p factors, only a small number of factors (e.g., m with $m < p$) that account for a relatively large proportion of the total item variation have the substantial meanings and considered as the non-error factors. Each item can contribute differently to a non-error factor and thus items may have different factor loadings on the factor (i.e., the weights represent the contribution of items to the factor). Moreover, principal components are ordered so that the first component exhibits the greatest amount of the variation, the second component exhibits the second greatest amount of the variation, and so on.

There are two general cases in principal component approach to reliability estimation. The first case is to assume a single factor solution. That is, the first principal component is sufficient for accounting for the variation of the scale. The reliability coefficient theta of the composite scores based on this single factor solution is expressed as

$$\theta = \left(\frac{p}{p-1}\right)\left(1 - \frac{1}{\lambda_1}\right), \tag{35}$$

where λ_1 is the first eigenvalue of a principal component solution. It is mathematically equivalent to the maximum possible value of alpha (the alpha for a composite scale formed by weighting items according to their principal component factor loadings) (Lord, 1958).

The second situation is to assume a multiple-factor solution with rotated factors. The formula for the reliability coefficient theta based on the multiple-factor solution with rotated factors is given by

$$\theta_k = \left(\frac{p}{p-1}\right)\left(1 - \sum_{h=1}^{m} \frac{\phi_{hk}^2}{\lambda_h}\right), \tag{36}$$

where ϕ_{hk}^2 refers to the squared correlation between the original unrotated scores for factor h and the rotated scores for the new factor k. θ_k is the proper formula for estimating reliability when the complete set of the rotated factor scores (i.e., $m = p$)

from a principal components analysis is used (Armor, 1974).

Based on Armor's work, Şimşek and Noyan (2013) proposed generalized theta. Generalized theta is based on the eigenvalues of the principal components up to a pre-specified number of factors. The formula for generalized theta is written as

$$\theta_G = \left(\frac{p}{p-m}\right)\left(1 - \frac{m}{\sum_{i=1}^{m}\lambda_i}\right), \tag{37}$$

where m refers to a pre-specified number of factors. Generalized theta is a generalized version of Armor's theta and is equal to the true reliability when the dimensions are orthogonal and the items clustered within each dimension are parallel.

2.2.4. The EFA Approach

Although principal components analysis (PCA) and exploratory factor analysis (EFA) are often referred to collectively as factor analysis (FA), they differ in both mathematical and conceptual terms. The difference between PCA and EFA in mathematical terms lies in the diagonal elements of the correlation matrix for analysis. In PCA, all diagonal elements in the correlation matrix are 1s meaning that all of the variance in the matrix is to be accounted for by principal components. In contrast, in EFA, all diagonal elements are equal to what are called "communalities" meaning that only the variance shared with other variables is to be accounted for. The difference between PCA and EFA in conceptual terms is that PCA analyzes variance and EFA analyzes covariance (Tabachnick and Fidell, 2007, p. 635). Introduced by McDonald (1978) under the exploratory factor analytic framework, reliability can be estimated by coefficient omega, which is expressed as

$$\omega = 1 - \frac{\sum_{i=1}^{n}(1-h_i^2)}{\mathbf{1}'\Sigma_X\,\mathbf{1}}, \tag{38}$$

where h_i^2 is the communality of the *ith* item, assuming items have been standardized. McDonald's omega uses the estimates of uniqueness variances from factor analysis to represent error variances. Revelle and Zinbarg (2009) proposed to use a particular exploratory factor analytic approach to estimate omega, specifically, the application of higher order factor analysis with a Schmid–Leiman transformation (Schmid & Leiman, 1957). This procedure of estimating omega proposed by Revelle and Zinbarg (2009) is based on a decomposition of the variance of a test score into variances due to a general factor, variances due to a set of group factors, and uniqueness variances, which is the sum of undistinguishable specific variance and random error variance. The whole estimation procedure for omega was programmed in the omega function in the R package "psych" (Revelle, 2013). Except EFA with the Schmid–Leiman transformation, other EFA methods have also been adopted in reliability estimation. For example, Şimşek and Noyan (2013) used principal factor analysis with Promax rotation to compute omega.

The estimation of coefficient omega has also been extended into the field of confirmatory factor analytic models or structural equation models (McDonald, 1999). Omega derived from a specific confirmatory factor analytic model or structural equation model is usually in the form of

$$\omega = \mathbf{1}'\boldsymbol{F\Phi F}'\mathbf{1}/\mathbf{1}'\textstyle\sum_X \mathbf{1}, \tag{39}$$

where $\boldsymbol{F}$ represents factor loading matrix and $\boldsymbol{\Phi}$ factor correlation matrix. Currently, most researchers prefer to treat the coefficients derived from CFA or SEM as the SEM reliability coefficients than as coefficient omega (Green & Hershberger, 2000; Green and Yang, 2009; Komaroff, 1997; Miller, 1995; Raykov, 1997a, 1997b, 1998, 2000, 2001; Raykov & Shrout, 2002).

2.2.5. The SEM Approach

Structural Equation Modeling (SEM) is a general framework for modeling of relationships in multivariate data (Bollen, 1989). It can be roughly understood as a combination of two well-known classical statistical techniques: factor analysis and path analysis (regression). Consequently, SEM comprises two parts: a measurement model part (confirmatory factor analysis), and a structural model part (path analysis). Confirmatory factor analysis (CFA) is frequently used as a first step to assess the proposed measurement model in a structural equation model.

The general representation of structural equations with latent variables is expressed as (Bollen, 1989):

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}, \tag{40}$$

where $\mathbf{B}$ is the m x m coefficient matrix, $\boldsymbol{\Gamma}$ is the m by n coefficient matrix, $\boldsymbol{\eta}$ is a m x 1 vector that contains m latent variables, $\boldsymbol{\xi}$ is a n x 1 vector that contains n latent independent variables, and $\boldsymbol{\zeta}$ is the p x 1 vector of errors (residuals) in the equations. The measurement model of SEM specifies the relationship of the latent variables to the observed variables. The mathematical formulas of the measurement model of SEM are as follows:

$$\mathbf{y} = \boldsymbol{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\epsilon}, \tag{41}$$

$$\boldsymbol{x} = \boldsymbol{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta}, \tag{42}$$

where $\mathbf{y}$ is p x 1 vector of endogenous variables (dependent variables), $\boldsymbol{\eta}$ is a m x 1 vector that contains m latent variables ($m \leq p$), Λ_y is a p x m matrix that contains the factor loadings from m latent variables to p observed dependent variables , $\boldsymbol{\epsilon}$ is a p x 1 vector that contains the measurement errors for the observed dependent variables. Similarly, $\mathbf{x}$ is q x 1 vector of exogenous variables (independent variables), $\boldsymbol{\xi}$ is a n x 1 vector that contains n latent independent variables ($n \leq q$), $\boldsymbol{\Lambda}_x$ is a q x n matrix that contains the factor loadings from n latent independent variables to q observed independent variables, $\boldsymbol{\delta}$ is a q x 1 vector that contains the measurement errors for the observed independent variables. The observed variables, regardless of whether they are dependent or independent variables, are named as indicators or manifest variables in the literature of SEM.

The covariance matrix for $\mathbf{x}$ is the expected value of $\mathbf{xx}'$, mathematically, that is,

$$\Sigma_{\mathbf{x}} = \mathrm{E}(\mathbf{xx}') = \mathrm{E}(\boldsymbol{\Lambda}_{\mathbf{x}}\boldsymbol{\xi} + \boldsymbol{\delta})(\boldsymbol{\Lambda}_{\mathbf{x}}\boldsymbol{\xi} + \boldsymbol{\delta})' = \boldsymbol{\Lambda}_{\mathbf{x}}\boldsymbol{\Phi}\boldsymbol{\Lambda}_{\mathbf{x}}' + \boldsymbol{\Theta}_{\delta}, \tag{43}$$

where $\boldsymbol{\Phi}$ is the covariance matrix of the latent factors $\boldsymbol{\xi}$ and $\boldsymbol{\Theta}_{\delta}$ is the q x q matrix containing the error variances and covariances for the *x* variables. The covariance matrix for $\mathbf{y}$ is the expected value of $\mathbf{yy}'$, which is derived as follows:

$$\Sigma_{\mathbf{y}} = \mathrm{E}(\mathbf{yy}') = \mathrm{E}(\boldsymbol{\Lambda}_{\mathbf{y}}\boldsymbol{\eta} + \boldsymbol{\epsilon})(\boldsymbol{\Lambda}_{\mathbf{y}}\boldsymbol{\eta} + \boldsymbol{\epsilon})' = \boldsymbol{\Lambda}_{\mathbf{x}}\mathbf{E}(\boldsymbol{\eta\eta}')\boldsymbol{\Lambda}_{\mathbf{x}}' + \boldsymbol{\Theta}_{\epsilon}, \quad (44)$$

and

$$\mathbf{E}(\boldsymbol{\eta\eta}') = (\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\Gamma\Phi\Gamma}' + \boldsymbol{\Psi})(\mathbf{I} - \mathbf{B})^{-1\prime}, \quad (45)$$

thus

$$\Sigma_{\mathbf{y}} = \boldsymbol{\Lambda}_{\mathbf{y}}[(\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\Gamma\Phi\Gamma}' + \boldsymbol{\Psi})(\mathbf{I} - \mathbf{B})^{-1'}]\boldsymbol{\Lambda}_{\mathbf{y}}' + \boldsymbol{\Theta}_{\epsilon}, \quad (46)$$

where $\boldsymbol{\Psi}$ is the covariance matrix for $\boldsymbol{\zeta}$ and $\boldsymbol{\Theta}_{\epsilon}$ is the p by p matrix containing the error variances and covariances for the *y* variables. Suppose $\mathbf{B}$ and $\boldsymbol{\Gamma}$ are zero matrix, that is, no direct cause specified to latent variables $\boldsymbol{\eta}$, then $\boldsymbol{\eta} = \boldsymbol{\zeta}$ and $\Sigma_{\mathbf{y}}$ is simplified as $\boldsymbol{\Lambda}_{\mathbf{y}}\boldsymbol{\Psi}\boldsymbol{\Lambda}_{\mathbf{y}}' + \boldsymbol{\Theta}_{\epsilon}$. Under that situation the SEM model is equal to the CFA model.

The general formula for estimating reliability in SEM can be expressed as

$$\rho_{SEM} = 1 - \frac{\mathbf{1}'\Theta_{\varepsilon}\,\mathbf{1}}{\mathbf{1}'\Sigma\mathbf{1}}, \quad (47)$$

where $\sum$ is the estimated population covariance matrix and $\boldsymbol{\Theta}_{\varepsilon}$ is a general matrix which represents the $\boldsymbol{\Theta}_{\delta}$ when the measures are observed independent variables and the $\boldsymbol{\Theta}_{\epsilon}$ when the measures are observed dependent variables.

SEM for Reliability Estimation---Unidimensional Models

A number of researchers have discussed reliability estimation within an SEM framework when items are unidimensional (Fleishman & Benson, 1987; Green & Hershberger, 2000; Komaroff, 1997; Miller, 1995; Raykov, 1997a, 1997b, 1998, 2001; Zimmerman, Zumbo, & Lalonde, 1993; Graham, 2006). Among all the models raised by these researchers, Graham (2006)'s and Raykov (1997)'s models are reviewed here because of the generality of these models.

In Graham (2006)'s model, the composite observed variable (X), which is the sum of the scores of components, is created. The variance of the composite observed variable X is obtained by adding the variances of the individual observed variables (X_1, X_2, etc.) while taking into account the shared variance of the individual observed variables. Graham (2006)'s model mainly follows Miller (1995)'s essentially tau equivalent model except there are no constraints between the composite true variable (T) to the individual item variables (X_1, X_2, etc.). The path diagram of this model is represented in the following figure.

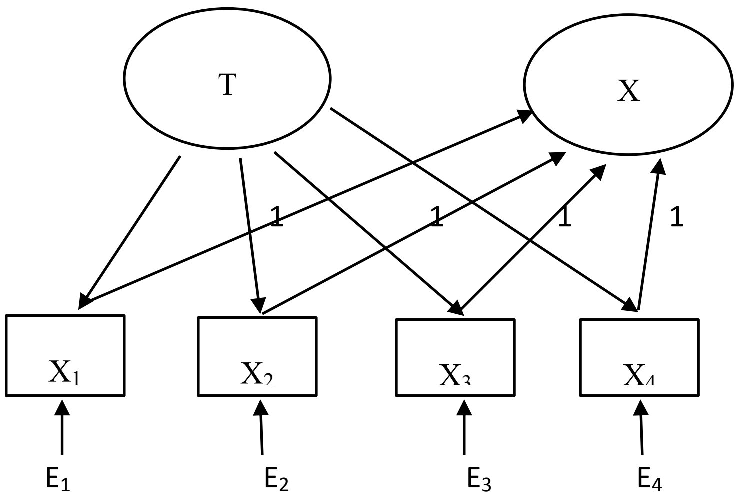

Figure 1. Graham (2006)'s Structural Equation Model for Reliability Estimation

The model illustrated in Figure 1 nests the three classical measurement models, respectively: the parallel model, the (essentially) tau-equivalent model and the congeneric model. That is, we can transform a general structural model to any of the classical measurement models by corresponding specifications. To achieve a parallel model, each of the paths from the composite true variable (T) to the individual item variables (X_1, X_2, etc.) are set to 1 in Figure 1, implying that each item variable measures the same latent variable in the same degree. Additionally, the individual item error variances are constrained to be equal to each other. The specification of the (essentially) tau-equivalent model is identical to the parallel model path diagram, except that error variances are not constrained to equality. To specify the congeneric model, the path from the latent true variable to one of the measured items is set to 1 (which is specified by default in SEM programs) or set the variance of the latent variable to 1, whereas the other paths from the true variable to the items are set free to be estimated. Any of the measured items can be chosen as the scaling variable, with no effect on the outcome of the model.

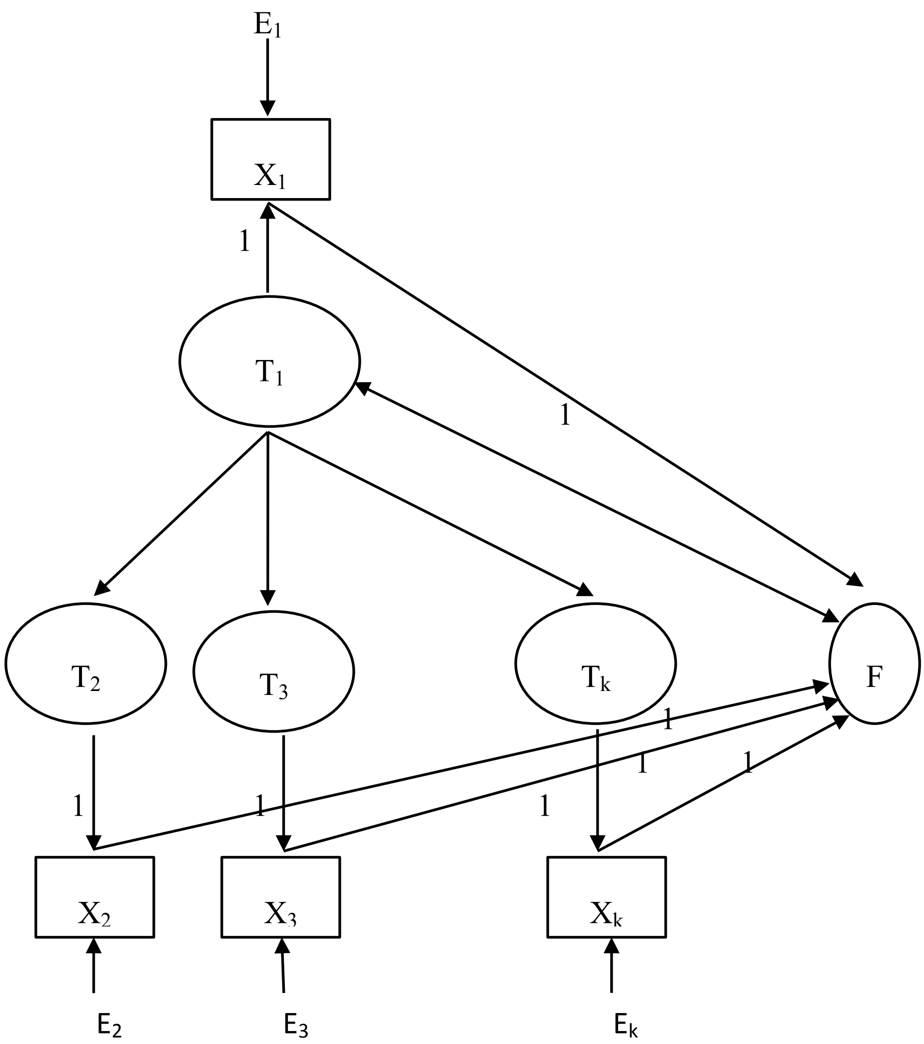

Figure 2. Raykov (1997)'s Structural Equation Model for Reliability Estimation

Raykov (1997) proposed a structural equation model for estimation of composite reliability (illustrated in Figure 2), which is equivalent to Graham's model. In effect, Raykov (1997) complicated the model presented in Figure 1 by: (a) adding a group of true score variables (T_2, T_3, etc.), (b) naming the composite observed variable X as a phantom variable F_2 and (c) estimating the correlation between the latent variable T_1 (equal to the composite true score variable in the model in Figure 1) and the composite variable F_2

(equal to the X in the model presented in Figure 1). The only difference between Graham (2006)'s and Raykov (1997)'s model is that the correlation between T_1 and F_2 (the determinant of reliability) is estimated in Raykov (1997)'s model. The square of the correlation between T_1 and F_2 in Raykov's model is equal to the ratio of the true score variance to the observed score variance, and this ratio can be calculated using Graham's model. In this study, Graham's analysis model is used for unidimensional models analysis because the ratio of the true score variance to the observed score variance is normally used to calculate reliability in SEM estimates of reliability (see Yang & Green, 2010; Yang & Green, 2011) and it is consistent with the formula used for multidimensional models analysis.

SEM for Reliability Estimation---Multidimensional Models

Researchers have also proposed SEM methods for estimating reliability when items are multidimensional (Brunner & Sub, 2005; Raykov, 1998; Raykov & Shrout, 2002; Yang & Green, 2010). Brunner and Sub (2005) specified a "nested-factor model", following the terminology of Gustafsson and Balke (1993), to represent the structure of the Berlin Intelligence Structure Test (BIS Test). The nested factor model defined by Gustafsson and Balke (1993) is in effect the hierarchical factor model, which is the more traditional terminology adopted by McDonald (1985), Schmid & Leiman (1957), Tucker (1940), and Wherry (1959). The hierarchical factor models or the nested factor models, unlike the higher-order factor models, are the models with all factors at the same (first-order) level but different in their clusters of related manifest variables (Yung, Thissen, & McLeod, 1999). As illustrated in Figure 3, the factors in the first layer (F_1 and F_2) partition the manifest variables into clusters so that each factor has a distinct cluster of related manifest variables. The next layer of factors (G) in the hierarchical factor model again partitions the manifest variables into clusters. However, this time each cluster contains at least two clusters of manifest variables that are formed in the previous layer. Brunner & Sub (2005)'s model has eight orthogonal factors in their model, including the general factor (G), four operative factors (Mental Speed, Memory, Reasoning, and Creativity), and three content factors (Figural Ability, Verbal Ability, and Numerical Ability). The two-layer hierarchical pattern with the second layer containing all the clusters of manifest variables is equivalent to the bifactor model (Holzinger & Swineford, 1937; Rindskopf & Rose, 1988).

