THE EFFECTS OF MISSPECIFICATION TYPE AND NUISANCE VARIABLES ON THE BEHAVIORS OF POPULATION FIT INDICES USED IN STRUCTURAL EQUATION MODELING

by

Claudia Mahler

B.A., The University of British Columbia, 2011

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS in

THE FACULTY OF GRADUATE STUDIES

(Psychology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

July 2011

© Claudia Mahler, 2011
Abstract

The present study examined the performance of population fit indices used in structural equation modeling. Index performances were evaluated in multiple modeling situations that involved misspecification due to either omitted error covariances or to an incorrectly modeled latent structure. Additional nuisance parameters, including loading size, factor correlation size, model size, and model balance, were manipulated to determine which indices’ behaviors were influenced by changes in modeling situations over and above changes in the size and severity of misspecification. The study revealed that certain indices (CFI, NNFI) are more appropriate to use when models involve latent misspecification, while other indices (RMSEA, GFI, SRMR) are more appropriate in situations where models involve misspecification due to omitted error covariances. It was found that the performances of all indices were affected to some extent by additional nuisance parameters. In particular, higher loading sizes led to increased sensitivity to misspecification and model size affected index behavior differently depending on the source of the misspecification.
Table of Contents

Abstract.............................................................................................................................. ii
Table of Contents .............................................................................................................. iii
List of Tables .................................................................................................................... vii
List of Figures ................................................................................................................... viii
List of Illustrations .......................................................................................................... x
Acknowledgements ........................................................................................................... xi
Dedication .......................................................................................................................... xii

Chapter 1: Introduction .................................................................................................. 1

1.1 Structural Equation Modeling .................................................................................. 1
1.2 Goals of the Present Research ................................................................................ 3
1.3 Thesis Structure ........................................................................................................ 5
1.4 Estimation and Model Fit ....................................................................................... 6
  1.4.1 Estimation .......................................................................................................... 7
  1.4.2 Assessing model fit .......................................................................................... 8
    1.4.2.1 The chi-square test statistic ..................................................................... 9
    1.4.2.2 Fit indices .................................................................................................. 10
1.5 Literature Review .................................................................................................... 11
  1.5.1 Assessing model fit .......................................................................................... 14
    1.5.2.1 Sample size .............................................................................................. 14
    1.5.2.2 Normality .................................................................................................. 15
  1.5.2 Fit indices .......................................................................................................... 15
1.5.2.1 Sample size ............................................................................................................. 15
1.5.2.1.1 Class 1............................................................................................................. 15
1.5.2.1.2 Class 2............................................................................................................. 16
1.5.2.1.3 Class 3............................................................................................................. 17
1.5.2.1.4 Class 4............................................................................................................. 18
1.5.2.1.5 Class 5............................................................................................................. 18
1.5.2.1.6 Class 6............................................................................................................. 19
1.5.2.1.7 Class 7............................................................................................................. 20
1.5.2.2 Omitted indices ...................................................................................................... 20
1.5.3 Concerns regarding fit indices .................................................................................. 21
1.5.3.1 Theoretical and methodological concerns .......................................................... 21
1.5.3.1.1 Choosing an index.......................................................................................... 21
1.5.3.1.2 Cutoff values.................................................................................................. 22
1.5.3.2 Sample size, estimation method, and model features .......................................... 26
1.5.3.2.1 Sample size .................................................................................................. 26
1.5.3.2.2 Estimation method ....................................................................................... 27
1.5.3.2.3 Model size.................................................................................................... 29
1.5.3.2.4 Parameter values ........................................................................................ 34
1.5.3.2.5 Source of misspecification ......................................................................... 36
1.5.3.2 Summary ............................................................................................................ 41
1.6 Methods ...................................................................................................................... 42

Chapter 2: One-Factor Models ......................................................................................... 46

2.1 Effects of Loading Size .............................................................................................. 46
2.1.1 Homogeneous loadings................................................................. 46
2.1.2 Heterogeneous loadings ................................................................. 50
2.1.3 Omitted error correlation ................................................................. 53
2.1.4 Continuous loadings ......................................................................... 56
2.1.5 Increasing number of large loadings......................................................... 61
2.2 Effects of an Increasing Degree of Misspecification .......................................... 65
   2.2.1 Second omitted error covariance ......................................................... 65
   2.2.2 Increasing number of omitted error covariances ......................................... 69
2.3 Model Size ............................................................................................. 73
2.4 Summary and Discussion ......................................................................... 77