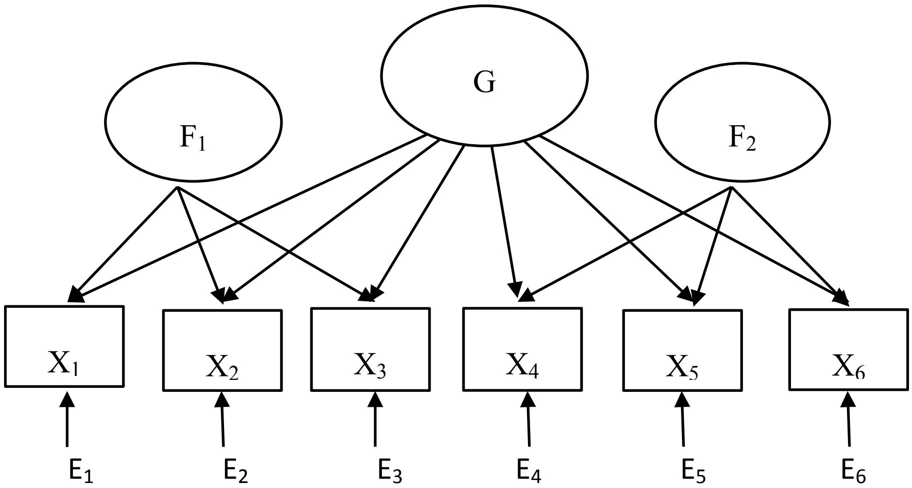

Figure 3. A Two-layer Hierarchical Factor Model (the Bifactor Model)

Raykov and Shrout (2002) proposed an SEM method to estimate reliability for measures with an underlying correlated factors structure. In Raykov and Shrout's model, the set of test parts is not homogeneous overall, but each subscale is substantially to highly interrelated with one another and these subscales are congeneric. As shown in Figure 4, there are two latent traits η_1 and η_2 with the manifest variables X_1 to X_4 loaded on η_1 and X_3 to X_6 loaded on η_2. Instead of using the model in Figure 4, they proposed the model in Figure 5 for the composite reliability estimation. In the model in Figure 5, the first added latent variable η_3 is the composite score of the manifest variables from X_1 to X_6 with all the loadings set to 1, while the second added latent variable η_4 is the composite latent score of the latent variables η_1 and η_2. To relate this variable to the measured components, the path from η_1 to η_4 is set equal to the sum of the loadings of the manifest measures X_1 to X_4 that assess η_1 (i.e., $\lambda_{11} + \lambda_{21} + \lambda_{31} + \lambda_{41}$), and the path from η_2 into η_4 is constrained to be equal to the sum of the loadings of the manifest variables X_3 to X_6 measuring η_2 (i.e., $\lambda_{32} + \lambda_{42} + \lambda_{52} + \lambda_{62}$). Although the model in Figure 5 seems more complex than the one in Figure 4, these two models are equivalent in terms of estimation since there is no new free parameter to be estimated in the model in Figure 5. In this study, the model in Figure 4 is used as the analysis model with constraints modeling η_3 and η_4, hence it is equivalent to the model in Figure 5. When purely using the model in Figure 4, the general formula expressed in equation (47) (see page 27) can be used to calculate the SEM reliability coefficient.

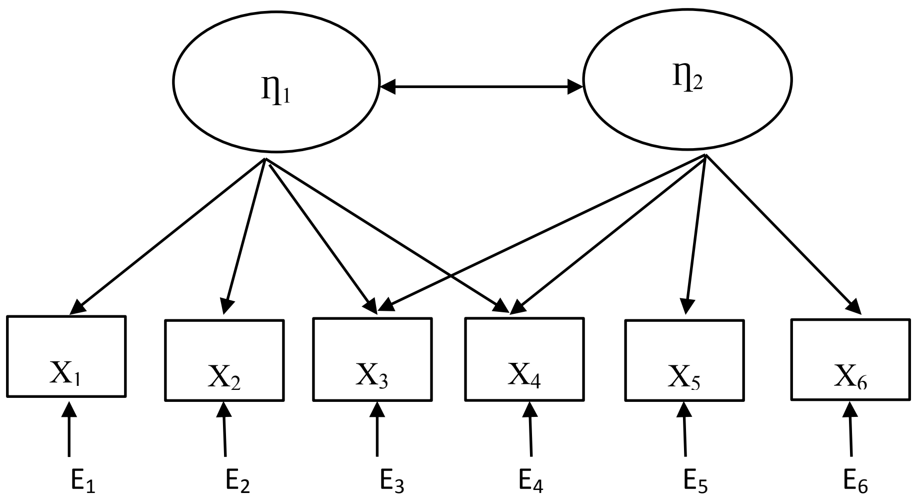

Figure 4. A Correlated Factor Model in Raykov and Shrout (2002)'s

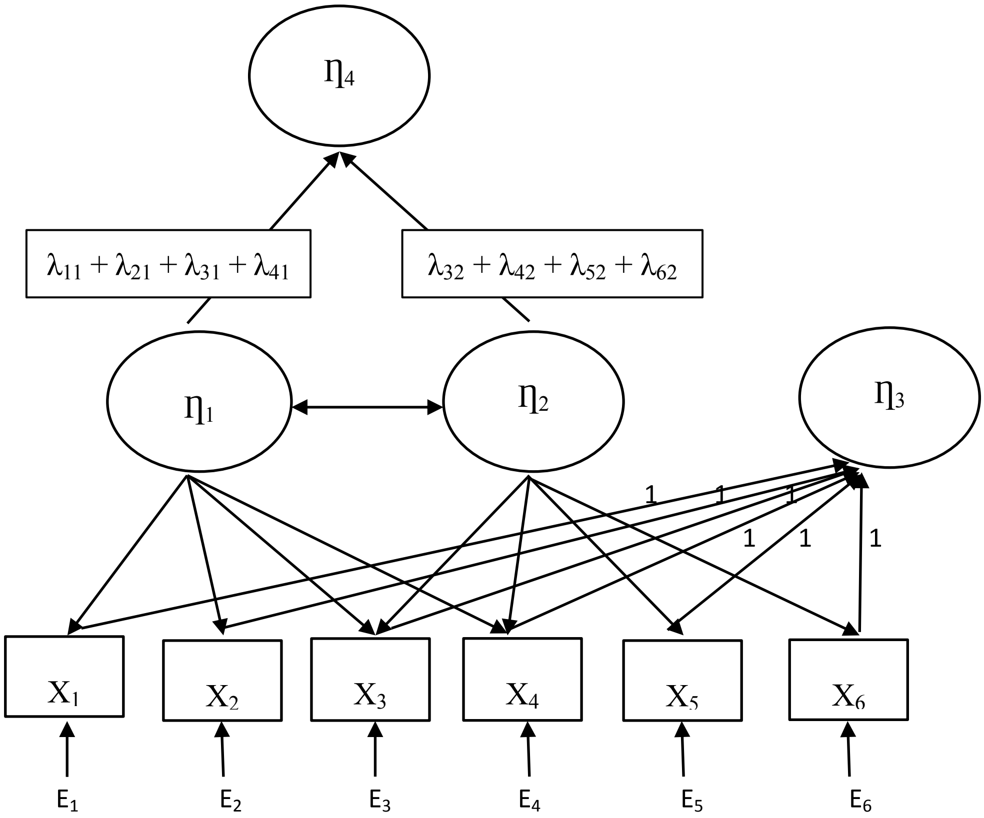

Figure 5. Raykov and Shrout (2002)'s Structural Equation Model for Reliability Estimation.

SEM for Reliability Estimation---Correlated Error Models

In SEM, correlated errors are frequently treated as due to factors left unspecified by a researcher's model rather than pure measurement errors. In that case, the error covariance between items is not counted as the source of unreliable variance of the test. However, some researchers (Green & Hershberger, 2000; Green, 2003; Raykov, 1998; Rozeboom, 1966, 1989; Zimmerman, Zumbo, & Lalonde, 1993) argued that the correlated errors may be due to random measurement errors, an *unreliable* component of measures. Green & Hershberger (2000) indicated that if random measurement errors on earlier items were allowed to affect directly or indirectly scores on later items, the errors would contribute to the covariances between items. For example, Becker (2000) and Green (2003) argued that transient conditions, such as mood, can cause correlated errors among items.

Green and Hershberger (2000) proposed two structural equation models: the model with direct effect on the measurement error of an item and the one with indirect effect. The former (see Figure 6) is the model in which only the parameters for immediately preceding error terms can be nonzero and covariances between error composites for this model would be zero except for adjacent items. The second (see Figure 7) is distinct from the former by adding an autoregressive component rather than a moving average component to the classical test theory model. In the second model each item score would be linked to previous item scores and only indirectly to previous error scores. To be consistent with the previous research studies (e.g., Zimmerman, Zumbo, & Lalonde, 1993; Yang & Green, 2010), the direct effect on measurement errors is modeled in my dissertation. In addition, the direct effect model is easier to be interpreted in practice than the indirect effect model of correlated errors.

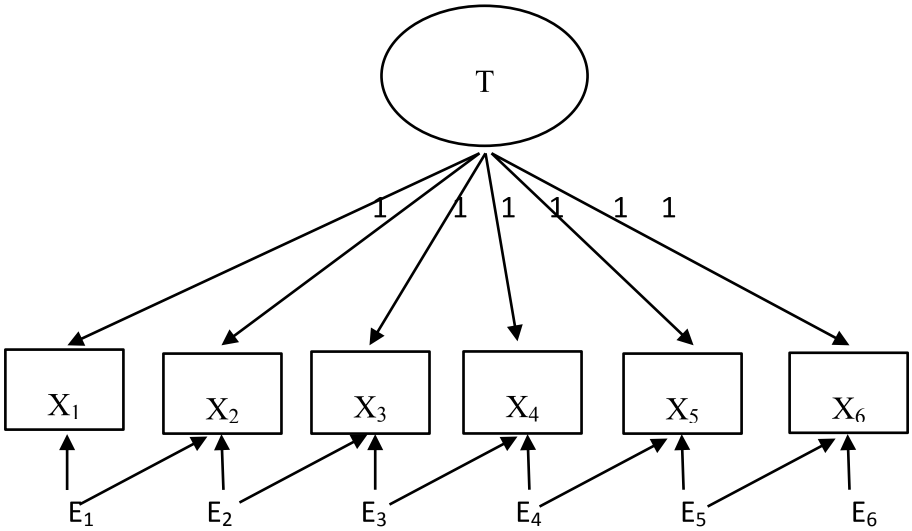

Figure 6. The Model with Moving Average Component of Order 1 in Green & Hershberger (2000)'s

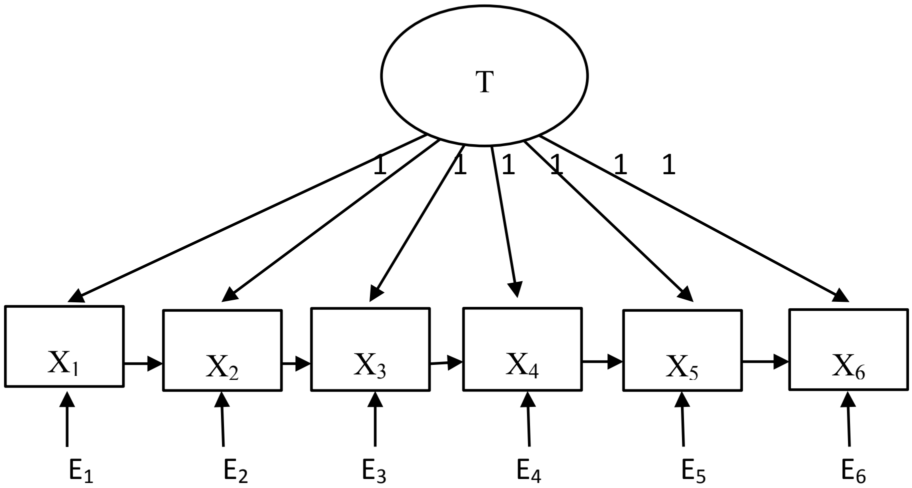

Figure 7. The Autoregressive Model in Green & Hershberger (2000)'s

2.3. Methods Used in Previous Reliability Coefficients Comparison Studies

Although many researchers have conducted studies to compare different reliability coefficients, the results are not always consistent and the whole picture of how these reliability coefficients perform in comparison is unrevealed. This inconsistency or incommensurability among reliability studies is mainly because study methods differ from one study to another and hence the indices used to evaluate the quality of reliability estimators also differ. Some studies examine reliability coefficients at the *population* level, some at the *sample* level (e.g., Callender & Osburn, 1979; Green & Hershberg, 2000; Osburn, 2000; Shapiro & Ten Berge, 2000; Ten Berge & Sočan, 2004; Zinbarg, Revelle, Yovel, & Li, 2005).

Studies at the population level examine the bias of the population values of reliability coefficients from the true reliability of the test under manipulated conditions, that is, no sample data is generated and analyzed. For example, Osburn (2000) examined up to ten reliability coefficients by varying the degree of heterogeneity of test items. All the coefficients were calculated based on the population correlation matrices and then compared to the corresponding true reliability. The bias obtained from this type of study design is consistent with the bias of a point estimator $\hat{\rho}$ defined in mathematical statistics. Statistically, the bias of a point estimator $\hat{\rho}$ is given by $\boldsymbol{B}(\hat{\rho}) = \boldsymbol{E}(\hat{\rho}) - \rho$. Here ρ represents the true reliability and $\boldsymbol{E}(\hat{\rho})$ the population reliability that is calculated using the known population covariance matrix. Studies at the population level provide the theoretical values of reliability coefficients; however, they could not provide the empirical values of reliability coefficients. Thus, studies at the sample level are necessary and they offer more implications for practice.

However, studies at the sample level vary remarkably. The first type of studies at the sample level directly compares different reliability coefficient values of various real data sets without examining the sampling distribution of reliability coefficients or the deviations from the true reliability. For example, Revelle and Zinbarg (2009) used nine data sets in total, including six examples (S1–S2c) from Sijtsma (2009), two examples (B&W 1 & 2) from Bentler and Woodward (1980), and the De Leeuw (1983) dataset to compare 13 reliability coefficients. They suggested that contrary to what the name implies, the glb was not the greatest lower bound estimate of reliability, but somewhat lower than omega. Their study showed that omega was the greatest under 7 examples out of 9 while maximized λ_4 was the greatest under the remained 2 examples, and the glb never produced the greatest estimates of reliability. However, the glb value in this study was a sample estimate of reliability. We need both true reliability and expected value of the glb to determine whether it is the greatest lower bound to reliability or not. Since the true reliability values were not provided in this study, no definite conclusion on which coefficient was the greatest lower bound could be drawn. It should also be noted that these nine examples only had the length of 4 to 8 items, the coefficients maximized lambda4 and omega might be sensitive to small item number and produced a positively biased sample estimate higher than the glb. In short, this type of comparison tells little information about the bias in reliability coefficients and the results may change remarkably as sample data changes.

The second type compares the deviation of the mean of sample reliability estimates relative to the corresponding population coefficient values and defines this deviation as the sample bias in reliability estimates. Mathematically, the bias is given by $B(\hat{\rho}) = E(\hat{\rho}) - \frac{\sum \hat{\rho}}{N}$ and $E(\hat{\rho})$ is obtainable from the population covariance matrix. In this type of study, either the covariance matrix of a real dataset or an artificial covariance matrix is used as the population covariance matrix. Then random samples with varied sample sizes are drawn from the population covariance matrix. For example, Ten Berge and Sočan (2004) generated 500 samples of size 100, 250, 500, and 1000 out of De Leeuw (1983)'s the correlation matrix of a real data with 119 subjects and 6 items. The authors compared the sample bias in coefficient alpha, the maximized lambda4 and the glb in terms of the sample coefficient values relative to the corresponding coefficient values of the real dataset. The results showed that 1) all the sample bias values were low (ranged from -0.004 to 0.008); 2) alpha had the least sample bias (in absolute value), the glb had the largest sample bias, and the sample bias in the maximized lambda4 was in the middle. Shapiro and Ten Berge (2000) generated 500 samples of sample size respectively 100, 500, 2000 from the multivariate normal $N\ (0,\ \mathbf{I}_{\mathrm{p}})$ distribution (p is the number of test items) and postmultiply each vector by $\sum^{1/2}$ ($\sum$ was a given population covariance matrix) to yield data following the multivariate normal $N\ (0,\ \sum)$ distribution. The given population covariance matrix was respectively of 5 and 10 items from a scholastic achievement. The results showed that the sampling bias in the glb increased as the number of test items changed from 5 to 10 and decreased as the sample size increased from 100 to 2000. The sample bias was found to range from 1.8% to 13.5% higher than the population values, and the largest bias occurred under the condition of 10 items and 100 cases. The sample bias defined in this type of study tells about the deviation of sample estimates of reliability from expected reliability estimate. Statistically, this deviation is measured by $\mathrm{E}[(\hat{\rho} - \mathrm{E}(\hat{\rho}))^2]$, that is, the variance of the sample estimates. However, it tells little about the deviation of sample estimates of reliability with regard to the true reliability.

The third type of studies evaluates reliability coefficients by examining the degree of the bias of the mean of sample estimates relative to the true reliability of a test. For example, Yang and Green (2010) systematically studied the relative bias [calculated by $(\bar{\hat{\rho}} - \rho)/\rho]$ of sample SEM estimates of reliability in comparison to coefficient alpha by a Monte Carlo study under the tau-equivalent, congeneric, bifactor and correlated error models. They found the SEM approach showed minimal (relative) bias when the model was correctly specified and items were relatively well defined by their underlying factor(s). When the model was misspecified, particularly underspecified, greater bias occurs. Although they proposed that the SEM estimates may be unstable and biased with small sample sizes, their study results showed that when the model is correctly specified, SEM estimates of reliability demonstrated less bias than coefficient alpha for small sample sizes (i.e. 50 observations) and small number of items (i.e., 6 items) even under the tau-equivalent model. In addition, they reported the average deviations of the sample estimate from the estimated expected reliability and true reliability as the indices for stability of SEM reliability estimates. The formulas for the two types of deviations are respectively $\sqrt{\frac{(\hat{\rho}-\bar{\hat{\rho}})^2}{N}}$ and $\sqrt{\frac{(\hat{\rho}-\rho)^2}{N}}$.

From the review of the methods in previous reliability studies, we can see that the method in the third type of studies at the sample level is most rigorous in terms of statistical standards for evaluating the quality of reliability coefficients. Therefore, this study adopted the method in the third type of studies at the sample level and closely followed the method section in Yang and Green (2010)'s study. However, some design factors were changed in this study. The details are presented in Chapter 3 section 3.1 and rationales for major changes are also provided in that section.

2.4. Study Purpose and Research Questions

To date, there is no study that has compared the performances of the most often used or recommended reliability coefficients from each of the five approaches. Therefore, the major purpose of the current study is to evaluate the performances of the reliability coefficients from the five approaches under simulated conditions. The selected reliability coefficients are coefficient alpha, the most often used reliability coefficient (the internal consistency approach); the recommended reliability coefficients: the glb (the lower bound approach), recommended by Green and Yang (2009), Sijtsma (2009), and Ten Berge and Sočan (2004), theta and generalized theta (PCA approach), recommended by Şimşek and Noyan (2013), omega (EFA approach), recommended by Revelle and Zinbarg (2009), and SEM estimates of reliability (SEM approach) under different analysis models (see details in Chapter 3), recommended by Green and Yang (2009), and Yang and Green (2011). Therefore, the first and main research question addressed in this simulation study is *(1) Which reliability coefficient generates the best estimate of reliability considering the manipulated conditions?* To answer this research question, a Monte Carlo study was conducted to compare the accuracy and stability of these reliability coefficients.

Yang and Green's (2010) study has shown that the correctly specified SEM estimates of reliability are very promising alternatives to coefficient alpha. However, the Monte Carlo simulation study on SEM estimates of reliability has not been replicated on various multidimensional measurement models and the performance quality of SEM estimates of reliability has not fully exposed. For example, we do not know how SEM estimates of reliability will perform when the assumption of unidimensionality is violated in different degrees. Moreover, multidimensional data is hard to be correctly specified in practice without prior knowledge of a measure or test; even with prior knowledge or theoretical rationale, it is not easy to correctly represent the internal structure of a measure. In addition, a complicated multidimensional model often incurs estimation problems (see

Yang & Green, 2010). Thus, the under-specified models seem to be more advantageous since model specifications are comparably simple and fewer parameters are required to be estimated. Therefore, the second research question is raised with a focus on the performances of the under-specified SEM estimates of reliability under the violation of unidimensionality in varied degree: *(2) When the assumption of unidimensionality is violated, will unidimensional SEM estimates of reliability approximate the correctly specified SEM estimates of reliability?*

Graham's (2006) study indicated that when the congeneric measurement model was used to analyze the data generated from a tau-equivalent model, the estimated reliability was almost as accurate as the one estimated by the correct model (tau-equivalent model). The congeneric model can provide approximately accurate estimates of the parameter loadings as the tau-equivalent model, given that there is no estimation problem, hence providing an approximately accurate estimate of reliability. Yang and Green's (2010) study also found when the model was underspecified, greater bias occurs, whereas the over-specified models generated less bias. Therefore, I hypothesize that a general model can produce robust estimates of reliability when it is used to analyze data generated from the less specified models. The corresponding research question is: *(3) If the analysis model is an over-specified model, will its estimates of reliability approximate those using the correctly specified analysis model?*

However, the disadvantage of using a general model is that more parameters need to be estimated, which usually causes estimation problems (e.g., instability and convergence failure). Although ordinary least squares (OLS) and maximum likelihood (ML) are the most frequently used estimation methods in SEM, they tend to cause sampling bias and experience estimation problems for data with small sample sizes. To avoid these problems in parameter estimation when sample sizes are small, Bayesian estimation methods can be used for conditions where estimation problems occur. Bayesian estimation methods have been found to work well with small sample sizes, and usually do not produce inadmissible parameter estimates such as negative variances (Lee, 2007; Hox, Van de Schoot & Matthijsse, 2012). However, Bayesian analysis requires complex statistical specifications and that hinders its application. Therefore, in this study, Mplus 7 was used due to its simple analysis specifications with convenient defaults (i.e., diffuse or non-informative priors). The corresponding and final research question is *(4) Can Bayesian estimation with non-informative priors overcome estimation problems in SEM estimates of reliability using ML? If yes, will the estimates of reliability using Bayesian estimation with non-informative priors be more accurate and stable than those using ML?*

CHAPTER 3 METHODS

Chapter 3 presents the study design, specifically the manipulated design factors, the methods and programs for data generation and analysis, and the evaluation criteria for assessing the quality of the reliability coefficients selected in this comparison study.