Chapter 3: Results for Two-Factor Models......................................................... 84
3.1 Parameter Size .......................................................................................... 85
   3.1.1 Location of omitted error covariance ...................................................... 85
   3.1.2 Factor correlation .................................................................................. 88
3.2 Second Omitted Error Covariance ............................................................... 92
3.3 Model Size ............................................................................................... 107
3.4 Model Balance ........................................................................................... 112
   3.4.1 Location of misspecification ................................................................. 117
   3.4.2 Factor size and misspecification ............................................................. 120
3.5 Summary and Discussion ......................................................................... 128

Chapter 4: Results for Misspecified Latent Structure Models............................... 134
4.1 One-Factor Model Fit to Two- and Three-Factor Data ..................................... 135
   4.1.1 Model size .......................................................................................... 140
4.1.2 Model balance ............................................................................................................. 150

4.2 Increasing Number of Latent Factors ......................................................................... 154
   4.2.1 Indicator to factor ratio ......................................................................................... 164

4.3 Summary and Discussion ............................................................................................. 174

Chapter 5: Conclusion and Overall Discussion ................................................................. 181

5.1 Effects of Model Components .................................................................................... 181
   5.1.1 Loading size ........................................................................................................... 181
   5.1.2 Factor correlation size ............................................................................................ 182
   5.1.3 Model size ............................................................................................................. 183
   5.1.4 Model balance ....................................................................................................... 184

5.2 Index-Specific Effects .................................................................................................. 185
   5.2.1 CFI and NNFI ........................................................................................................ 185
   5.2.2 GFI and AGFI ....................................................................................................... 188
   5.2.3 RMSEA and gamma ............................................................................................. 189
   5.2.4 SRMR .................................................................................................................... 190

5.3 Concluding Remarks .................................................................................................. 192

Bibliography ...................................................................................................................... 195
List of Tables

Table 1  Names, Sample Definitions, and Population Definitions of Fit Indices ............... 12
# List of Figures

| Figure 1 | Index values vs. omitted error covariance with homogeneous loadings | 48 |
| Figure 2 | Index values vs. omitted error covariance with heterogeneous loadings | 51 |
| Figure 3 | Index values vs. omitted error correlation | 54 |
| Figure 4 | Index values vs. loading size | 58 |
| Figure 5 | Index values vs. number of high loadings | 63 |
| Figure 6 | Index values vs. second omitted error covariance | 66 |
| Figure 7 | Index values vs. number of omitted error covariances | 70 |
| Figure 8 | Index values vs. number of indicators | 74 |
| Figure 9 | Index values vs. omitted error covariance for a two-factor model | 86 |
| Figure 10 | Index values vs. omitted error covariance for different factor correlation sizes | 89 |
| Figure 11 | Index values vs. second omitted error covariance (I) | 98 |
| Figure 12 | Index values vs. second omitted error covariance (II) | 101 |
| Figure 13 | Index values vs. second omitted error covariance (III) | 105 |
| Figure 14 | Index values vs. number of indicators for a two-factor model | 108 |
| Figure 15 | Index values vs. omitted error covariance for imbalanced models (I) | 114 |
| Figure 16 | Index values vs. omitted error covariance for imbalanced models (II) | 118 |
| Figure 17 | Index values vs. number of indicators for imbalanced models | 121 |
| Figure 18 | Index values vs. size of factor containing an omitted error covariance | 125 |
| Figure 19 | Index values vs. factor correlation for a latent structure misspecification | 136 |
| Figure 20 | Index values vs. number of indicators for a latent structure misspecification | 142 |
| Figure 21 | Index values vs. number of indicators for different factor correlations | 148 |
| Figure 22 | Index values vs. factor correlation for imbalanced models | 152 |
Figure 23  Index values vs. number of factors for a latent structure misspecification ....... 156
Figure 24  Index values vs. number of factors when inter-item correlation is constant ...... 161
Figure 25  Index values vs. number of factors when $p:k$ is held constant ...................... 166
Figure 26  Index values vs. number of factors for different sizes of $p$ .............................. 170
List of Illustrations

Illustration 1  Model diagram for Figure 11 ................................................................. 94
Illustration 2  Model diagram for Figure 12 ................................................................. 95
Illustration 3  Model diagram for Figure 13 ................................................................. 96
Acknowledgements