3.1. Study Design

There were two major stages in this study. In stage 1, the selected reliability coefficients were evaluated across all the simulated conditions introduced below. First, three types of measurement models, including the (essentially) tau-equivalent model, the congeneric model and the correlated factor model, were chosen for data simulation. The (essentially) tau-equivalent model was selected to serve the purpose of comparison since coefficient alpha can correctly estimate reliability under the (essentially) tau-equivalent model. The correlated factor model instead of the bifactor model in Yang and Green's study was used because (1) it is the basic model in the family of multidimensional models, of which the bifactor model and the higher order factor model can be mathematically derived (Yung, Thissen, & McLeod, 1999), (2) it is easier to manipulate because it has less parameters that need to be estimated compared to the corresponding bifactor model and higher order factor model, and (3) it is a more interpretable model than the corresponding bifactor model when manipulating test heterogeneity.

In addition to measurement models, other factors that have been known to affect reliability estimation, including sample size, item number, factor correlation, and error correlation, were manipulated as follows (a) three levels of sample size: 50, 150, and 500, (b) two levels of number of items: 6 and 12, (c) two levels of factor correlation: 0.2 and 0.6 for the correlated factor model, and (d) three levels of measurement error correlation: 0, 0.2 and 0.5.

Conditions with a large number of items and large sample sizes were not considered because bias in most reliability coefficients decrease systematically as item number increases and stability of reliability coefficients increases as sample size increases (Tang & Cui, 2012, 2014). Therefore, differences in various reliability estimates became smaller and less noticeable as item number and sample size increased. Give that large numbers of sample sizes and item numbers had already been considered in previous simulation studies (Tang & Cui, 2012, 2014), this study focused on smaller sample sizes and item numbers where the selection of an appropriate reliability estimate matters more.

The factor correlation was manipulated no more than 0.6 since a high factor correlation indicated the existence of unidimensionality. Constructing highly correlated factor models was redundant in a simulation study where unidimentional models were already included.

The models with correlated errors were also considered because the assumption of independent errors is hard to hold when items on a test are administered on a single occasion. Rozeboom (1966) has argued that when items on a test are administered on a single occasion, errors among items are likely to be positively correlated and that these correlated errors yield spuriously high coefficient alphas. Correlated errors under this situation should be treated purely as measurement errors and considered as an unreliable component of measures. The size of error correlation was manipulated to be no more than 0.5, following Yang and Green (2010)'s study. The underlying reason is a high correlation of measurement errors in a measure indicates the measure has very poor quality, which is not representative of most tests.

Table 1.

Model Design

Model	Model Type	ϕ_{12}	λ_{i1}	λ_{i2}	$\rho_{\varepsilon\varepsilon'}$
1	Tau-equivalent Model	0	All .6	0	0
2		0	All .6	0	0.2
3		0	All .6	0	0.5
4	Congeneric Model	0	.4, .5, .6, .6, .7, .8*	0	0
5		0	.4, .5, .6, .6, .7, .8*	0	0.2
6		0	.4, .5, .6, .6, .7, .8*	0	0.5
7	Correlated Model	0.2	.5, .6, .7, .0, .0, .0*	.0, .0, .0, .5, .6, .7*	0
8		0.2	.5, .6, .7, .0, .0, .0*	.0, .0, .0, .5, .6, .7*	0.2
9		0.2	.5, .6, .7, .0, .0, .0*	.0, .0, .0, .5, .6, .7*	0.5
10		0.6	.5, .6, .7, .0, .0, .0*	.0, .0, .0, .5, .6, .7*	0
11		0.6	.5, .6, .7, .0, .0, .0*	.0, .0, .0, .5, .6, .7*	0.2
12		0.6	.5, .6, .7, .0, .0, .0*	.0, .0, .0, .5, .6, .7*	0.5

Note 1. $\rho_{\varepsilon\varepsilon'}$ refers to the error correlation; λ_{i1} and λ_{i2} refer to the factor loadings on factor 1 and factor 2 respectively; ϕ_{12} refers to the correlation between the two factors.

Note 2. * represents λ_{i1} of six items in a measure. λ_{i1} is replicated for a measure with 12 items.

After combining the design factors: types of measurement models, factor correlations, and error correlations, there were in total 12 specific measurement models (displayed in Table 1) for data generation. As shown in Table 1, the factor loadings were all 0.6 in the (essentially) tau-equivalent model, and they ranged from 0.4 to 0.8 in the congeneric model. In the correlated factor models, two factors with simple structure were generated with factor loadings (varying from 0.5 to 0.7) loading on distinctive items. The average of the factor loadings was controlled at 0.6, the medium high level, in these models. In total, two levels of number of items, three levels of number of subjects, and twelve measurement models were considered in the simulation study so as to produce a total of 2 x 3 x 12= 72 simulation conditions. Each simulation condition was replicated 1,000 times.

In stage 1, ML was used in all the SEM estimates of reliability; in stage 2, only the conditions with serious estimation issues were further examined using Bayesian estimation with non-informative priors. There are two reasons for using non-informative priors. First, it should not be assumed that researchers always have knowledge of the distributions of the parameters from previous research. Second, the convenience of using Bayesian estimation matters in application. Mplus 7 provides convenient defaults of non-informative priors. Specifically, these non-informative priors are the normal distributions with mean of 0 and infinite variance for free parameters like loadings and intercepts, and inverse Gamma distributions for free variance parameters. For a more detailed explanation of choosing priors for Bayesian SEM, see Dunson, Palomo, and

Bollen (2005). To estimate posteriors, the Markov Chain Monte Carlo technique (MCMC), or more specifically, Gibbs sampling, was used in Mplus 7.

3.2. Data Generation and Analysis

Responses to all the items were assumed to be multivariate normally distributed. For simplicity, all population item responses were assumed to have the mean of 0 and variance of 1, and thus the item covariance matrix was equal to correlation matrix. The population covariance matrix was generated for each specific model by the following equation,

$$\sum = \boldsymbol{\Lambda}_y \boldsymbol{\Psi} \boldsymbol{\Lambda}_y' + \boldsymbol{\Theta}_\varepsilon, \tag{48}$$

where $\sum$ is the population covariance matrix, $\boldsymbol{\Lambda}_y$ is the factor loadings matrix, $\boldsymbol{\Psi}$ is the correlation matrix of common factors, and $\boldsymbol{\Theta}_\varepsilon$ is the error variance-covariance matrix. The true reliability was calculated with each defined population covariance matrix.

The sample observed score matrix was then generated by the R (R Development Core Team, 2008) package "mvtnorm" (Genz, A., Bretz, F., Miwa, T., Mi, X, Leisch, F., Scheipl, F., & Hothorn, T., 2014) to obtain multivariate normal item data using the defined population covariance matrix. After sample observed score matrices were generated, sample reliability estimates were computed. Coefficient alpha, theta, and generalized theta were calculated by their corresponding formulas. The glb and omega were obtained using the R package "psych" (Revelle, 2013).

For obtaining the SEM estimates of reliability, SEM software, Mplus 7 (Muthén &

Muthén, 1998-2014), was used for estimating parameters of structural equation models and SEM reliability estimates were calculated by the general formula: equation (47) reviewed in Chapter Two. Each generated data set was fit to the following analysis models in Mplus 7. For unidimensional models, three analysis models were specified for running SEM: (1) the tau-equivalent model (i.e., equality constraints on the factor loadings), (2) the congeneric model (i.e., without equality constraints on factor loadings), (3) the unifactor congeneric model with correlated errors. For multidimensional models, five analysis models were specified for running SEM: (1) the tau-equivalent model (i.e., equality constraints on the factor loadings), (2) the congeneric model (i.e., without equality constraints on factor loadings), (3) the unifactor congeneric model with correlated errors, (4) the correlated factor model (with constraints of zero loadings on items to obtain the simple structure), and (5) the correlated factor model with correlated errors. The obtained SEM estimates of reliability from the above mentioned analysis models were named respectively as SEM.tau, SEM.cong, SEM. CE, SEM.CF and SEM.CFCE in the first stage of data analysis using the ML estimation method. In the second stage, the Bayesian estimation method was used and the BSEM estimates of reliability were named in a similar way as the corresponding SEM estimates of reliability. For example, SEM.CFCE was named as BSEM.CFCE when the Bayesian estimation method was selected.

Each simulation condition (72 conditions in total) was replicated 1000 times. The percentage relative bias, standard error and root mean square error of each reliability coefficient were calculated for each condition in R environment. The rounded percentages of estimation failure numbers for each SEM reliability coefficient under all the manipulated conditions were also obtained. This information was used to select the conditions with serious estimation issues for the second stage analysis. Finally, the percentage relative bias, standard error and root mean square error of Bayesian SEM estimates of reliability for the selected conditions were calculated.

3.3. Evaluation Criteria

The primary evaluation criterion used in this study is the accuracy of reliability estimation. To measure the accuracy of reliability estimation, the relative bias was calculated by the formula $(\overline{\hat{\rho}} - \rho)/\rho$, where $\overline{\hat{\rho}}$ is the mean of reliability estimates across 1,000 replications under each condition.

In addition to a smaller bias, a preference exists for an estimator that has a distribution with a smaller variance ensuring in repeated sampling a higher portion of values of $\hat{\rho}$ will be closer to ρ. Given two unbiased estimators of a parameter ρ and all other things being equal, we would select the estimator with the smaller variance. Thus, the secondary evaluation criterion is the stability or precision of reliability estimation. To measure the stability of reliability estimation, the standard error, $\sqrt{\frac{\Sigma(\hat{\rho}-\overline{\hat{\rho}})^2}{N}}$, of the sampling distribution of $\hat{\rho}$ was calculated under each condition. Smaller standard errors indicate more stable estimates across samples.

Finally, the root mean square error, $\sqrt{\frac{\sum(\hat{\rho}-\rho)^2}{N}}$, was considered because it takes account of both bias and precision of the estimator distribution (Wackerly, Mendenhall, & Scheaffer, 2008). If a reliability coefficient has inconsistent values in its relative bias and standard error (i.e., neither high nor low simultaneously), this criterion can be referred to given that it is the function of both bias and standard error.

CHAPTER 4 RESULTS

4.1. Results in Stage One

The indices for measuring the accuracy and precision of the reliability coefficients are summarized in Tables 2 to 13. Each table (from Table 2 to 13) presents the percentage relative biases (hereafter referred to as biases), standard errors (SEs) and root mean square errors (RMSEs) of the reliability coefficients under the conditions that item number is respectively 6 and 12 and sample size is respectively 50, 150 and 500 for each of the twelve specified measurement models described in Table 1 of Chapter 3. For SEM reliability coefficients, the rounded percentages of the numbers of estimation failures are presented in Tables 14 to 18. Each table (from Table 14 to 18) presents the rounded percentages of the estimation failure numbers for each of the five types of SEM reliability coefficient examined in this study (i.e., SEM.tau, SEM.cong, SEM. CE, SEM.CF and SEM.CFCE), under the conditions that item number is 6 and 12, and sample size is 50, 150 and 500.

4.1.1 Reliability Estimates for Tau-Equivalent Models

The relative percentage biases, SEs and RMSEs of all the reliability coefficients for Model 1 (the tau-equivalent model with independent errors), Model 2 (the tau-equivalent model with error correlation at 0.2) and Model 3 (the tau-equivalent model with error correlation at 0.5) are respectively presented in Tables 2, 3, and 4. The results are summarized in the order of bias, SE and RMSE under each measurement model with respect to the conditions of 6 items and 12 items.

6-Item Scale Model 1

As shown in Table 2 (Model 1), alpha, theta, generalized theta (hereafter shorten as theta.g), SEM.tau (SEM estimate of reliability using the correctly specified analysis model, hereafter shortened as CSM), SEM.cong (SEM estimate of reliability using the over-specified analysis model, hereafter shortened as OSM) and SEM.CE (OSM) had quite accurate reliability estimates. They all had their biases below or around 1% under all the conditions of 6 items in Model 1. However, the biases of the glb and omega were considerably high (respectively 8.03% and 9.44%) under the condition that sample size was 50. As sample size increased, the biases in the glb and omega all decreased and the glb had higher rate of decreasing than omega. When sample size was 500, the biases of the glb and omega were respectively 2.96% and 6.68%, which were still higher than the remaining coefficients.

The SEs of these reliability coefficients were all small, ranging from 0.01 to 0.06. SEM.tau (CSM) had the largest SE (0.06) and the glb and omega had the smallest SE (0.04) when sample size was 50. As sample size increased, the SEs of these reliability coefficients all decreased, and SEM.tau had the highest rate of decreasing. When sample size was 500, SEM.tau, SEM.cong (OSM) and SEM.CE (OSM) had the smallest SE

(0.01), and the remaining coefficients had a SE of 0.02.

The RMSEs of these reliability coefficients were all considerably small, ranging from 0.01 to 0.08. Omega had the largest RMSE and the glb had the second largest RMSE regardless of sample size. Alpha, theta, theta.g, SEM.cong (OSM) and SEM.CE (OSM) had the smallest RMSE (0.05) when sample size was 50. When sample size increased to 150, except the glb and omega, all the coefficients had the smallest RMSE (0.03). When sample size continued to increase to 500, SEM.tau (CSM), SEM.cong and SEM.CE had the smallest RMSE (0.01).

12-Item Scale Model 1

The biases, SEs and RMSEs of these reliability coefficients all decreased as item number increased from 6 to 12. When item number increased from 6 to 12, omega's bias dramatically decreased, whereas the glb's bias only dropped unremarkably. On the other hand, simply increasing sample size significantly decreased the glb's bias while it only slightly decreased omega's bias. Under the conditions of 12 items, the glb had the largest bias and omega had the second largest bias regardless of sample size. The remaining coefficients had similarly small biases, the magnitude of which were all below or around 1%.

The SEs of these reliability coefficients were all small, ranging from 0.01 to 0.03. As sample size increased, the SEs of these reliability coefficients all decreased. However, if the SEs were at the level of 0.01, the decreasing rate were approaching zero.

The RMSEs of these reliability coefficients ranged from 0.01 to 0.07. The glb had the largest RMSEs regardless of sample size. The RMSEs of the remaining coefficients were almost equally low, especially when sample sizes were large.

Table 2

Percentage Bias, SE and RMSE of Reliability Estimates of the Data Generated from the Tau-equivalent Model with Independent Errors (Model 1)

6-Item scale	n=50			n=150			n=500		
EC=0; ρ=0.77	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
Alpha	-1.06	0.05	0.05	-0.36	0.03	0.03	0.12	0.02	0.02
Glb	8.03	0.04	0.07	4.61	0.03	0.05	2.96	0.02	0.03
Theta	-0.35	0.05	0.05	-0.13	0.03	0.03	0.19	0.02	0.02
Theta.g	-0.35	0.05	0.05	-0.13	0.03	0.03	0.19	0.02	0.02
Omega	9.44	0.04	0.08	7.34	0.03	0.06	6.68	0.02	0.06
SEM.tau	-0.65	0.06	0.06	-0.39	0.03	0.03	0.13	0.01	0.01
SEM.cong	-0.31	0.05	0.05	-0.95	0.03	0.03	0.18	0.01	0.01
SEM.CE	-0.71	0.05	0.05	-0.29	0.03	0.03	1.04	0.01	0.01
12-Item scale	n=50			n=150			n=500		
EC=0; ρ=0.87	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE

Alpha	-1.06	0.03	0.03	-0.28	0.01	0.01	0.06	0.01	0.01
Glb	7.51	0.02	0.07	4.60	0.01	0.04	2.69	0.01	0.03
Theta	-0.32	0.03	0.03	-0.24	0.02	0.02	0.01	0.01	0.01
Theta.g	-0.32	0.03	0.03	-0.24	0.02	0.02	0.01	0.01	0.01
Omega	2.59	0.03	0.04	1.71	0.01	0.02	1.40	0.01	0.02
SEM.tau	-0.75	0.03	0.03	-0.17	0.01	0.01	0.09	0.01	0.01
SEM.cong	-0.36	0.03	0.03	-0.07	0.01	0.01	0.13	0.01	0.01
SEM.CE	-0.13	0.03	0.03	-0.07	0.02	0.02	0.13	0.01	0.01

Note 1. The bias values are in the form of percentage bias.

Note 2. Theta.g represents generalized theta

6-Item Scale Model 2

Under the conditions of 6 items in Model 2 (Table 3), all the sample estimates of the reliability coefficients had positive bias except SEM.CE (CSM). It had negative bias when sample size was small (e.g., 50), and positive bias when sample size was enlarged. Among all the coefficients compared in this study, SEM.CE (CSM) stood out for its accurate reliability estimates with its bias below 1% under the conditions of 6 items in Model 2. However, the glb, omega had serious biases, ranging from 19.61% to 22.21%. Again, omega had the largest bias under the conditions of 6 items. The remaining reliability coefficients: alpha, theta, theta.g, SEM.tau (SEM estimate of reliability using the under-specified analysis model, hereafter shortened as USM) and SEM.cong (USM) had similar biases ranging from 9.94% to 11.75%.

The SEs of these reliability coefficients were all small, ranging from 0.01 to 0.05. SEM.CE (CSM) had the largest SE (0.05) and SEM.tau (USM) had the smallest SE (0.02) when sample size was 50. As sample size increased, the SEs of these reliability coefficients all decreased. When sample size was 500, SEM.CE had a SE of 0.02, and the remaining coefficients had a SE of 0.01.

Compared with the corresponding conditions in Model 1 (Table 2), the RMSEs of these reliability coefficients were larger as there were correlated errors in Model 2 (Table 3). SEM.CE (CSM) had the smallest RMSEs under all the conditions (ranging from 0.01 to 0.05). The glb and omega had similar larger RMSE values, ranging from 0.14 to 0.16. The remaining reliability coefficients had their RMSEs ranging from 0.07 to 0.09.

12-Item Scale Model 2

The biases, SEs and RMSEs of these reliability coefficients all decreased as item number increased from 6 to 12. As similar as in Model 1, omega’s bias dramatically decreased, whereas the glb’s bias only dropped unremarkably when item number increased from 6 to 12. Under the conditions of 12 items, the glb had the largest bias regardless of sample size. SEM.CE (CSM) had the smallest biases, the magnitude of which were all below or around 1%. The remaining coefficients had similar biases ranging from 5.17% to 6.18%.

The SEs of these reliability coefficients were all small, ranging from 0.01 to 0.04. When sample size was 50, SEM.CE (CSM) had the largest SE (0.04) and the glb had the smallest SE (0.02). As sample size increased, the SEs of these reliability coefficients all decreased until they were at the level of 0.01.

The RMSEs of these reliability coefficients ranged from 0.01 to 0.12. The glb had larger RMSE values (respectively 0.12, 0.10 and 0.09) than omega (respectively 0.08, 0.07 and 0.07). SEM.CE (CSM) had the smallest RMSEs (respectively 0.04, 0.02 and 0.01). The remaining reliability coefficients (i.e., alpha, theta, theta.g, SEM.tau [USM], and SEM.cong [USM]) all had their RMSEs at 0.05 regardless of sample size.

Table 3

Percentage Bias, SE and RMSE of Reliability Estimates of the Data Generated from the Tau-equivalent Model with Error Correlation at 0.2 (Model 2)

6-Item scale	n=50			n=150			n=500		
EC=0.2; ρ=0.72	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
Alpha	10.40	0.04	0.08	11.28	0.02	0.08	11.67	0.01	0.08
Glb	21.53	0.03	0.16	19.83	0.02	0.14	19.61	0.01	0.14
Theta	11.08	0.04	0.09	11.51	0.02	0.09	11.75	0.01	0.09
Theta.g	11.32	0.04	0.09	11.51	0.02	0.09	11.75	0.01	0.09

Omega	22.21	0.03	0.16	21.21	0.02	0.15	20.90	0.01	0.15
SEM.tau	9.94	0.02	0.07	10.85	0.01	0.08	11.21	0.01	0.08
SEM.cong	10.67	0.03	0.08	11.10	0.02	0.08	11.31	0.01	0.08
SEM.CE	-0.96	0.05	0.05	-0.08	0.03	0.03	0.24	0.02	0.02
12-Item scale		n=50			n=150			n=500	
EC=0.2; ρ=0.83	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
Alpha	5.17	0.03	0.05	5.78	0.01	0.05	5.89	0.01	0.05
Glb	13.93	0.02	0.12	12.14	0.01	0.10	11.24	0.01	0.09
Theta	5.40	0.03	0.05	5.93	0.02	0.05	5.92	0.01	0.05
Theta.g	5.40	0.03	0.05	5.93	0.02	0.05	5.92	0.01	0.05
Omega	8.57	0.03	0.08	8.31	0.01	0.07	7.84	0.01	0.07
SEM.tau	5.46	0.03	0.05	5.92	0.01	0.05	6.07	0.01	0.05
SEM.cong	5.83	0.03	0.05	5.95	0.01	0.05	6.18	0.01	0.05
SEM.CE	-0.19	0.04	0.04	0.05	0.02	0.02	0.39	0.01	0.01

6-Item Scale Model 3

As the error correlation rose to 0.5 in Model 3 (Table 4), all sample estimates of the reliability coefficients had larger positive bias except SEM.CE (CSM). Still, SEM.CE (CSM) had the smallest bias in magnitude. As sample size increased, the magnitude of its bias decreased trivially. The glb was the least accurate reliability coefficient in Model

3, the bias of which ranged from 49.03% to 49.34% under the conditions of 6 items. Although omega had slightly less bias than the glb under all the conditions, it was less accurate than the remaining reliability coefficients (i.e., alpha, theta, theta.g, SEM.tau [USM], and SEM.cong [USM]). The remaining reliability coefficients had their biases ranging from 26.92% to 29.63% under the conditions of 6 items.

Except SEM.CE (CSM), all the reliability coefficients had small SEs ranging from 0.01 to 0.06. SEM.CE (CSM) had the largest SE (0.11) and the glb had the smallest SE (0.01) when sample size was 50. As sample size increased, the SEs of these reliability coefficients all decreased. When sample size was 500, SEM.CE had a SE of 0.03, and the remaining coefficients had their SEs at 0.01 or below.

The RMSEs of these reliability coefficients continued to increase as error correlation increased to 0.5 in Model 3 (Table 4). SEM.CE (CSM) had the smallest RMSEs under all the conditions (ranging from 0.03 to 0.11), and the glb had the largest RMSE (0.32) irrespective of sample size. Compared with the glb, omega had a slightly smaller RMSE value (0.29) under all the sample size conditions. The remaining reliability coefficients had similar RMSEs ranging from 0.18 to 0.20.

12-Item Scale Model 3

The biases, SEs and RMSEs of these reliability coefficients all decreased as item number increased from 6 to 12 except that the bias of SEM.CE slightly increased in magnitude under the condition of 500 observations. When item number was 12, SEM.CE (CSM) had the smallest bias (below 1%), and the glb had the largest bias (around 27%). Compared with the glb, omega had quite smaller bias values (around 18%) under all the sample size conditions. The remaining reliability coefficients had similar biases near 14% to 15%.

The SEs of these reliability coefficients were all small, ranging from 0.01 to 0.05. When sample size was 50, SEM.CE (CSM) had the largest SE (0.05) and the glb had the smallest SE (0.00). When sample size was 500, the SE of SEM.CE decreased to 0.01, which was the same as the other reliability coefficients except the glb. The glb had the smallest SE (0.00) under all the conditions.

The RMSEs of these reliability coefficients ranged from 0.01 to 0.21. The glb had largest RMSEs (0.21), and SEM.CE (CSM) had the smallest RMSEs (respectively 0.05, 0.03 and 0.01). Omega had its RMSE in the range of 0.14 to 0.15, and the remaining reliability coefficients had their RMSEs in the range of 0.11 to 0.12 under the conditions of 12 items.

Table 4

Percentage Bias, SE and RMSE of Reliability Estimates of the Data Generated from the Tau-equivalent Model with Error Correlation at 0.5 (Model 3)

6-Item scale	n=50	n=150	n=500

EC=0.5; ρ=0.65	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
Alpha	26.92	0.04	0.18	29.37	0.02	0.19	29.22	0.01	0.19
Glb	49.03	0.01	0.32	49.34	0.01	0.32	49.23	0.00	0.32
Theta	27.71	0.04	0.18	29.63	0.02	0.19	29.40	0.01	0.19
Theta.g	29.35	0.06	0.20	29.63	0.02	0.19	29.40	0.01	0.19
Omega	44.25	0.02	0.29	44.37	0.01	0.29	44.20	0.00	0.29
SEM.tau	27.95	0.02	0.18	28.75	0.01	0.19	29.09	0.01	0.19
SEM.cong	27.80	0.05	0.19	28.62	0.02	0.19	29.05	0.01	0.19
SEM.CE	-2.11	0.11	0.11	-0.57	0.08	0.08	-0.18	0.03	0.03

12-Item scale		n=50			n=150			n=500	
EC=0.5; ρ=0.78	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
Alpha	14.31	0.02	0.11	14.62	0.01	0.11	14.86	0.01	0.12
Glb	27.28	0.00	0.21	27.18	0.00	0.21	27.15	0.00	0.21
Theta	14.12	0.02	0.11	14.64	0.02	0.12	14.91	0.01	0.12
Theta.g	14.12	0.02	0.11	14.64	0.02	0.12	14.91	0.01	0.12
Omega	18.88	0.02	0.15	18.68	0.01	0.15	18.51	0.01	0.14
SEM.tau	14.58	0.02	0.12	14.71	0.01	0.12	14.90	0.01	0.12
SEM.cong	14.92	0.02	0.12	14.83	0.01	0.12	14.94	0.01	0.12
SEM.CE	0.15	0.05	0.05	0.04	0.03	0.03	0.29	0.01	0.01

In summary, the reliability coefficient SEM.CE had the smallest bias and RMSE under almost all the conditions in the tau-equivalent models (Models 1 to 3). Under the conditions in Model 1 (the tau-equivalent model with independent errors), alpha, theta, theta.g, SEM.tau, and SEM.cong also had the smallest or nearly the smallest bias and RMSE. However, as error correlation occurred in Model 2 (the tau-equivalent model with error correlation at 0.2) and increased in Model 3 (the tau-equivalent model with error correlation at 0.5), their biases and RMSEs were much larger than SEM.CE. The glb had the largest bias and RMSE values under the conditions of 12 items in Models 1 and 2 and under all the conditions in Model 3, while omega had the largest bias and RMSE values under the remaining conditions. The two coefficients (i.e., the glb and omega) consistently had larger biases and RMSEs than the other coefficients.

All the reliability coefficients, except SEM.CE, had very small SEs ranging from 0.00 to 0.06 under all the conditions. In fact, SEM.CE had similarly small SEs as other coefficients under most conditions. It only had slightly larger SE values than other coefficients when there were correlated errors as in Models 2 and 3 and when sample size was 50 and item number were 6, the discrepancies were very trivial unless under the condition when error correlation was 0.5. Under that condition, SEM.CE had its largest SE of 0.11. When sample size was 50 and item number was 12, the glb consistently had the smallest SE values across Models 1 to 3. In short, the reliability coefficients had similarly small SEs and their SEs decreased as sample size increased from 50 to 150 or more.

4.1.2 Reliability Estimates for Congeneric Models

The relative percentage biases, SEs and RMSEs obtained for Model 4 (the congeneric model with independent errors), Model 5 (the congeneric model with error correlation at 0.2) and Model 6 (the congeneric model with error correlation at 0.5) are presented in Tables 5, 6 and 7. The results are summarized in the order of bias, SE and RMSE under each measurement model with respect to the conditions of 6 items and 12 items.

6-Item Scale Model 4

Under the conditions of 6 items in Model 4 (Table 5), the most accurate reliability coefficients were theta, theta.g and SEM.cong (CSM). The three coefficients all had their percentage bias below 0.5% under all the conditions of 6 items. Coefficient alpha, SEM.tau (USM) and SEM.CE (OSM) had slightly larger bias than theta, theta.g and SEM.cong. However, the discrepancies among these coefficients were not remarkable, especially when the sample size was large. Similar as their performances in Model 1, the biases of the glb and omega were still higher than those of the other coefficients (i.e., alpha, theta, theta.g, SEM.tau, SEM.cong and SEM.CE), and omega had the largest bias under the conditions of 6 items.

The SEs of these reliability coefficients were all small, ranging from 0.01 to 0.06.

alpha had the largest SE (0.06) and the glb and omega had the smallest SE (0.04) when sample size was 50. As sample size increased, the SEs of these reliability coefficients all decreased. When sample size was 500, alpha, SEM.tau (UCM), SEM.cong (CSM) and SEM.CE (OCM) had the smallest SE (0.01), and the remaining coefficients had a SE of 0.02.

The RMSEs of these reliability coefficients were all considerably small, ranging from 0.01 to 0.08. Omega had the largest RMSE and the glb had the second largest RMSE irrespective of sample size. Theta, theta.g, SEM.Tau (UCM), SEM.cong (CSM) and SEM.CE (OCM) had the smallest RMSE (0.05) and alpha had a RMSE of 0.06 when sample size was 50. Their RMSEs all decreased as sample size increased. When sample size increased to 500, alpha, SEM.tau, SEM.cong and SEM.CE had the smallest RMSE (0.01), while theta and theta.g had a slightly larger RMSE (0.02).

12-Item Scale Model 4

The biases, SEs and RMSEs of these reliability coefficients all decreased as item number increased from 6 to 12. When item number increased from 6 to 12, omega's bias remarkably decreased, whereas the glb's bias only changed trivially. On the other hand, simply increasing sample size significantly decreased the glb's bias while it only slightly decreased omega's bias. Under the conditions of 12 items in Model 4 (Table 5), the glb had the largest bias and omega had the second largest bias regardless of sample size. The remaining coefficients had similarly small biases, the magnitude of which were all below or around 1%.

The SEs of these reliability coefficients were all small, ranging from 0.01 to 0.03. As sample size increased, the SEs of these reliability coefficients all decreased. When sample size was 500, all the coefficients had a SE of 0.01.

The RMSEs of these reliability coefficients were also small in Model 4 (Table 5), ranging from 0.01 to 0.07. The glb had the largest RMSE values (respectively 0.07, 0.04 and 0.02). The RMSEs of the remaining coefficients were almost equally low, especially when item number was large.

Table 5

Percentage Bias, SE and RMSE of Reliability Estimates of the Data Generated from the Congeneric Model with Independent Errors (Model 4)

6-Item scale	n=50			n=150			n=500		
EC=0; ρ=0.78	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
Alpha	-2.68	0.06	0.06	-2.27	0.04	0.04	-0.90	0.01	0.01
Glb	7.79	0.04	0.07	4.53	0.03	0.05	2.83	0.02	0.03
Theta	-0.27	0.05	0.05	-0.08	0.03	0.03	0.18	0.02	0.02
Theta.g	-0.27	0.05	0.05	-0.08	0.03	0.03	0.18	0.02	0.02

Omega	9.18	0.04	0.08	7.14	0.03	0.06	5.85	0.02	0.05
SEM.tau	-1.78	0.05	0.05	-1.79	0.03	0.03	-0.55	0.01	0.01
SEM.cong	-0.28	0.05	0.05	-0.85	0.03	0.03	0.12	0.01	0.01
SEM.CE	-1.45	0.05	0.05	-1.05	0.03	0.03	0.23	0.01	0.01
12-Item scale	n=50			n=150			n=500		
EC=0; ρ=0.87	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
Alpha	-1.44	0.03	0.03	-0.69	0.01	0.01	-0.39	0.01	0.01
Glb	7.28	0.02	0.07	4.55	0.01	0.04	2.62	0.01	0.02
Theta	0.09	0.03	0.03	0.13	0.01	0.01	0.30	0.01	0.01
Theta.g	0.09	0.03	0.03	0.13	0.01	0.01	0.30	0.01	0.01
Omega	2.52	0.03	0.04	1.70	0.01	0.02	1.24	0.01	0.01
SEM.tau	-1.03	0.03	0.03	-0.51	0.01	0.01	-0.28	0.01	0.01
SEM.cong	-0.36	0.03	0.03	-0.08	0.01	0.01	0.09	0.01	0.01
SEM.CE	-0.16	0.03	0.03	-0.09	0.02	0.02	0.09	0.01	0.01

6-Item Scale Model 5

The results for Model 5 (Table 6) resembled those of Model 2 (Table 3). All the sample estimates of the reliability coefficients had positive bias values except SEM.CE (CSM). SEM.CE (CSM) had its bias under 0.5% under all the conditions except when sample size was 50. Under that condition, SEM.CE (CSM) had a bias of 4.04%, which was still much smaller than the biases of the other coefficients. Again, omega had the largest bias values under the conditions of 6 items (respectively 21.43%, 20.19% and 20.08%), which were slightly larger than the corresponding bias values of omega (respectively 21.13%, 19.29% and 19.06%). The remaining reliability coefficients (i.e., alpha, theta, theta.g, SEM.tau [USM] and SEM.cong [USM]) had similar biases ranging from 10.04% to 12.10%.

Under the condition of 6 items In Model 5 (Table 6), SEM.CE (CSM) had an abnormally large SE (0.21) when sample size was 50. When sample size increased to 150, the SE of SEM.CE (CSM) became as normal as the other coefficients. All the other coefficients had low SEs ranging from 0.01 to 0.04 and their SEs all decreased slightly as sample size increased.

Compared with the corresponding conditions in Model 4 (Table 5), the RMSEs of these reliability coefficients were larger as there were correlated errors in Model 5 (Table 6). SEM.CE (CSM) had the smallest RMSEs under all the conditions (ranging from 0.01 to 0.04) except for the condition of 50 observations. Under that condition, it had the largest RMSE (0.21) and alpha had the smallest RMSE (0.08). The glb and omega had similar larger RMSE values, ranging from 0.14 to 0.16. The remaining reliability coefficients had their RMSEs ranging from 0.08 to 0.10.

12-Item Scale Model 5

Under the conditions of 12 items in Model 5 (Table 6), SEM.CE (CSM) had the smallest bias (below 0.5%) and the glb had the largest bias (above 10%), irrespective of sample size. Omega consistently had a smaller bias than the glb but larger bias than the remaining reliability coefficients (i.e., alpha, theta, theta.g, SEM.tau [USM] and SEM.cong [USM]), which had similar biases ranging from 4.58% to 5.95%.

When item number was 12, the SEs of these reliability coefficients were all small, ranging from 0.01 to 0.04. As sample size increased, the SEs of these reliability coefficients all decreased. When sample size was 500, all the coefficients had a SE of 0.01.

The RMSEs of these reliability coefficients ranged from 0.01 to 0.11. The glb had larger RMSE values (respectively 0.11, 0.10 and 0.09) than omega (respectively 0.07, 0.07 and 0.06). SEM.CE (CSM) had the smallest RMSEs (respectively 0.04, 0.02 and 0.01). The remaining reliability coefficients (i.e., alpha, theta, theta.g, SEM.tau [USM], and SEM.cong [USM]) all had similar RMSEs ranging from 0.04 to 0.06.

Table 6

Percentage Bias, SE and RMSE of Reliability Estimates of the Data Generated from the Congeneric Model with Error Correlation at 0.2 (Model 5)

6-Item scale	n=50	n=150	n=500

EC=0.2; ρ=0.72	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
Alpha	10.21	0.04	0.08	10.04	0.03	0.08	10.67	0.01	0.08
Glb	21.13	0.03	0.16	19.29	0.02	0.14	19.06	0.01	0.14
Theta	10.92	0.04	0.09	11.11	0.02	0.08	11.35	0.01	0.08
Theta.g	11.40	0.04	0.09	11.11	0.02	0.08	11.35	0.01	0.08
Omega	21.43	0.03	0.16	20.19	0.02	0.15	20.08	0.01	0.14
SEM.tau	10.93	0.04	0.09	10.43	0.03	0.08	10.93	0.04	0.09
SEM.cong	12.10	0.04	0.10	11.08	0.03	0.09	11.53	0.01	0.08
SEM.CE	4.04	0.21	0.21	0.33	0.04	0.04	0.25	0.02	0.02

12-Item scale		n=50			n=150			n=500	
EC=0.2; ρ=0.84	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
Alpha	4.58	0.03	0.05	5.36	0.01	0.05	5.21	0.01	0.04
Glb	13.43	0.02	0.11	11.61	0.01	0.10	10.76	0.01	0.09
Theta	5.62	0.03	0.06	5.87	0.02	0.05	5.87	0.01	0.05
Theta.g	5.62	0.03	0.06	5.87	0.02	0.05	5.87	0.01	0.05
Omega	8.27	0.02	0.07	7.82	0.01	0.07	7.36	0.01	0.06
SEM.tau	4.93	0.03	0.05	5.32	0.01	0.05	5.52	0.01	0.05
SEM.cong	5.57	0.02	0.05	5.64	0.01	0.05	5.95	0.01	0.05
SEM.CE	-0.19	0.04	0.04	0.01	0.02	0.02	0.49	0.01	0.01

6-Item Scale Model 6

As the measurement error correlation rose to 0.5 in Model 6 (Table 7), all the sample estimates of the reliability coefficients had considerably larger positive bias except SEM.CE (CSM). SEM.CE (CSM) had the negative and smallest bias in absolute value. The glb was the least accurate reliability coefficient, the bias of which ranged from 48.20% to 48.55% under the conditions of 6 items. Omega was less biased than the glb under all the conditions, but it was less accurate than the remaining reliability coefficients (i.e., alpha, theta, theta.g, SEM.tau [USM], and SEM.cong [USM]). The remaining reliability coefficients had similar biases ranged from 27.18% to 28.78%.

In Model 6 (Table 7), SEM.CE (CSM) had the largest large SE (0.23) when sample size was 50. When sample size increased to 150 or more, the SEs of SEM.CE became much smaller and comparable with the other coefficients. All the other coefficients had small SEs ranging from 0.00 to 0.05 and their SEs all decreased slightly as sample size increased.

In Model 6 (Table 7), SEM.CE (CSM) had the smallest RMSE (ranging from 0.01 to 0.04) except for the condition of 50 observations. Under that condition, it had a large RMSE (0.23). The glb had the largest RMSE value (0.32); omega had a smaller RMSE (0.28) than the glb but greater than the remaining coefficients (ie., alpha, theta, theta.g, SEM.tau [USM], and SEM.cong [USM]). The remaining coefficients' RMSEs ranged from 0.18 to 0.19 under the conditions of 6 items.

12-Item Scale Model 6

Under the conditions of 12 items in Model 6 (Table 7), SEM.CE (CSM) still had the smallest bias and the glb had the largest bias, irrespective of sample size. Omega consistently had a smaller bias than the glb but larger bias than the remaining reliability coefficients (i.e., alpha, theta, theta.g, SEM.tau [USM] and SEM.cong [USM]), which had similar biases ranging from 13.47% to 14.54%.

The SEs of these reliability coefficients were all small (ranging from 0.01 to 0.02) except SEM.CE (CSM) under the condition of 50 observations. Under that condition, it had a comparatively large SE (0.15). As sample size increased to 150, SEM.CE's SE decreased to 0.02, as small as those of the other coefficients. When sample size was 500, all the coefficients had a SE of 0.01 or below.

The RMSEs of these reliability coefficients ranged from 0.01 to 0.21. The glb had largest RMSE (0.21) and omega had the second largest RMSE (0.14). Coefficient alpha, theta, theta.g, SEM.tau [USM], and SEM.cong [USM] all had a RMSE of 0.11. SEM.CE (CSM) had a rather large RMSE (0.15) when sample size was 50. When sample size increased to 150 or more, SEM.CE had the smallest RMSE.

Table 7

Percentage Bias, SE and RMSE of Reliability Estimates of the Data Generated from the Congeneric Model with Error Correlation at 0.5 (Model 6)

6-Item scale		n=50			n=150			n=500	
EC=0.5; ρ=0.65	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
Alpha	27.20	0.04	0.18	28.02	0.02	0.18	28.29	0.01	0.18
Glb	48.20	0.01	0.31	48.55	0.01	0.32	48.42	0.00	0.31
Theta	26.98	0.04	0.18	28.78	0.02	0.19	28.51	0.01	0.19
Theta.g	28.18	0.05	0.19	28.78	0.02	0.19	28.51	0.01	0.19
Omega	43.35	0.02	0.28	43.72	0.01	0.28	43.66	0.00	0.28
SEM.tau	27.63	0.04	0.18	28.29	0.02	0.18	28.45	0.01	0.19
SEM.cong	27.18	0.05	0.18	27.37	0.03	0.18	27.63	0.01	0.18
SEM.CE	5.97	0.23	0.23	0.52	0.04	0.04	0.60	0.02	0.02

12-Item scale		n=50			n=150			n=500	
EC=0.5; ρ=0.78	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
Alpha	13.47	0.02	0.11	13.85	0.01	0.11	14.04	0.01	0.11
Glb	26.53	0.00	0.21	26.41	0.00	0.21	26.38	0.00	0.21
Theta	13.81	0.02	0.11	14.36	0.01	0.11	14.54	0.01	0.11
Theta.g	13.81	0.02	0.11	14.36	0.01	0.11	14.54	0.01	0.11
Omega	18.35	0.02	0.14	18.24	0.01	0.14	18.06	0.01	0.14
SEM.tau	13.79	0.02	0.11	13.99	0.01	0.11	14.10	0.01	0.11
SEM.cong	14.37	0.02	0.11	14.40	0.01	0.11	14.38	0.01	0.11
SEM.CE	-2.14	0.15	0.15	0.24	0.02	0.02	0.26	0.01	0.01

In summary, the reliability coefficient SEM.CE had the smallest bias under most conditions in the congeneric models (Models 4 to 6). Under the conditions in Model 4 (the congeneric model with independent errors), theta, theta.g, and SEM.cong had the smallest bias or near the smallest bias. The biases of alpha, SEM.tau and SEM.CE were slightly larger than theta, theta.g and SEM.cong, but their biases became similar as sample size increased. However, as error correlation occurred in Model 2 (the tau-equivalent model with error correlation at 0.2) and in Model 3 (the tau-equivalent model with error correlation at 0.5), only SEM.CE had the smallest bias. The biases of theta, theta.g, alpha, SEM.tau and SEM.cong were distinguishable larger than SEM.CE when there were error correlations, and the discrepancies between the bias of SEM.CE and those of the other coefficients became larger as error correlation increased. Similar as in the tau-equivalent models (Models 1 to 3), the glb and omega had larger biases than the other coefficients in Models 4 to 6. The glb had the largest bias values under the conditions of 12 items in Models 4 and 5 and under all the conditions in Model 6, while omega had the largest bias values under the remaining conditions.

All the reliability coefficients had very small SEs ranging from 0.00 to 0.06 under all the conditions in Model 1. As there were correlated errors in Models 2 and 3, SEM.CE had rather large SE values when sample size was 50. When sample size increased, the SE of SEM.CE became similarly small as the other coefficients. The glb had the smallest SE values under most conditions across Models 1 to 3. In short, the reliability coefficients had similarly small SEs and their SEs decreased as sample size increased from 50 to 150 or more.

Similar to the tau-equivalent models (Models 1 to 3), SEM.CE had the smallest RMSE under most conditions in the congeneric models (Models 4 to 6). Under the conditions of 50 observations in Models 4 and 5, SEM.CE had rather large RMSE values. The glb had the largest RMSE values under the conditions of 12 items in Models 4 and 5 and under all the conditions in Model 6, SEM.CE had the largest RMSE under the condition of 6 items and 50 observations in Model 5, and omega had the largest RMSE values under the remaining conditions. Alpha, theta, theta.g, SEM.tau, and SEM.cong had comparatively moderate RMSE values. The RMSEs of all the coefficients increased as error correlation increased or item number decreased.

4.1.3 Reliability Estimates for Correlated Models

The relative percentage biases, SEs and RMSEs of all the reliability coefficients for Model 7 (the correlated factor model with factor correlation at 0.2 and independent errors), Model 8 (the correlated factor model with factor correlation at 0.2 and error correlation at 0.2), Model 9 (the correlated factor model with factor correlation at 0.2 and error correlation at 0.5), Model 10 (the correlated factor model with factor correlation at 0.6 and independent errors), Model 11 (the correlated factor model with factor correlation at 0.6 and error correlation at 0.2) and Model 12 (the correlated factor model with factor correlation at 0.6 and error correlation at 0.5) are presented in Table 8 to 13. The results are summarized in the order of bias, SE and RMSE under each measurement model with respect to the conditions of 6 items and 12 items.

6-Item Scale Model 7

Under the conditions of 6 items in Model 7, the glb, omega, theta.g, SEM.CF (CSM) and SEM.CFCE (OSM) had positive bias values; alpha, theta, SEM.tau (USM) and SEM.cong (USM) had negative bias; SEM.CE (USM) had a positive bias when sample size was 50 and negative bias values when sample size was 150 or more. None of the coefficients consistently had the smallest bias. Theta.g and SEM.CE had the smallest bias when sample size was 50, and SEM.CF had the smallest bias when sample size was 150 or more. The glb, omega had larger bias values than SEM.CF under all the conditions of 6 items, but their biases were smaller than theta.g when sample size was 150 or more. Coefficient alpha, theta, SEM.tau and SEM.cong all had similar negative biases ranging from -9.73% to -14.73%. SEM.CFCE had the abnormally largest bias irrespectively of sample size.

When sample size was 50, the non-SEM estimates of reliability (i.e., alpha, the glb, theta, theta.g and omega) all had smaller SEs (ranging from 0.06 to 0.09) than the SEM estimates of reliability (ranging from 0.20 to 1.81). When sample size was 150 or more, SEM.tau (USM), SEM.cong (USM) and SEM.CF (CSM) had similar SEs as the non-SEM estimates of reliability, whereas SEM.CE (USM) and SEM.CFCE (OSM) still had much larger SEs (ranging from 0.38 to 0.69) than the non-SEM estimates of reliability (ranging from 0.02 to 0.07).

The RMSEs of the non-SEM estimates of reliability all had smaller RMSEs (ranging from 0.03 to 0.15) than the SEM estimates of reliability except for SEM.CF. SEM.CF had the smallest RMSE when sample size was 150 or more, but its RMSE was larger than the non-SEM estimates of reliability when sample size was 50. When sample size was 50, the glb had the smallest RMSE (0.09), theta, theta.g and omega had a similar but slightly larger RMSE (0.10), and alpha had a RMSE of 0.15. As to the SEM estimates of reliability, SEM.tau (USM) and SEM.CF (CSM) had comparatively smaller RMSEs, especially when sample size was large. SEM.cong (USM) had large RMSE (0.71) when sample size was 50, but its RMSE dropped at the level of the non-SEM estimates of reliability when sample size increased to 150 or more. SEM.CE (USM) also had large RMSEs (respectively 0.82, 0.69 and 0.38) and SEM.CFCE (OSM) had the largest RMSEs (respectively 1.87, 0.73 and 0.63).

12-Item Scale Model 7

When the item number increased to 12, theta, theta.g, omega, SEM.CF (CSM) and SEM.CFCE (OSM) had their biases below 5% regardless of sample size. Among them, omega and SEM.CF (CSM) were least biased. Alpha, SEM.tau (USM) and SEM.cong (USM) had slightly larger biases (in absolute value) than theta (in absolute value) and theta.g, and rather larger biases than omega and SEM.CF. On the other hand, SEM.CE (USM) had the bias (in absolute value) much higher than all the other coefficients. The glb had the second largest bias when sample size was 150 or less, but its bias dropped fast as the sample size increased.

Under the conditions of 12 items, the SEs of the non-SEM estimates of reliability ranged from 0.01 to 0.07, and those of the SEM estimates of reliability ranged from 0.01 to 0.15. The SEs of the SEM estimates of reliability dropped quickly as item number increased to 12. When sample size was 50, all the reliability coefficients, except SEM.CE (USM) (0.15) and SEM.CFCE (OSM) (0.12), had small SEs ranging from 0.04 to 0.07. When sample size increased to 150 or more, all the coefficients had similarly small SEs ranging from 0.01 to 0.04.

When sample size was 50, theta and SEM.CF (CSM) had the smallest RMSE (0.05), whereas SEM.CE (USM) had the largest RMSE (0.21). As to the remaining coefficients, SEM.CFCE (OSM), the glb and alpha had larger RMSEs (ranging from 0.09 to 0.12) than theta.g, omega, SEM.tau and SEM.cong (ranging from 0.05 to 0.07). When sample size was 150, SEM.CF had the smallest RMSE (0.02), SEM.CE had the largest RMSE (0.15), and the remaining coefficients had their RMSEs in the range of 0.03 to 0.07. When sample size was 500, omega, SEM.CF and SEM.CFCE all had the smallest RMSE (0.02), SEM.CE had the largest RMSE (0.15), and the remaining coefficients had a RMSE of 0.04 or 0.05.

Table 8

Percentage Bias, SE and RMSE of Reliability Estimates of the Data Generated from the Correlated Factor Model with Factor Correlation at 0.2 and Independent Errors (Model 7)

6-Item scale	n=50			n=150			n=500		
EC=0; ρ=0.67	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
Alpha	-17.43	0.09	0.15	-14.72	0.07	0.12	-13.97	0.02	0.10
Glb	10.52	0.06	0.09	5.54	0.04	0.05	3.09	0.02	0.03
Theta	-9.73	0.07	0.10	-12.88	0.05	0.10	-12.88	0.03	0.09
Theta.g	6.28	0.09	0.10	8.21	0.06	0.08	8.96	0.04	0.07
Omega	10.21	0.07	0.10	5.73	0.07	0.08	5.66	0.06	0.07
SEM.tau	-16.54	0.11	0.16	-16.49	0.06	0.13	-13.82	0.02	0.09
SEM.cong	-14.84	0.70	0.71	-13.70	0.06	0.11	-12.45	0.02	0.09
SEM.CE	6.28	0.82	0.82	-6.13	0.69	0.69	-22.21	0.38	0.41
SEM.CF	8.67	0.20	0.21	1.69	0.04	0.04	0.55	0.02	0.02
SEM.CFCE	67.28	1.81	1.87	40.58	0.68	0.73	38.31	0.57	0.63
12-Item scale	n=50			n=150			n=500		
EC=0; ρ=0.80	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
Alpha	-8.85	0.06	0.09	-6.61	0.03	0.05	-6.26	0.02	0.05
Glb	11.90	0.03	0.11	7.26	0.02	0.07	4.16	0.01	0.04

Theta	-3.76	0.04	0.05	-5.80	0.02	0.04	-5.94	0.01	0.04
Theta.g	4.13	0.04	0.06	3.66	0.02	0.04	3.51	0.02	0.04
Omega	-0.11	0.07	0.07	0.16	0.04	0.04	0.79	0.02	0.02
SEM.tau	-7.83	0.05	0.07	-6.58	0.03	0.05	-6.18	0.02	0.05
SEM.cong	-6.18	0.05	0.06	-6.06	0.03	0.05	-6.04	0.02	0.05
SEM.CE	-19.35	0.15	0.21	-18.84	0.06	0.15	-19.34	0.03	0.15
SEM.CF	-0.11	0.05	0.05	0.40	0.02	0.02	0.43	0.01	0.02
SEM.CFCE	2.30	0.12	0.12	0.59	0.03	0.03	0.50	0.01	0.02

6-Item Scale Model 8

Under the conditions of 6 items in Model 8 (Table 9), SEM.cong (USM) had the smallest bias (below 3% in absolute value) and SEM.CFCE (CSM) had the largest bias (above 50%). Coefficient alpha and SEM.tau (USM) had rather low biases (ranging from 3.97% to 6.16%). Theta had a lightly larger bias than alpha and SEM.tau, but it bias was smaller than SEM.CF (USM) (ranging from 21.85% to 24.36%). The glb, theta.g and omega had similarly large biases (ranging from 28.26% to 32.59%). SEM.CE (USM) had very inconsistent bias values. When sample size was 50, SEM.CE had a positive bias of 5.75%; when sample size was 150, it had a negative bias of -0.90%, which was small in magnitude; when sample size was 500, SEM.CE had a bias of -21.70%, which was rather large in magnitude.

Except SEM.cong, SEM.CE and SEM.CFCE, all the reliability coefficients had small SEs (ranging from 0.02 to 0.08). When sample size was 50, SEM.cong (0.20) had a much smaller SE than SEM.CE (0.85) and SEM.CFCE (1.5). When sample size was 150 or more, SEM.cong had a similarly small SE as the other coefficients (i.e., alpha, theta, theta.g, the glb, omega, SEM.tau and SEM.CF), whereas SEM.CE and SEM.CFCE still had rather large SEs (ranging from 0.23 to 0.90).

The RMSEs of these reliability coefficients in Model 8 (Table 9) were a little different from their corresponding values in Model 7 (Table 8) due to the discrepancies among non-SEM estimates of reliability. Alpha and theta had comparatively small RMSEs (ranging from 0.04 to 0.08), whereas the glb, omega and theta.g all had large RMSEs (ranging from 0.17 to 0.20) under all the conditions of 6 items in Model 8. As to the SEM estimates of reliability, SEM.tau (USM) and SEM.cong (USM) had comparatively smaller RMSEs, especially when sample size was large (150 or more). SEM.CF (USM) had slightly larger RMSEs than SEM.tau and SEM.cong, ranging from 0.13 to 0.16 under all the conditions of 6 items in Model 8. SEM.CE (USM) had the second large RMSEs (respectively 0.85, 0.63 and 0.27) and SEM.CFCE (CSM) had the largest RMSEs (respectively 1.54, 0.83 and 0.95).

12-Item Scale Model 8

When item number increased to 12, all the coefficients, except SEM.cong and

SEM.CE, had decreased biases. SEM.CFCE (CSM) had minimal bias values (ranging from 0.08% to 1.43%). Alpha, theta, SEM.tau (USM), SEM.cong (USM) also had rather small biases in absolute values (ranging from 1.95 to 5.37%). Theta.g, omega and SEM.CF (USM) had rather large bias (ranging from 9.60% to 14.04%). SEM.CE (USM) and the glb had the largest bias in absolute values (ranging from 18.29% to 22.61%). However, SEM.CE had negative bias and the glb had positive bias.

When item number was 12, the reliability coefficients all had quite small SEs ranging from 0.01 to 0.06, except SEM.CE (USM) (0.12) and SEM.CFCE (CSM) (0.12) under the condition of 50 observations. When sample size was 150 or more, SEM.CE and SEM.CFCE also fell into the range of 0.01 to 0.06.

Under the conditions of 12 items, alpha, theta, SEM.tau (USM) and SEM.cong (USM) had small RMSEs ranging from 0.03 to 0.06; omega, theta.g and SEM.CF (USM) had moderate RMSEs ranging from 0.09 to 0.13; the glb and SEM.CE (USM) had comparatively large RMSEs ranging from 0.15 to 0.19. SEM.CFCE (CSM) had a moderate RMSE (0.12) when sample size was 50 and the smallest RMSE (no more than 0.03) when sample size was 150 or more.

Table 9

Percentage Bias, SE and RMSE of Reliability Estimates of the Data Generated from the Correlated Factor Model with Factor Correlation at 0.2 and Error Correlation at 0.2 (Model 8)

6-Item scale		n=50			n=150			n=500	
EC=0.2; ρ=0.61	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
Alpha	6.16	0.07	0.08	3.97	0.05	0.06	5.43	0.02	0.04
Glb	32.59	0.04	0.20	28.95	0.03	0.18	27.92	0.02	0.17
Theta	9.05	0.05	0.07	5.84	0.04	0.05	5.56	0.03	0.05
Theta.g	31.87	0.08	0.21	31.23	0.04	0.19	31.79	0.03	0.20
Omega	31.66	0.06	0.20	28.26	0.06	0.18	29.43	0.03	0.18
SEM.tau	5.69	0.08	0.09	4.62	0.05	0.06	5.57	0.02	0.04
SEM.cong	0.80	0.20	0.20	-1.72	0.04	0.04	-2.82	0.02	0.03
SEM.CE	5.75	0.85	0.85	-0.90	0.63	0.63	-21.70	0.23	0.27
SEM.CF	24.36	0.06	0.16	22.51	0.03	0.14	21.85	0.02	0.13
SEM.CFCE	55.26	1.50	1.54	48.75	0.77	0.83	50.46	0.90	0.95
12-Item scale		n=50			n=150			n=500	
EC=0.2; ρ=0.75	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
Alpha	1.95	0.05	0.06	3.39	0.03	0.05	3.24	0.02	0.04
Glb	22.61	0.02	0.19	19.68	0.02	0.17	18.29	0.01	0.15

Theta	4.83	0.04	0.06	4.32	0.03	0.05	3.67	0.01	0.04
Theta.g	13.45	0.04	0.12	14.76	0.03	0.13	14.04	0.02	0.12
Omega	10.15	0.06	0.11	10.61	0.05	0.11	11.99	0.01	0.10
SEM.tau	2.29	0.05	0.06	3.65	0.03	0.05	3.32	0.02	0.04
SEM.cong	-5.37	0.04	0.05	-3.43	0.03	0.04	-3.01	0.02	0.03
SEM.CE	-18.91	0.12	0.19	-18.85	0.06	0.15	-19.27	0.04	0.15
SEM.CF	9.60	0.04	0.09	10.13	0.02	0.09	10.19	0.01	0.09
SEM.CFCE	1.43	0.12	0.12	0.13	0.03	0.03	0.08	0.02	0.02

6-Item Scale Model 9

As the error correlation further increased to 0.5 in Model 9 (Table 10), all the coefficients had positive and sizable biases except SEM.CE (USM). SEM.CE (USM) had positive bias when sample size was no more than 150. In addition, SEM.CE had the least amount of bias (respectively 12.42%, 1.21%, -2.68%) in absolute values under condition of 6 items. The glb had the largest bias under all the conditions of 6 items. Omega and theta.g had smaller bias than the glb but much larger bias than the remaining coefficients (i.e., alpha, theta, SME.tau [USM] and SEM.cong [USM]). The remaining coefficients had similar large biases ranging from 32.51% to 35.58%.

Under the conditions of 6 items in Model 9 (Table 10), the non-SEM estimates of reliability (i.e., alpha, the glb, theta, theta.g and omega), SEM.tau (USM), and SEM.CF

(USM) had small SEs (ranging from 0.01 to 0.08), whereas SEM.cong (USM), SEM.CE (USM) and SEM.CFCE (CSM) had large SEs (all above 0.30).

The RMSEs of all the reliability coefficients in Model 9 (Table 10), except SEM.CE (USM) and SEM.CFCE (CSM), became larger than their corresponding values in Model 8 (Table 9) as the error correlation increased from 0.2 to 0.5. Alpha, theta and SEM.tau had the smallest RMSE values under the conditions of 6 items. The glb, omega, theta.g and SEM.CF (USM) had large RMSEs, ranging from 0.32 to 0.47. SEM.CE (USM) had larger RMSEs than the glb, omega, theta.g and SEM.CF (USM), but smaller RMSEs than SEM.cong (USM). SEM.cong (USM) had the largest RMSE values (ranging from 0.62 to 0.73). The RMSE of SEM.CFCE (CSM) changed rather differently from other reliability coefficients. When sample size was 50, SEM.CFCE (CSM) had very large RMSE (0.70); however, as sample size increased, its RMSE dropped dramatically (0.26 for 500 observations).

12-Item Scale Model 9

When item number went up from 6 to 12, the bias in these coefficients, except SEM.CE (USM), all decreased dramatically, and SEM.CFCE (CSM) became the coefficient with the smallest bias values (respectively 0.63%, 0.38%, and 0.25%). SEM.CE (USM), on the other hand, had sizeable negative bias values (above 20%), the absolute values of which were slightly smaller than those of theta.g and SEM.CF

(USM). The glb had the largest bias under all the conditions of 6 items. The remaining coefficients (i.e., alpha, theta, SEM.tau [USM] and SEM.cong [USM]) had similar large biases ranging from 17.24% to 18.97% under the conditions of 12 items.

When item number increased to 12, the SEs of SEM.cong (USM), SEM.CE (USM) and SEM.CFCE (CSM) dropped markedly, and the SEs of the other coefficients all decreased. All the coefficients, except SEM.CE, had their SEs in the range of 0.00 to 0.08. Although SEM.CE had the largest SE values (respectively 0.15, 0.09 and 0.05), they were considerably smaller compared with its corresponding SE values under the conditions of 6 items.

When item number was 12, SEM.CFCE (CSM) had the smallest RMSE values (respectively 0.07, 0.04 and 0.02). Omega, theta.g, SEM.CE (USM) and SEM.CF (USM) had larger RMSEs than alpha, theta, SEM.tau (USM) and SEM.cong (USM), but smaller RMSEs (ranging from 0.18 to 0.25) than the glb (0.37). The glb had the largest RMSEs under the conditions of 12 items.

Table 10

Percentage Bias, SE and RMSE of Reliability Estimates of the Data Generated from the Correlated Factor Model with Factor Correlation at 0.2 and Error Correlation at 0.5 (Model 9)

6-Item scale	n=50	n=150	n=500

EC=0.5; ρ=0.53	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
Alpha	32.51	0.07	0.19	33.81	0.04	0.18	34.68	0.02	0.18
Glb	75.23	0.02	0.40	75.74	0.01	0.40	75.43	0.01	0.40
Theta	35.43	0.05	0.19	35.58	0.04	0.19	34.49	0.02	0.18
Theta.g	67.53	0.08	0.37	67.57	0.05	0.36	68.02	0.03	0.36
Omega	69.21	0.03	0.37	70.47	0.02	0.37	70.28	0.01	0.37
SEM.tau	33.09	0.06	0.19	34.53	0.04	0.19	34.89	0.02	0.19
SEM.cong	32.25	0.71	0.73	23.17	0.65	0.66	15.53	0.61	0.62
SEM.CE	12.42	0.50	0.50	1.21	0.50	0.50	-2.68	0.37	0.37
SEM.CF	67.08	0.07	0.36	62.45	0.02	0.33	61.17	0.01	0.32
SEM.CFCE	52.15	0.64	0.70	40.83	0.45	0.50	20.62	0.26	0.28

12-Item scale		n=50			n=150			n=500	
EC=0.5; ρ=0.68	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
Alpha	17.24	0.04	0.15	17.01	0.04	0.15	18.12	0.01	0.15
Glb	44.94	0.00	0.37	44.79	0.00	0.37	44.76	0.00	0.37
Theta	18.97	0.03	0.16	18.38	0.03	0.16	18.46	0.01	0.16
Theta.g	29.03	0.03	0.24	29.75	0.03	0.25	30.32	0.01	0.25
Omega	27.72	0.04	0.23	27.90	0.04	0.23	29.49	0.01	0.24
SEM.tau	17.71	0.04	0.15	17.81	0.02	0.15	18.19	0.01	0.15
SEM.cong	20.85	0.08	0.19	18.13	0.02	0.15	18.09	0.01	0.15

SEM.CE	-21.96	0.15	0.22	-21.16	0.09	0.18	-21.96	0.05	0.17
SEM.CF	24.37	0.04	0.21	22.49	0.03	0.19	21.91	0.02	0.18
SEM.CFCE	0.63	0.07	0.07	0.38	0.04	0.04	0.25	0.02	0.02

Models 10 to 12 (Tables 11 to 13) resembled Models 7 to 9 (Tables 8 to 10) in every way except the factor correlation increased to 0.6 in Model 10 to 12. Therefore, the results presented in Tables 11 to 13 are not summarized in detail. The differences in the pairs of models (i.e., Model 7 and Model 10, Model 8 and Model 11, Model 9 and Model 12) are pointed out, in addition with a brief summary of the estimates of reliability with the largest or smallest bias, SE and RMSE.

In Model 10 (Table 11), all the coefficients, except theta.g, had considerably smaller biases, SEs and RMSEs than their corresponding values in Model 7 (Table 8). Similar to Model 7, when item number was 6, SEM.CF (CSM) had the smallest bias (ranging from 0.45% to 1.18%) and SEM.CFCE (OSM) had the largest bias (22.74%) when sample size was 50, but the bias of SEM.CFCE (OSM) was substantially smaller than the corresponding values in Model 7. When item number was 12, SEM.CFCE (OSM) had the smallest bias (ranging from -0.60% to -0.52%), which was smaller than SEM.CF (CSM) (ranging from -5.95% to -5.75%), whereas theta.g had the largest bias (ranging from 7.58% to 22.11%), which was different from the results in Model 7 where SEM.CE (USM) had the largest bias. In terms of the SEs of these coefficients, the glb had the smallest SEs, whereas SEM.CFCE (OSM) had largest SEs under most of the conditions. When sample size was 500, all the coefficients had low SEs ranging from 0.01 to 0.05. As to the RMSEs of these coefficients, SEM.CF (CSM) had the smallest values under all the conditions, and other SEM estimates (except SEM.tau [USM]) had substantially larger RMSE values when both item number and sample size were small.

Table 11

Percentage Bias, SE and RMSE of Reliability Estimates of the Data Generated from the Correlated Factor Model with Factor Correlation at 0.6 and Independent Errors (Model 10)

6-Item scale	n=50			n=150			n=500		
EC=0; ρ=0.73	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
Alpha	-8.18	0.08	0.10	-7.47	0.05	0.07	-5.44	0.02	0.04
Glb	8.26	0.05	0.08	4.37	0.04	0.05	2.41	0.02	0.03
Theta	-5.26	0.06	0.07	-5.14	0.04	0.05	-4.78	0.02	0.04
Theta.g	16.41	0.09	0.15	13.36	0.07	0.12	14.90	0.06	0.12
Omega	9.10	0.05	0.08	6.86	0.04	0.06	5.82	0.03	0.05
SEM.tau	-7.18	0.08	0.10	-7.12	0.05	0.07	-5.30	0.02	0.04
SEM.cong	3.88	0.37	0.37	-6.37	0.05	0.07	-4.95	0.02	0.04
SEM.CE	0.95	0.50	0.50	-12.18	0.06	0.11	-13.11	0.03	0.10

SEM.CF	1.18	0.06	0.06	0.93	0.03	0.03	0.45	0.02	0.02
SEM.CFCE	22.74	0.39	0.42	9.27	0.21	0.22	2.12	0.05	0.05
12-Item scale		n=50			n=150			n=500	
EC=0; ρ=0.85	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
Alpha	-3.88	0.04	0.05	-2.79	0.02	0.02	-2.40	0.01	0.02
Glb	8.87	0.02	0.08	5.42	0.01	0.05	3.11	0.01	0.03
Theta	-2.35	0.04	0.04	-2.35	0.02	0.02	-2.15	0.01	0.01
Theta.g	22.11	0.06	0.19	10.53	0.04	0.10	7.58	0.01	0.07
Omega	1.80	0.04	0.05	1.31	0.02	0.03	0.92	0.01	0.02
SEM.tau	-3.42	0.04	0.04	-2.62	0.02	0.02	-2.33	0.01	0.01
SEM.cong	-2.66	0.03	0.03	-2.34	0.02	0.02	-2.16	0.01	0.01
SEM.CE	-5.95	0.05	0.06	-5.88	0.03	0.05	-5.74	0.01	0.04
SEM.CF	-1.11	0.04	0.04	-0.69	0.02	0.02	-0.58	0.01	0.01
SEM.CFCE	-0.60	0.04	0.04	-0.58	0.02	0.02	-0.52	0.01	0.01

In Model 11 (Table 12), alpha, theta.g, SEM.tau (USM) and SEM.cong (USM) had larger biases, whereas the remaining had smaller biases than their corresponding values in Model 8 (Table 9). All the coefficients had smaller than or equally low SEs as the corresponding value in Model 8. These coefficients, accordingly, had smaller RMSEs than those in Model 8 except for alpha, theta.g, and SEM.tau (USM). When item number was 6, SEM.CE (USM) had the smallest, which was different from the results in Model 8 where SEM.cong (USM) had the smallest bias. The coefficient with the largest bias was theta.g in Model 11, whereas it was the glb that had the largest bias in Model 8. When item number was 12, SEM.CFCE (CSM) produced the most accurate reliability estimate, which was consistent with the result in Model 8. However, in Model 11 theta.g was the least accurate reliability coefficient, which was different from the result in Model 8 where the glb was the least accurate reliability coefficient. In terms of the SEs of these coefficients, the glb had the smallest SE, whereas SEM.CFCE (OSM) had largest SEs under most of the conditions. When sample size was 500, all the coefficients had low SEs ranging from 0.01 to 0.05. As to the RMSEs of these coefficients, SEM.CFCE (CSM) had the largest RMSEs when sample size was no more than 150 and item number was 6, and theta.g had the largest RMSEs under the remaining conditions. However, SEM.CFCE (CSM) had the smallest RMSEs when item number was increased to 12 or sample size was increased to 500 while keeping item number at 6.

Table 12

Percentage Bias, SE and RMSE of Reliability Estimates of the Data Generated from the Correlated Factor Model with Factor Correlation at 0.6 and Error Correlation at 0.2 (Model 11)

6-Item scale	n=50	n=150	n=500

EC=0.2; ρ=0.67	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
Alpha	8.54	0.06	0.08	8.30	0.03	0.06	9.12	0.02	0.06
Glb	25.31	0.04	0.17	22.70	0.03	0.16	22.01	0.01	0.15
Theta	8.91	0.05	0.08	9.01	0.03	0.07	9.28	0.02	0.07
Theta.g	35.73	0.07	0.25	35.46	0.04	0.24	36.67	0.02	0.25
Omega	25.69	0.04	0.18	24.18	0.02	0.16	23.48	0.02	0.16
SEM.tau	9.36	0.06	0.09	8.63	0.03	0.07	9.27	0.02	0.07
SEM.cong	10.99	0.05	0.09	9.34	0.03	0.07	9.60	0.02	0.07
SEM.CE	2.94	0.40	0.40	-10.64	0.05	0.09	-11.52	0.02	0.08
SEM.CF	18.21	0.05	0.13	17.78	0.03	0.12	17.48	0.01	0.12
SEM.CFCE	30.18	0.79	0.82	10.73	0.23	0.24	2.04	0.05	0.05

12-Item scale		n=50			n=150			n=500	
EC=0.2; ρ=0.80	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
Alpha	3.68	0.04	0.05	4.85	0.02	0.05	4.71	0.01	0.05
Glb	16.99	0.02	0.15	14.79	0.01	0.13	13.73	0.01	0.12
Theta	4.68	0.04	0.06	5.14	0.02	0.05	5.03	0.01	0.05
Theta.g	32.44	0.06	0.27	22.28	0.05	0.19	16.16	0.02	0.14
Omega	9.75	0.03	0.09	9.54	0.02	0.09	9.14	0.01	0.08
SEM.tau	4.11	0.04	0.06	5.01	0.02	0.05	4.79	0.01	0.05
SEM.cong	4.80	0.03	0.06	5.29	0.02	0.05	4.94	0.01	0.05

SEM.CE	-6.33	0.06	0.07	-5.36	0.03	0.05	-5.14	0.02	0.04
SEM.CF	7.05	0.04	0.08	7.55	0.02	0.07	7.68	0.01	0.07
SEM.CFCE	-0.49	0.05	0.05	0.05	0.02	0.02	0.10	0.01	0.01

In Model 12 (Table 13), all the coefficients, except theta.g, had considerably smaller biases, SEs and RMSEs than their corresponding values in Model 9 (Table 10). Similar to Model 9, when item number was 6, SEM. CE (USM) had the smallest bias and SEM.CFCE (CSM) had the largest bias when sample size was 50, but the bias of SEM.CFCE (CSM) was substantially smaller than the corresponding values in Model 9. When item number was 12, SEM.CFCE (OSM) had the smallest bias, whereas theta.g had the largest bias, which was different from the results in Model 9 where the glb had the largest bias. As to the SEs of these coefficients, the glb had the smallest SEs, whereas SEM.CFCE (OSM) had largest SEs under most of the conditions. When sample size was 500, all the coefficients had low SEs ranging from 0.01 to 0.03. With regard to the RMSEs of these coefficients, theta.g had the largest RMSEs under all the conditions; SEM.CE (USM) had the smallest RMSEs when item number was 6 and sample size was 150 or less, and SEM.CFCE (CSM) had the smallest RMSEs under the remaining conditions.

Table 13

Percentage Bias, SE and RMSE of Reliability Estimates of the Data Generated from the Correlated Factor Model with Factor Correlation at 0.6 and Error Correlation at 0.5 (Model 12)

6-Item scale	n=50			n=150			n=500		
EC=0.5; ρ=0.60	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
Alpha	29.35	0.05	0.18	30.62	0.03	0.19	31.03	0.02	0.19
Glb	58.00	0.02	0.35	58.45	0.01	0.35	58.25	0.00	0.35
Theta	28.68	0.05	0.18	31.27	0.03	0.19	30.88	0.02	0.19
Theta.g	68.30	0.10	0.42	63.68	0.04	0.38	63.55	0.02	0.38
Omega	53.12	0.03	0.32	53.68	0.01	0.32	53.47	0.01	0.32
SEM.tau	30.05	0.05	0.19	30.98	0.03	0.19	31.22	0.02	0.19
SEM.cong	27.05	0.07	0.18	25.58	0.04	0.16	25.35	0.03	0.16
SEM.CE	3.87	0.28	0.28	-6.73	0.08	0.09	-8.18	0.03	0.06
SEM.CF	47.88	0.08	0.30	44.20	0.02	0.27	43.12	0.01	0.26
SEM.CFCE	23.37	0.45	0.47	8.07	0.25	0.25	1.00	0.03	0.03
12-Item scale	n=50			n=150			n=500		
EC=0.5; ρ=0.74	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
Alpha	15.12	0.03	0.12	15.50	0.02	0.12	15.81	0.01	0.12
Glb	33.65	0.00	0.26	33.54	0.00	0.26	33.53	0.00	0.26

Theta	15.16	0.03	0.12	15.74	0.02	0.13	16.07	0.01	0.13
Theta.g	53.80	0.07	0.41	52.80	0.06	0.40	54.82	0.06	0.42
Omega	22.46	0.03	0.18	22.51	0.01	0.17	22.42	0.01	0.17
SEM.tau	15.51	0.03	0.13	15.65	0.02	0.12	15.88	0.01	0.13
SEM.cong	14.89	0.03	0.12	15.30	0.02	0.12	15.72	0.01	0.12
SEM.CE	-5.78	0.11	0.12	-4.22	0.03	0.04	-4.01	0.02	0.03
SEM.CF	17.16	0.03	0.14	16.41	0.03	0.13	16.42	0.01	0.13
SEM.CFCE	-0.36	0.06	0.06	0.04	0.03	0.03	0.12	0.02	0.02

In summary, none of the reliability coefficients had the smallest bias under all the conditions in the correlated models. Under the correlated factor models and when the assumption of error independence held as in Models 7 and 10, SEM.CF generated the most accurate estimates of reliability except for the condition that the factor correlation was 0.2, item number was 6 and sample size was 50. Under that condition, theta.g and SEM.CE had the smallest bias. When there were correlated errors but error correlation was low (0.2) as in Models 8 and 11, SEM.cong had the smallest bias when factor correlation was 0.2 and item number was 6, and alpha had the smallest bias when factor correlation was 0.6 and item number was 6. When item number was 12, SEM.CFCE was the least biased. When there were correlated errors and error correlation was moderate (0.5) as in Models 9 and 12, SEM.CE had the smallest bias when item number was 6. When item number was 12, SEM.CFCE had the smallest bias regardless of sample size.

One obvious change in the correlated factor models, if compared with the previous models (i.e., the tau-equivalent models and congeneric models), was theta and theta.g had exactly opposite biases. In the previous models, theta and theta.g had very similar biases. However, in the correlated models, theta had the bias in negative values and theta.g had the bias in positive values; besides, the absolute values of theta's bias increased and those of theta.g decreased as sample size went up.

Under the correlated factor models and when the assumption of error independence held as in Models 7 and 10, all the SEM estimates of reliability, except SEM.tau and SEM.CF, had very large SEs, whereas the non-SEM estimates of reliability all had comparatively small SEs. As sample size or item number increased, the SEs of these reliability coefficients all decreased. As factor correlation increased from 0.2 in Model 7 to 0.6 in Model 10, the SEs of all the reliability coefficients decreased. When error correlation was increased from 0 to 0.2 as in Models 8 and 11, all the reliability coefficients, except SEM.CFCE under the conditions of 6 items, had slightly smaller SEs. Under the conditions of 6 items, SEM.CFCE had increased SEs, especially when sample size was 150 or more. When low error correlation (0.2) as in Models 8 and 11 was increased to moderate error correlation (0.5) as in Models 9 and 12, the SEs of non-SEM estimates and most SEM estimates of reliability decreased trivially. However, the SE of SEM.CFCE decreased remarkably, whereas the SE of SEM.CE increased.

Under the correlated factor models and when the assumption of error independence held as in Models 7 and 10, SEM.CF (CSM) generated the smallest RMSEs except for one condition that the factor correlation was 0.2, item number was 6, and sample size was 50. Under that condition, theta.g had the smallest RMSE. When there were low error correlation (0.2) as in Models 8 and 11, alpha had the smallest RMSEs when item number was 6. When item number was 12, SEM.CFCE (CSM) had the smallest RMSE except under the condition that sample size was 50 and factor correlation was 0.2. When error correlation was increased to 0.5 as in Models 9 and 12, alpha, theta and SEM.tau had very similar RMSEs when item number was 6. Although their RMSEs were smaller than the remaining coefficients, they were still considerably large (ranging from 0.18 to 0.19). When item number was 12, SEM.CFCE (CSM) had the smallest RMSE values regardless of sample size.

4.1.4 Percentage of Estimation Failure Number

Tables 14 to 18 present the rounded percentages of the numbers of estimation failure for each SEM reliability coefficient. Each estimation failure means that there was no convergence for parameter estimation and thus no reliability estimate was obtained from that replication of data. For SEM reliability coefficients of the unidimensional analysis model, that is, SEM.tau, SEM.cong and SEM.CE, the rounded percentages of estimation failure were reported across Models 1 to 12. As shown in Table 14, there was no estimation failure for SEM.tau. For SEM.cong (see Table 15), there were serious estimation convergence problems in Model 9 when item number was 6. The percentage of failure numbers were 59%, 38% and 34% as sample size increased from 50 to 150 and to 500. Although increasing sample size could alleviate the estimation issues, there were still severe estimation problems under the conditions of small item numbers and complicated models. For SEM.CE, the estimation problem became even worse. There were serious estimation convergence problem in Models 7 to 9 when item number was 6. Also, the percentages of estimation failure increased from Model 7 to Model 9. The highest failure percentage, 61%, occurred in Model 9 when sample size was 50 and item number was 6, which was almost as twice as the failure percentage in Model 7 under the same condition. When item number was 12, the estimation problems decreased dramatically, and the worst case was 5% for SEM.CE in Model 9 when sample size was 50.

The SEM reliability coefficients using the multidimensional analysis model, that is, SEM.CF and SEM.CFCE, differed greatly in their percentages of estimation failures. SEM.CF had no severe estimation problems under all the conditions. When sample size was 150 or more, there was almost zero convergence failure across 1000 replications. However, SEM.CFCE had severe estimation problems, especially in Models 7, 8 and 9 when item number was 6. It had similar percentages of the estimation failure among Models 7, 8 and 9 although it used the correctly specified analysis model for data generated from Models 8 and 9.

Table 14

The Rounded Percentage of Estimation Failures in SEM.tau Estimates of Reliability

	6-Item scale			12-Item scale		
Model	n=50	n=150	n=500	n=50	n=150	n=500
1 (CSM)*	0	0	0	0	0	0
2 (USM)*	0	0	0	0	0	0
3 (USM)	0	0	0	0	0	0
4 (USM)	0	0	0	0	0	0
5 (USM)	0	0	0	0	0	0
6 (USM)	0	0	0	0	0	0
7 (USM)	0	0	0	0	0	0
8 (USM)	0	0	0	0	0	0
9 (USM)	0	0	0	0	0	0
10 (USM)	0	0	0	0	0	0
11 (USM)	0	0	0	0	0	0
12 (USM)	0	0	0	0	0	0

Note: (CSM)* means SEM.tau correctly specify the model, and (USM)* means SEM.tau underspecify the model.

Table 15

The Rounded Percentage of Estimation Failures in SEM.cong Estimates of Reliability

	6-Item scale			12-Item scale		
Model	n=50	n=150	n=500	n=50	n=150	n=500
1 (OSM)*	0	0	0	0	0	0
2 (USM)	0	0	0	0	0	0
3 (USM)	2	0	0	0	0	0
4 (CSM)	0	0	0	0	0	0
5 (USM)	0	0	0	0	0	0
6 (USM)	0	0	0	0	0	0
7 (USM)	2	0	0	0	0	0
8 (USM)	3	0	0	0	0	0
9 (USM)	59	38	34	4	0	0
10 (USM)	0	0	0	0	0	0
11 (USM)	0	0	0	0	0	0
12 (USM)	11	9	1	1	0	0

Note: (OSM)* means SEM.cong over-specify the model.

Table 16

The Rounded Percentage of Estimation Failures in SEM.CE Estimates of Reliability

	6-Item scale			12-Item scale		
Model	n=50	n=150	n=500	n=50	n=150	n=500
1 (OSM)	0	0	0	0	0	0
2 (OSM)	1	0	0	0	0	0
3 (OSM)	1	0	0	0	0	0
4 (OSM)	1	0	0	0	0	0
5 (CSM)	1	0	0	0	0	0
6 (CSM)	3	0	0	0	0	0
7 (OSM)	33	24	4	1	0	0
8 (USM)	39	31	13	3	0	0
9 (USM)	61	48	43	5	0	0
10 (OSM)	6	0	0	0	0	0
11 (USM)	9	0	0	1	0	0
12 (USM)	17	2	0	1	0	0

Table 17

The Rounded Percentage of Estimation Failures in SEM.CF Estimates of Reliability

	6-Item scale			12-Item scale		
Model	n=50	n=150	n=500	n=50	n=150	n=500
7 (CSM)	9	0	0	0	0	0

8 (USM)	5	0	0	0	0	0
9 (USM)	7	1	0	5	0	0
10 (CSM)	0	0	0	0	0	0
11 (USM)	0	0	0	0	0	0
12 (USM)	6	1	0	3	0	0

Table 18

The Rounded Percentage of Estimation Failures in SEM.CFCE Estimates of Reliability

	6-Item scale			12-Item scale		
Model	n=50	n=150	n=500	n=50	n=150	n=500
7 (OSM)	59	43	16	1	0	0
8 (CSM)	60	43	17	1	0	0
9 (CSM)	60	48	13	6	0	0
10 (OSM)	16	2	0	1	0	0
11 (CSM)	19	3	0	0	0	0
12 (CSM)	19	4	0	2	3	0

4.2. Results in Stage Two

As indicated in stage one results, when item number was 6, SEM.CE and SEM.CFCE had serious estimation problems in Models 7, 8 and 9, and SEM.cong had serious estimation problems in Model 9. Even when there were no convergence problem and the parameters were estimated for some replications of data, the standard errors of these estimated SEM reliability coefficients were quite high, indicating that these estimates were not stable. The warning in Mplus 7 reported that the residual covariance matrix of the generated sample data is not positive definite, suggesting that the estimated parameters were neither reliable nor accurate. This finding was consistent with the large biases and standard errors in the three coefficients presented in Tables 8, 9 and 10. Therefore, Bayesian estimation was used in the second stage of this study, attempting to solve the estimation problems in the three coefficients when analyzing data generated from Models 7, 8 and 9.

The biases, SEs and RMSEs of BSEM.cong, BSEM.CE and BSEM.CFCE for Model 7 (the correlated factor model with factor correlation at 0.2 and independent errors), Model 8 (the correlated factor model with factor correlation at 0.2 and error correlation at 0.2), Model 9 (the correlated factor model with factor correlation at 0.2 and error correlation at 0.5) are presented in Tables 19 to 21. The results are summarized with the focus on the differences between ML SEM estimates of reliability and Bayesian SEM estimates of reliability regarding their respective biases, SEs and RMSEs. Since no estimation failures were detected when using the Bayesian estimation method, the rounded percentages of estimation failures were not reported for BSEM.cong, BSEM.CE and BSEM.CFCE.

Table 19

Percentage Bias, SE and RMSE of Bayesian SEM Reliability of the Data Generated from the Correlated Factor Model with Factor Correlation at 0.2 and Independent Errors (Model 7)

6-Item scale	n=50			n=150			n=500		
EC=0; ρ=0.67	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
BSEM.cong	-21.69	0.13	0.19	-20.48	0.08	0.16	-19.70	0.04	0.14
BSEM.CE	-48.73	0.16	0.36	-40.91	0.13	0.30	-34.36	0.06	0.24
BSEM.CFCE	-34.93	0.13	0.27	-21.81	0.11	0.18	-10.28	0.10	0.12

Table 20

Percentage Bias, SE and RMSE of Bayesian SEM Reliability Estimates of the Data Generated from the Correlated Factor Model with Factor Correlation at 0.2 and Error Correlation at 0.2 (Model 8)

6-Item scale	n=50			n=150			n=500		
EC=0.2; ρ=0.61	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
BSEM.cong	-3.75	0.11	0.12	-1.56	0.06	0.06	-2.82	0.04	0.04
BSEM.CE	-45.80	0.15	0.32	-39.67	0.13	0.27	-34.44	0.06	0.22
BSEM.CFCE	-27.44	0.13	0.21	-15.18	0.10	0.14	-5.46	0.08	0.09

Table 21

Percentage Bias, SE and RMSE of Bayesian SEM Reliability Estimates of the Data Generated from the Correlated Factor Model with Factor Correlation at 0.2 and Error Correlation at 0.5 (Model 9)

6-Item scale	n=50			n=150			n=500		
EC=0.5; ρ=0.53	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
BSEM.cong	25.53	0.10	0.17	22.02	0.05	0.13	19.21	0.03	0.11
BSEM.CE	-37.87	0.16	0.25	-33.23	0.12	0.21	-29.94	0.08	0.18
BSEM.CFCE	-12.06	0.13	0.14	1.68	0.08	0.08	1.28	0.05	0.05

As shown in Tables 19, 20 and 21, the Bayesian SEM estimates of reliability tended to be lower than the corresponding ML SEM estimates of reliability (see Tables 8, 9 and 10). Comparing Table 19 and Table 8 (Model 7), BSEM.cong and SEM.cong both had negative biases, and BSEM.cong had bias larger than SEM.cong in absolute value. BSEM.CE had negative bias under all the conditions, but SEM.CE had positive bias when sample size was 50. When sample size was small, BSEM.CE had much lower estimates of reliability than SEM.CE. On the other hand, BSEM.CFCE had its bias totally different from the corresponding SEM.CFCE bias. SEM.CFCE had abnormally large positive bias, whereas BSEM.CFCE had negative bias that was sizeable in magnitude, especially when sample size was small. The SEs of Bayesian SEM estimates of reliability were all smaller than the corresponding SEM estimates of reliability except for BSEM.cong under the conditions with 150 or more observations. The RMSEs of BSEM.cong and BSEM.CFCE were lower than the corresponding values of SEM.cong and SEM.CFCE, whereas the RMSEs of BSEM.CE were higher than the corresponding values of SEM.CE.

When there were correlated errors as in Model 8 (see Tables 20 and Table 9), BSEM.cong had similar bias as SEM.cong, especially when sample size was large. BSEM.CE continued to have negative bias values that were much larger in magnitude than the corresponding SEM.CE regardless of sample sizes. BSEM.CFCE had negative bias comparatively smaller in magnitude (especially when sample size was large), which was different from SEM.CFCE under the same conditions in Model 8 where SEM.CFCE had large positive bias. The SEs and RMSEs of Bayesian SEM estimates of reliability were all smaller than the corresponding SEM estimates of reliability except for BSEM.cong under the conditions with 150 or more observations.

As error correlation continued to increase as in Model 9 (see Table 21 and 9), BSEM.cong had relatively similar bias as SEM.cong, BSEM.CE had negative bias much larger in magnitude than the corresponding SEM.CE, and BSEM.CFCE had much smaller bias than SEM.CFCE under the same conditions. The SEs and RMSEs of Bayesian SEM estimates of reliability were all smaller than the corresponding SEM estimates of reliability under all the conditions in Table 21.

CHAPTER 5 DISCUSSION AND CONCLUSIONS

Chapter 5 is organized in the order of the five approaches (i.e., internal consistency, lower bound, principle component analysis, exploratory factor analysis and structural equation modeling) from which the reliability coefficients in this study are derived. Accordingly, the reliability coefficients discussed below are (1) coefficient alpha, (2) the glb, (3) theta and theta.g, (4) omega, (5) SEM estimates of reliability and BSEM estimates of reliability. Then, conclusions are summarized to address the research questions raised in Section 2.4 in Chapter 2. Finally, limitations are pointed out and suggestions for future directions are provided.

5.1 Discussion on the Studied Reliability Coefficients

In this section, the findings of the reliability coefficients studied in my dissertation are discussed, with a focus on the consistencies and contradictions between the findings in this study and previous studies. The unique findings in this study are emphasized and the possible reasons for these findings are discussed.

5.1.1 Coefficient alpha

Coefficient alpha is the most intensively studied reliability coefficient and often used as a reference coefficient when researchers are studying and examining a new reliability coefficient. In this study, the findings related to coefficient alpha are consistent with those in previous studies. First, alpha is negatively biased when the assumption of (essential) tau-equivalence is violated. The degree of the violation of (essential) tau-equivalence is associated with the degree of internal consistency. Specifically, when the assumption of essential tau-equivalence is seriously violated, the internal consistency of a test component tends to be low (see Tang, Cui, & Babenko, 2014). The coefficients based on the internal consistency approach then have a negative bias and are frequently quoted as the lower bounds to reliability (see the literature review for the difference between internal consistency approach and the lower bound approach). Second, when the assumption of independent measurement errors is violated while the assumption of (essential) tau equivalence holds, coefficient alpha has positive bias. Zimmerman et al. (1993) examined variability of coefficient alpha under violation of the two assumptions and found that alpha underestimated reliability under violation of the assumption of (essential) tau equivalence and overestimated reliability under violation of the assumption of uncorrelated.

Komaroff (1996, 1997) studied the interactive effects of simultaneous violations of the two assumptions by simulated true and error scores with known properties. As essential tau -equivalence is violated, alpha decreases. However, as the spread and magnitude of correlations among error scores is increased from zero, the reduction in alpha is attenuated to the point that alpha equals or overestimates the classical reliability coefficient. Coefficient alpha did not differentiate between the two sources of observed inter-item covariances. This finding is also confirmed in my study. Particularly, Models

5, 6, 8, 9, 11 and 12 were designed to study the effects of violating the two assumptions simultaneously.

The results in my study further indicate that violating the assumption of the uncorrelated measurement error affects the bias in coefficient alpha more than violating the assumption of (essential) tau-equivalence. For example, if comparing Model 7 (the (essential) tau equivalence assumption being violated), Model 8 (both assumptions being violated with error correlation of 0.2) and Model 9 (both assumptions being violated with error correlation of 0.5), the bias in alpha changed dramatically from negative 11.68% in model 7 to positive 3.76% in Model 8 and then to positive 17.23% in model 9 when item number was 6 and sample size was 50. The possible reason for these results may lie in the manipulation of error correlation and factor loadings in this study. The factor loadings were set with the average at 0.6, whereas error correlation was set with two levels (0.2 and 0.6). If the average of factor loadings was set at 0.3, violating (essential) tau-equivalent might affect more of the bias in coefficient alpha.

5.1.2 Glb

The glb is the theoretically optimal lower bound to reliability from a mathematical perspective. However, its bias was substantially larger than coefficient alpha's bias under most of the conditions in this study. Besides, its sample values overestimated true reliability under all the conditions in this study. Although the glb is lacking practical meaning due to its large positive sample bias (especially under the conditions of small sample sizes), it has some attractive theoretical properties.

First, its asymptotic property indicates that it is the most accurate and stable reliability coefficient when sample size goes to infinity. Thus, the glb is the greatest lower bound to reliability theoretically (Jackson & Agunwamba, 1977; Sijtsma, 2009; Ten Berge & Sočan, 2004; Woodhouse & Jackson, 1977). However, the glb had large sample bias in this simulation study, which confirmed the findings in Shapiro and Ten Berge (2000)'s study. When the assumption of error independence held as in Models 1, 4, 7 and 10, the magnitude of the bias of the glb was remarkably larger than that of alpha as in Model 1(tau equivalent model) and Model 4 (congeneric model). In Model 7 (the correlated factor model with low factor correlation), the glb had much smaller bias in magnitude than alpha, especially when sample size was 150 or more. As the factor correlation increased to 0.6 as in Model 10, the glb had similar bias as alpha except that the signs of their bias were different: the glb had positive bias while alpha had negative bias. When the assumption of error independence was violated, the glb was also seriously affected. The bias in the glb increased dramatically as error correlation increased when keeping other conditions the same.

Second, the violation of (essential) tau equivalence assumption or unidimensional assumption has little effect on the glb sample estimates of reliability (Sijtsma, 2009). When purely examining the effect of (essential) tau equivalence assumption without the interference of the assumption of error independence, we can compare the glb's bias in

Models 1, 4, 7 and 10. The results suggested that the glb had slightly smaller bias as the tau equivalent model was changed into the congeneric model. As the unidimensional model was transformed into multidimensional model, the bias in the glb increased trivially. As the factor correlation increased from Model 7 to Model 10, the bias in the glb decreased slightly to the level as in the unidimensional models. When (essential) tau equivalence and error independence assumptions were both violated in Models 5, 6, 8, 9, 11 and 12, the bias in the glb increased remarkably due to the correlated errors.

This study examined, for the first time, the effect of violating the error independence assumption on the bias of the glb. Like the other non-SEM estimates of reliability (i.e., coefficient alpha, theta, theta.g and omega), the bias in the glb continued to increase as the error correlation increased from 0 to 0.2 and then to 0.5 when keeping other conditions the same. When the assumption of error independence held, the bias in the glb increased slightly as item number increased from 6 to 12. However, when there were correlated errors in the measurement models, the bias in the glb decreased substantially as item number increased.

5.1.3 Theta and Theta.g

Theta and theta.g are derived from the principal components analysis approach. One advantage of using theta and theta.g is that they are easy to calculate. The information needed for calculating these two coefficients is the eigenvalues of the variance covariance matrix and the determined number of factors. Theoretically, theta is an unbiased estimator of true reliability when the assumption of unidimensionality holds (Armor, 1974) and theta.g is an unbiased estimator of true reliability when factors are orthogonal and items are parallel on each factor (Şimşek & Noyan, 2013). Şimşek and Noyan (2013) studied the performance of theta.g under the conditions that the assumptions of orthogonal dimensions and parallel items on each factor were violated. They found that theta.g outperformed alpha when factors had negative correlation or when factors had positive but small or moderate correlations. However, the robustness of theta and theta.g has not yet been studied under the conditions that assumptions of (essential) tau equivalence and error independence are both violated.

In this study, it was found that when both assumptions held as in Model 1, theta and theta.g had slightly smaller bias than alpha and they all had their percentage bias below 0.5% under all the conditions in Model 1. When tau-equivalence assumption held and error independence assumption was violated as in Models 2 and 3, theta and theta.g had slightly larger bias than alpha when measurement errors were correlated. Theta and theta.g had the same bias when the (essential) tau equivalence assumption held, except under the condition of small item number (e.g., 6) and sample size (e.g., 50). However, the difference between theta and theta.g under that condition was trivial, which was probably due to the inaccurate estimation of factor number under the condition of small item number and sample size.

When simply violating the assumption of (essential) tau-equivalence as in Model 4, theta and theta.g still had their percentage bias below 0.5% under all the conditions. When the assumptions of (essential) tau-equivalence and error independence were violated simultaneously, theta and theta.g had slightly larger bias than alpha as they did under the conditions that only the assumption of error independence was violated in Models 2 and 3. The findings suggest that if a test or subscale is unidimensional, the violation of (essential) tau equivalence barely has effect on theta and theta.g estimates of reliability.

When models became multidimensional, theta and theta.g had exactly opposite biases under the conditions in the correlated factor model. Theta had bias in negative values while theta.g had bias in positive values. In addition, the absolute values of theta's bias increased regardless of item number; however, as sample size went up, the bias of theta.g increased when item number was 6 and decreased when item number was 12. The bias in theta is caused by underestimating the number of factors and the bias in theta.g is mainly caused by factor correlation (Şimşek & Noyan, 2013). As factor correlation increased from 0.2 in Model 7 to 0.6 in model 10, the bias in theta.g increased substantially, but the bias in theta became smaller since the higher the factor correlation was, the closer it was to a unidimensional model. Still, violating the assumption of error independence added more positive bias in theta and theta.g.

5.1.4 Omega

Reliability can be estimated under the exploratory factor analytic framework

(McDonald, 1978) using coefficient ω. Revelle and Zinbarg (2009) compared 13 reliability coefficients by using nine data sets. They found that omega had slightly larger sample bias than the glb under all the conditions where 1) measurement error correlation was 0 or low; 2) item number was small and 3) factor correlation was high or the model is unidimensional.

In this study, it was found that omega performed similarly as the glb when item number was small (i.e., 6 in this study): its sample values overestimated the true reliability under all the conditions. However, the two coefficients still differed in two aspects: (1) the bias of omega decreased more slowly than that of the glb as sample size increased; (2) omega had slightly larger sample bias than the glb except when there was measurement error correlation, especially when error correlation was high. The findings suggested the rate of the bias increasing in omega was smaller than that in the glb as error correlation increased, although the increase of error correlation surged the bias of omega. When item number increased from 6 to 12, the bias in omega decreased remarkably. When the assumption of independent errors held, the bias in omega went down by around 70% in the tau equivalent model and congeneric model, by around 100% in correlated factor model with factor correlation of 0.2, and by around 90% in the correlated factor model with factor correlation of 0.6. When the assumption of independent errors was violated and error correlation started to increase, the decreasing rate of the bias in omega became smaller as item number went up.

Although omega was better estimate of reliability than the glb when item number became larger, it was not the most accurate reliability estimator under most of the conditions. Thus, as the glb, it has less value for practical application.

5.1.5 SEM Estimates of Reliability Using ML and Bayesian Estimation

To be consistent with Yang and Green's (2010) study, the SEM estimates of reliability using ML was named SEM estimates of reliability. The results of SEM estimates in this study were slightly inconsistent with Yang and Green's study. Yang and Green found that the correctly specified SEM estimates of reliability were more accurate than the misspecified SEM estimates of reliability. However, this finding was only confirmed under the unidimensional models when the assumptions of (essential) tau equivalence and error independence were both violated. When the two assumptions both held, SEM.cong and SEM.CE, as the SEM estimates of reliability using over-specified analysis models, were all quite accurate. When sample size was 50, they had even less bias than SEM.tau. When sample size was large (150 or more), the advantage of model data fit could be seen.

When the assumption of unidimensionality was violated, that is, under the correlated factor models in this study, the results were complicated and not completely inconsistent with Yang and Green (2010)'s findings. When there were no correlated errors in the correlated factor models, SEM.CF, the SEM coefficient using correctly specified analysis model, generated the most accurate estimate of reliability regardless of sample size or item number. When there were correlated errors, SEM.CFCE, the SEM coefficient using correctly specified analysis model, generated the most accurate estimate of reliability under the conditions of 12 items, but SEM.CFCE did not demonstrate its accuracy when item number was 6. On the contrary, under the conditions of 6 items, SEM.CFCE was the least accurate estimate of reliability when error correlation was low. This was caused by the estimation problems in complex models with small item number. Therefore, when item number was small and measurement model was complicated, the correctly specified SEM estimates of reliability were not necessarily the most accurate due to estimation issues.

This inconsistency between this study and Yang and Green (2010)'s study are probably caused by the following reasons. 1) The different multidimensional models were used in generating data. In Yang and Green's study, bifactor models were used for generating multidimensional data, whereas correlated factor models were used for generating multidimensional data in this study. Bifactor models assume there is a dominant general factor on which each item has a substantial loading. In that sense, bifactor models bear more resemblance to unidimensional models. In contrast, correlated factor models may have rather different factors with low correlations with each other. 2) The different softwares were used in data simulation and SEM analysis. Yang and Green's simulation study was conducted in SAS environment and this study was in R environment. Thus the results may have some minute differences in rounding and sampling.

BSEM approach to reliability estimation resembles SEM estimates of reliability except that they have different estimation methods. The SEM estimates of reliability using Bayesian estimation were simply named BSEM estimates of reliability in this section. The BSEM estimates of reliability in this study tended to be lower than the corresponding estimates using ML. Thus, if the SEM estimates of reliability using ML were higher than the true reliability, the BSEM estimates of reliability could reduce the positive bias. However, if the SEM estimates of reliability using ML were lower than the true reliability, the BSEM estimates of reliability would enlarger the magnitude of the negative bias. Therefore, using the Bayesian estimation method adopted in this study would not necessarily increase the accuracy of the corresponding reliability estimates. Nevertheless, the Bayesian estimation method used in this study solved estimation problems in SEM estimates of reliability. In addition, the correctly specified BSEM estimates of reliability (i.e., BSEM.CFCE in Models 8 and 9) were quite accurate when sample size was large.

Although the Bayesian estimation method used in this study did not eradicate the bias in SEM estimates of reliability under the conditions where estimation problems occurred, it reduced the standard errors of these SEM reliability coefficients to a great extent. The small standard errors suggested that there were no or fewer estimation problems when using BSEM estimates of reliability, which were confirmed by the results in stage two. Bayesian estimation was, for the first time, used in reliability estimation, and thus the findings in this study were of unique meaning.

5.2 Summary of Answers to Research Questions

Question 1: Which reliability coefficient generates the best estimate of reliability considering the manipulated conditions?

In order to select the best estimate of reliability, RMSE was screened first given that it combined the information of accuracy and stability, that is, only the coefficients that were both accurate and stable could have a small RMSE value. If two or more coefficients had similar RMSEs, their biases were further examined to determine the best performing reliability coefficient.

Under the conditions that both (essential) tau equivalence and error independence assumptions held as in Model 1, all the coefficients had small RMSEs ranging from 0.01 to 0.08. Thus, the biases of these coefficients were further examined. In fact, except the glb and omega, all the coefficients had trivial biases around or below 1%. All SEM estimates of reliability performed similarly as coefficient alpha, theta, and theta.g even if the analysis model was not correctly specified (e.g., SEM.cong and SEM.CE). When there were correlated errors as in Models 2 and 3, SEM.CE (CSM) consistently had the smallest RMSEs regardless of sample size or item number and hence the best estimate of reliability.

Under the conditions that both congeneric models and error independence held as in Model 4, SEM.cong (CSM) and SEM.tau (USM), yielded the same smallest RMSEs under all the conditions. However, SEM.tau was slightly less accurate than SEM.cong, and thus SEM.cong (CSM) was the best estimate of reliability under the conditions in the congeneric model with independent errors. When there were correlated errors as in Models 5 and 6, SEM.CE (CSM) had the smallest RMSEs under all the conditions except that when sample size was small (i.e., 50). Under the conditions of 50 observations, SEM.CE was not as stable as other coefficients although it was the least biased. Given that its RMSE was slightly larger than other coefficients' RMSEs but its bias was much smaller than others' biases, SEM.CE (CSM) was still regarded as the best estimate of reliability.

Under the correlated factor models and when the assumption of error independence held as in Models 7 and 10, SEM.CF (CSM) generated the smallest RMSEs and the most accurate estimates of reliability except for one condition that the factor correlation was 0.2, item number was 6 and sample size was 50. Under that condition, SEM.CF was not as stable as other coefficients and less accurate than theta.g and SEM.CE, and theta.g had the smallest RMSE. When there were correlated errors but error correlation was low (0.2) as in Models 8 and 11, alpha had the smallest RMSEs when item number was 6. When item number was 12, SEM.CFCE (CSM) had slightly larger RMSEs occasionally but it was the least biased. Thus, alpha was the best estimate of reliability when item number was 6 and SEM.CFCE was the best estimate of reliability when item number was 12 for the correlated factor models with low error correlation. When error correlation was moderate (0.5) as in Models 9 and 12, alpha, theta and SEM.tau all had very similar RMSEs and biases when item number was 6. Although their RMSEs were smallest, their estimates of reliability were far from accurate, especially under the conditions that factor correlation was low (0.2) as in Model 9. Under those conditions, the Bayesian estimation method provided much more accurate estimates of reliability when the analysis model was correctly specified. When item number was 12, SEM.CFCE (CSM) had the smallest RMSEs and was least biased regardless of sample size.

In short, correctly specified SEM estimates of reliability performed better than other coefficients unless estimation issues occurred. When estimation problems exist, Bayesian estimation can be used to obtain a comparatively more accurate and stable estimate of reliability when the analysis model is correctly specified.

Question 2: When the assumption of unidimensionality is violated, will unidimensional SEM estimates of reliability approximate the correctly specified SEM estimates of reliability?

When the assumption of unidimensionality was violated as in the correlated factor models, the results of the unidimensional SEM estimates of reliability (i.e., SEM.tau, SEM.cong and SEM.CE) were very different from those of the correctly specified SEM estimates of reliability. In Models 7 and 10 where errors were independent, SEM.CF (CSM) had much smaller bias than SEM.tau, SEM.cong and SEM.CE except under the condition that item number was 6 and sample size was 50. Under that condition, SEM.CE had the least bias. However, SEM.CE and SEM.cong were much less stable than SEM.CF although SEM.tau was slightly more stable than SEM.CF when sample size was small. When sample size was large, these coefficients were all stable unless estimation problems occurred as in SEM.CE.

For data generated from the model with both low factor correlation and error correlation (Model 8), SEM.tau, SEM.cong and SEM.CE were all less biased than SEM.CFCE (CSM) when item number was 6. Under the conditions of 6 items, SEM.cong demonstrated highest accuracy; its stability was also better than SEM.CE and SEM.CF, although no better than SEM.tau when sample size was 50. When item number was 12, SEM.tau, SEM.cong and SEM.CE were more biased than SEM.CFCE regardless of sample size.

For data generated from the model with moderate factor correlation and low correlated errors (Model 11), SEM.tau, SEM.cong and SEM.CE were less biased than SEM.CFCE (CSM) only under the conditions when item number was 6 and sample size was 150 or less. Under those conditions, SEM.CE demonstrated highest accuracy, but it was not very stable when sample size was 50. When item number was 12, SEM.tau, SEM.cong and SEM.CE were more biased than SEM.CFCE regardless of sample size.

For data generated from the model with low factor correlation and moderate error correlation (Model 9), SEM.tau, SEM.cong and SEM.CE were all less biased than SEM.CFCE (CSM) when item number was 6. Under the conditions of 6 items, SEM.CE demonstrated the highest accuracy. However, SEM.CE and SEM.cong were not nearly as stable as SEM.tau. When item number was 12, SEM.tau, SEM.cong and SEM.CE were still more biased than SEM.CFCE regardless of sample size.

For data generated from the model with both moderate factor correlation and error correlation (Model 12), SEM.CFCE (CSM) had smaller bias than SEM.tau, SEM.cong and SEM.CE except under the condition that item number was 6 and sample size was 150 or less. Under those conditions, SEM.CE had the least bias. However, the stability of SEM.CE was not as good as those of SEM.tau and SEM.cong, although it was better than that of SEM.CFCE.

In short, the unidimensional SEM estimates of reliability performed very differently from the correctly specified SEM estimate of reliability under the violation of unidimensionality, and none of the unidimensional SEM estimates of reliability could stand out with both satisfying accuracy and stability.

Question 3: If the analysis model is an over-specified model, will its estimates of reliability approximate those using the correctly specified analysis model?

Using an over-specified model as the analysis model could not generate estimates of reliability that were exactly the same as the correctly specified SEM estimates of reliability. However, the estimates of reliability by an over-specified model turned out to be very close to the correctly specified SEM estimates of reliability and even had less bias under some particular conditions unless estimation problems occurred.

Under the conditions of unidimensional models with independent errors, the analysis model for SEM.CE was an over-specified model for the data generated in Models 1 and 4. Its estimates of reliability were very close to SEM.tau (CSM) in Model 1(tau-equivalent model with independent errors) and SEM.cong (CSM) in Model 4 (congeneric model with independent errors). When sample size was 50 and item number was 12, it even had less bias than SEM.tau in Model 1 and SEM.cong in Model 4. Besides, the over-specified SEM estimates of reliability were as stable as the correctly specified SEM estimates of reliability.

For multidimensional models with independent errors (Models 7 and 10), SEM.CFCE was the SEM reliability coefficient using an over-specified analysis model. It had the estimates of reliability similar as those of SEM.CF (CSM) when item number was 12. However, it had estimation problems when item number was 6 and hence was seriously biased and instable.

In short, if there were no estimation problems, the over-specified SEM estimates of reliability could be used as alternatives to the correctly specified SEM estimates of reliability.

Question 4: Can Bayesian estimation with non-informative priors overcome estimation problems in SEM estimates of reliability using ML? If yes, will the estimates of reliability using Bayesian estimation with non-informative priors be more accurate and stable than those using ML?

The Bayesian estimation method used in this study (i.e., Bayesian estimation with non-informative priors) could overcome estimation problems in SEM estimates of reliability. In fact, no estimation problems were identified when using Bayesian estimation for the analysis conducted in the second stage of this study. However, BSEM estimates of reliability performed very differently from the corresponding SEM estimates of reliability using ML estimation.

If the analysis model was correctly specified, ML yielded the best SEM estimates of reliability unless estimation problems occurred. When estimation problems occurred, using Bayesian estimation with non-informative priors could reduce the bias in SEM estimates of reliability. However, the degree of the reduction in bias was different from one model to another. For Model 8 (the correlated factor model with low factor correlation and error correlation), BSEM.CFCE (CSM) had a satisfactory bias and standard error only when sample size was 500. For Model 9 (the correlated factor model with low factor correlation and moderate error correlation), BSEM.CFCE (CSM) had very small bias and moderate standard error when sample size was 150 or more.

For the misspecified SEM estimates of reliability, if the misspecified SEM reliability coefficient had very accurate estimates of reliability (e.g., SEM.cong in Model 8 under conditions of 6 items), using the Bayesian estimation method would neither improve nor worsen the accuracy of its estimates of reliability, especially when sample size was large. If the misspecified SEM reliability had inaccurate estimates of reliability (e.g., SEM.CE in Models 7, 8, and 9), the Bayesian estimation method in fact worsen the accuracy of SEM estimates of reliability instead of improving it.

5.3 Limitations and Suggestions for Future Research

As discussed afore, choosing different generation models or analysis models may result in different findings in SEM estimates of reliability. Thus, it is necessary to point out the differences in model design and the conditions under which new results are inconsistent with those in previous studies. In this study, not all the measurement models used in previous studies were replicated. If time and other conditions allow, it would be helpful to replicate previous studies and then compare the results with this study so that more convincing conclusions can be drawn.

Furthermore, future research can focus on solving the estimation problems as well as maintaining the accuracy of SEM estimates of reliability. The correctly specified SEM estimates of reliability did perform the best if parameters can be properly estimated. This study only identified the method for solving the estimation problems (i.e., using the

Bayesian estimation method) but could not maintain the absolute accuracy of SEM estimates of reliability. More types of priors (e.g., informative priors and weak informative priors) need to be tried and compared when using Bayesian estimation in future study in order to maintain the accuracy of SEM estimates of reliability as well as solving the estimation problems.

Last, more types of data (e.g., binary or categorical data) could be examined to see whether the same conclusions can be obtained. In this study, only continuous data was considered because in this way the results could be generalized to large psychological measures (e.g., Manual for the Adaptive Learning Scales) with various scales and subscales. That is, items in this study could be generalized to test parts at different levels (e.g., scales or subscales or some composite score of several items). Many published psychological measures transform raw scores into standardized scores like *z* or *t* scores. That is, in practice, continuous data like *z* or *t* scores are normally used for conducting psychometric analysis on these scales. However, it is still meaningful to compare the results in this study with the corresponding results for categorical data and examine whether the same conclusions can be obtained.

REFERENCES

Becker, G. (2000). How important is transient error in estimating reliability? Going beyond simulation studies. *Psychological Methods, 5,* 370-379.

Bentler, P.M. (1972). A lower-bound method for the dimension-free measurement of internal consistency. *Social Science Research, 1,* 343-357.

Bentler, P.M., & Woodward, J.A. (1980). Inequalities among lower-bounds to reliability: With applications to test construction and factor analysis. *Psychometrika, 45,* 249-267.

Bentler, P.A. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika, 74*(1), 137–143.

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

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology, 3*, 296–322.

Brunner, M., & Sub, H.M. (2005). Analyzing the reliability of multidimensional measures: an example from intelligence research. *Educational and Psychological Measrement, 65(2),* 227-240.

Callendar, J. C., & Osburn, H. G. (1979). An empirical comparison of coefficient alpha, Guttman's lambda-2, and msplit maximized split-half reliability estimates. *Journal of Educational Measurement, 16,* 89-99.

Chen F.F, West S.G., & Sousa K.H. (2006) A comparison of bifactor and second-order models of quality-of-life. *Multivariate Behavioral Research. 41*, 189–225.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98-104.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.

Cronbach, L. J. (1988) Internal consistency of tests. *Psychometrika, 53,* 63-70.

Cronbach, L. J., Gleser, G. C., Harinder, N., & Rajaratnam, N. (1972). *The Dependability of Behavioral Measurement: Theory of Generalizability for Scores and Profiles.* New York: Wiley.

Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. (1963). Theory of generalizability: A liberalization of reliability theory. *The British Journal of Statistical Psychology 16*: 137-163.

De Leeuw, J. (1983). Models and methods for the analysis of correlation coefficients. *Journal of Econometrics*, 22(1–2),113–137.

Feldt, L. S., & Brennan, R. L. (1989). Reliability. In: Linn, R. L., ed., *Educational Measurement.* 3rd ed. New York: Macmillan, pp. 105–146.

Fleishman, J., & Benson, J. (1987). Using LISREL to evaluate measurement models and scale reliability. *Educational and Psychological Measurement, 47,* 925-939.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis.* London: Chapman and Hall.

Gilmer, J. S., & Feldt, L. S. (1983). Reliability estimations for a test with parts of unknown lengths. *Psychometrika, 48,* 99-111.

Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability: What they are and how to use them. *Educational and Psychological Measurement, 66,* 930-944.

Green, S. B. (2003). A Coefficient Alpha for Test–Retest Data. *Psychological Methods, 8 (1),* 88-101.

Green, S. B., & Hershberger, S. L. (2000). Correlated errors in true score models and their effect on coefficient alpha. *Structural Equation Modeling*, *7(2),* 251-270.

Green, S.B., Lissitz, R.W., & Mulaik, S.A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement, 37,* 827-838.

Green, S. B., & Yang, Y. (2009). Commentary on coefficient alpha: A cautionary tale. *Psychometrika, 74*, 121-135.

Gulliksen, H. (1950). *Theory of mental tests.* New York: Wiley.

Gustafsson, J., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research, 28,* 407-434.

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika, 10*, 255-282.

Haertel, E. H. (2006). Reliability. In Brennan, R. L. (Ed.). *Educational Measurement.*

(4th ed.). National Council on Measurement in Education. American Council on Education. Westport, CT : Praeger Publishers.

Hattie, J. A. (1984a). An empirical study of various indices for determining unidimensionality. *Multivariate Behavioral Research, 19,* 49-78.

Hattie, J. A. (1995). Methodology review: assessing unidimensionality. *Applied Psychological Measurement, 9,* 139-164.

Holzinger, K. J., & Swineford, F. (1937). The Bi-Factor Method. *Psychometrika, 2,* 41–54.

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods & Research, 26,* 329-367.

Horst, P. (1953). Correcting the Kuder-Richardson reliability for dispersion of item difficulties. *Psychological Bulletin, 50,* 371-374.

Hox, J., Van de Schoot, R., & Matthijsse, S. (2012). How few countries will do? Comparative survey analysis from a Bayesian perspective. *Survey Research Methods, 6 (2),* 87-93.

Hoyt, C. (1941). Test reliability estimated by analysis of variance. *Psychometrika, 6,* 153-160.

Ibrahim, J. G. & Laud. P. W. (1991). On Bayesian analysis of general linear models using Jeffreys'prior. *Journal of the American Statistical Association, 86,* 981-986.

Jackson, P. H., & Agunwamba, C. C. (1977). Lower bounds for the reliability of the total score on a test composed of nonhomogeneous items. I: Algebraic lower bounds. *Psychometrika, 42,* 567-578.

Jackson, R. W. B. (1939). Reliability of mental tests. *British Journal of Psychology, 29,* 267-287.

Jöreskog, K.G. (1971). Statistical analysis of sets of congeneric test. *Psychometrika, 36*, 109-133.

Komaroff, E. (1997). Effect of simultaneous violations of essential t -equivalence and uncorrelated error on coefficient *a* . *Applied Psychological Measurement, 27(4),* 337-348.

Kristof, W. (1963). The statistical theory of stepped-up reliability coefficients when a test has been divided into several equivalent parts. *Psychometrika, 28,* 221-238.

Kristof, W. (1974). Estimation of reliability and true score variance from a split of a test into three arbitrary parts. *Psychometrika, 39,* 491–499.

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

Lee, S. (2007). *Structural equation modeling: a Bayesian approach.* Chicester, UK: Wiley.

Loevinger, J. (1948). The technique of homogeneous tests. *Psychological Bulletin, 45,* 507-529.

Lord, F. M., & Novick, M.R. (1968). *Statistical theories of mental test scores.* Menlo Park, CA: Addison-Wesley.

Maxwell, A. E. (1968). The effect of correlated errors on estimates of reliability coefficients. *Educational and Psychological Measurement, 28,* 803–811.

McDonald, R. P. (1978). Generalizability in factorable domains: "domain validity and generalizability": 1. *Educational and Psychological Measurement, 38* (1), 75-79.

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

McDonald, R. P. (1985). *Factor Analysis and Related Methods.* HillSEale, NJ: Lawrence Erlbaum Associates.

McDonald, R. P. (1999). *Test theory: A unified treatment.* HillSEale: Erlbaum.

Miller, M.B. (1995). Coefficient alpha: a basic introduction form the perspectives of classical test theory and structural equation modeling. *Structural Equation Modeling, 2,* 255-273.

Muthén, L.K., & Muthén, B.O. (1998-2014). *Mplus User's Guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Novick, M. R., & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika, 32,* 1-13.

Novick, M. R., Jackson, P. H., & Thayer, D. T. (1971). Bayesian inference and the classical test theory model: Reliability and true scores. *Psychometrika, 36(3),* 261-288.

Osburn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods, 5*, 343-355.

Raju, N. S. (1977). A generalization of coefficient alpha. *Psychometrika, 42,* 549-565.

Raykov, T. (1997a). Estimation of composite reliability for congeneric measures. Applied Psychological Measurement, 22, 173–184.

Raykov, T. (1997b). Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalencewith fixed congeneric components.Multivariate Behavioral Research, 32, 329–353.

Raykov, T. (1998). Coefficient alpha and composite reliability with interrelated nonhomogeneous items. *Applied Psychological Measurement, 22*(4), 375-385.

Raykov, T. (2001). Bias of Cronbach's coefficient alpha for congeneric measures with correlated errors. Applied Psychological Measurement, 25, 69–76.

Raykov, T., & Shrout, R.E. (2002). Reliability of scales with general structure: point and interval estimation using a structural equation modeling approach. *Structural Equation Modeling, 9(2),* 195-212.

R Development Core Team (2011). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria (ISBN 3-900051-07-0).

Reise S.P., Moore T.M., & Haviland M.G. (2010). Bifactor models and rotations:

Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment, 92,* 544–559.

Revelle, W., & Zinbarg, R.E. (2009). Coefficient alpha, beta, and the glb: comments on Sijtsma. *Psychometrika*, 74(1), 145-154.

Revelle, W. (2013). *psych: Procedures for Personality and Psychological Research* (R package version 1.3.2.).

Rindskopf, D., & Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behavioral Research, 23,* 51–67.

Rozeboom, W. W. (1966). *Foundations of the theory of prediction.* Homewood, IL: Dorsey.

Rozeboom,W.W. (1989). The reliability of a linear composite of nonequivalent subtests. *Applied Psychological Measurement, 13,* 277–283.

Rulon, P. J. (1939). A simplified procedure for determining the reliability of a test by split halves. *Harvard Educational Review*, *9*, 99-103.

Schmid, J.J., & Leiman, J.M. (1957). The development of hierarchical factor solutions. *Psychometrika*, 22(1), 83–90.

Schmitt, N. (1996). Uses and Abuses of Coefficient Alpha. *Psychological Assessment, 8,* 350-353.

Shapiro, A., & Ten Berge, J. M. F. (2000). The asymptotic bias of minimum trace factor analysis, with applications to the greatest lower bound to reliability. *Psychometrika,* *65*, 413-425.

Shrout, P. E., & Joseph, L. F. (1979). Intraclass Correlations: Using in Assessing Rater Reliability. *Psychologocal Bulletin, 86* (2), 420-428.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74,* 107-120.

Spearman, C. (1904). The proof and measurement of association between two things. A*merican Journal of Psychology, 15,* 72-101.

Spearman,C. (1910). Correlation calculated from faulty data. *British Journal of Psychology, 3*, 271–295.

Tang, W., Cui, Y., & Babenko, O. (2014). Internal consistency: Do we really know what it is and how to assess it? *Journal of Psychology and Behavioral Science, 2*, pp. 205-220.

Tang, W., & Cui, Y. (2014). *Does the greatest lower bound outperform coefficient alpha under violation of essential tau-equivalence?* Poster presented at the annual meeting of American Educational Research Association, Philadelphia, PA.

Tang, W., & Cui, Y. (2012). *Comparing four lower bounds of reliability under violation of the assumption of unidimensionality.* Paper presented at Annual Meeting of American Educational Research Association, Vancouver, BC.

Ten Berge, J. M. F., & Kiers, H.A.L. (1991). A numerical approach to the exact and the approximate minimum rank of a covariance matrix. *Psychometrika, 56,* 309-315.

Ten Berge, J. M. F., & Kiers, H.A.L. (2003). *The minimum rank factor analysis program MRFA* (Internal report). Department of Psychology, University of Groningen, The Netherlands.

Ten Berge, J.M.F., Snijders, T.A.B., & Zegers, F.E. (1981). Computational aspects of the greatest lower bound to reliability and constrained minimum trace factor analysis. *Psychometrika, 46,* 357-366.

Ten Berge, J. M. F., & Sočan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika, 69,* 613-625.

Ten Berge, J. M. F., & Zegers, F. E. (1978). A series of lower bounds to the reliability of a test. *Psychometrika, 43 (4),* 575-579.

Thorndike, R. L. (1951). Reliability. In E. F. Lindquist (ed.), *Educational Measurement* (pp. 560-620). Washington, DC: American Council on Education.

Traub, R. (1997). Classical Test Theory in Historical Perspective. *Educational Measurement: Issues and Practice, 16 (4),* 8-14.

Tucker, L. (1940). The role of correlated factors in factor analysis. *Psychometrika, 5,* 141-152.

Wackerly, D., Mendenhall, W., & Scheaffer, R. L. (2008). *Mathematical Statistics with Applications* (7th ed.). Belmont, CA, USA: Thomson Higher Education.

Wherry, R. J. (1959). Hierarchical factor solutions without rotation. *Psychometrika, 24,* 45-51.

Wilkinson, L., & American Psychological Association Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist. 54,* 594-604.

Woodhouse, B., & Jackson, P.H. (1977). Lower bounds for the reliability of the total score on a test composed of non-homogeneous items: II. A search procedure to locate the greatest lower bound. *Psychometrika, 42,* 579-591.

Yang, Y., & Green, S. B. (2010). A Note on Structural Equation Modeling Estimates of Reliability. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*, 66-81.

Yang Y., & Green S. B. (2011). Coefficient alpha: A reliability coefficient for the 21st century? *Journal of Psychoeducational Assessment, 29,* 377-392.

Yung, Y. F., Thissen, D., & McLeod, L. D. (1999). On the Relationship between the Higher-Order Factor Model and the Hierarchical Factor Model. *Psychometrika*, *64,* 113–128.

Zimmerman, D. W., Zumbo, B. D., & Lalonde, C. (1993). Coefficient alpha as an estimate of test reliability under violation of two assumptions. *Educational and Psychological Measurement, 53,* 33-49.