I offer my thanks to Dr. Savalei for introducing me to structural equation modeling and
guiding me on my path to a better understanding of the field. I also owe thanks to Dr.
Biesanz and Dr. Zumbo for serving on my committee and taking time to help me along the
way. Additional thanks go out to my parents, who have helped me tremendously throughout
my educational experience and without whose support I would have never made it this far.
Dedication

I dedicate this work to my mother, who has been a constant source of love, support, and inspiration throughout my life. I owe my deepest gratitude to her for everything she’s done for me over the past 23 years.
Chapter 1: Introduction

1.1 Structural Equation Modeling

Structural equation modeling (SEM) is a statistical modeling technique that lets researchers construct and test causal connections amongst variables. It allows researchers to express the covariance (or correlation) between two variables as a function of the parameters of a proposed model, typically a covariance structure model. SEM can be used as a tool for representing relationships between variables. These relationships can occur between latent variables, between observed variables, or between latent and observed variables and can be expressed as a series of structural equations in which model parameters are estimated. The restrictions imposed by these estimated parameters can then be applied to a given sample to determine whether or not the model suggesting these parameters holds in the population.

A structural equation model typically comprises two components: a structural model and a measurement model. In confirmatory factor analysis (CFA) models, the type of model focused on in the present study, the structural model describes the relationships among the $k$ latent variables (McDonald & Ho, 2002) and the measurement model represents the set of $p$ observable variables as indicators of the set of $k$ latent variables. The combination of these two components serves as a model of the causal connections between latent variables and between latent variables and their relevant indicator variables. The causal connections are functions of the model parameters. The model itself is a theory-based representation of how the variables contained within it relate to each other in reality.

The ability to test these theoretically derived models against empirical data is one of the main reasons SEM is growing in popularity amongst social scientists. Another reason behind its popularity is the fact that it allows for the modeling of latent variables and
correlated errors. Psychologists often require a way of relating observable variables (such as scores on an anxiety measure) to related latent factors (such as general anxiety).

The primary goal of SEM is to examine how well the causal inferences contained in a model proposed by a researcher actually match relationships found in the population. Traditionally, the chi-square test statistic has been used as the sole criterion by which model fit is judged. However, notable problems arise with the statistic’s performance in both large and small samples, under different estimation methods, and in cases where the assumptions that underlie its use are violated. In response to these problems, a multitude of goodness of fit indices have been developed to aid researchers in accurately assessing model fit in situations where the chi-square may prove inaccurate.

Though the majority of indices have been developed to overcome the problems associated with the chi-square, the use of these indices has not come without its own set of problems. One particular problem faced by researchers is the sheer number of indices that have been developed. Popular SEM programs such as LISREL and EQS can print upwards of seven or eight indices in addition to the chi-square test statistic. The availability of so many ways to evaluate fit may make it difficult for researchers to know which indices to report (Bollen & Long, 1992).

An additional problem stems from the fact that not all indices have been developed under the same theoretical rationales. For example, some indices have been developed to penalize for model complexity and thus adjust for the size of the model being evaluated. Other indices include no such adjustment. Therefore, indices may perform differently under different model or misspecification types. Differences in performance could lead to
conflicting conclusions regarding whether the fit of a model is appropriate (Fan, Thompson, & Wang, 1999).

There have been multiple demonstrations in the literature (e.g., Marsh, Hau, & Wen, 2004; Beauducel & Wittmann, 2005; Yuan, 2005) that indicate popular “cutoff values” used as thresholds of model acceptance/rejection may not be generalizable across all situations. Many other studies (e.g., Anderson & Gerbing 1988; Chau & Hocevar, 1995; Fan & Sivo, 2007; Hu & Bentler, 1999; Kenny & McCoach, 2003; Marsh & Balla, 1994) have revealed that factors such as sample size, model complexity, type of model misspecification, and estimation procedure all play a role in index performance.

In short, while attempts have been made to develop indices that perform accurately in situations where the chi-square does not, there still exist problems with methods of evaluating model fit, particularly when trying to do so across different model and misspecification types.

1.2 Goals of the Present Research

The aim of the current research is to investigate the performance of popular fit indices in various model and misspecification scenarios as well as to determine whether commonly-applied cutoff values can be used across different model situations. Examined are the performances of the comparative fit index (CFI), non-normed fit index (NNFI), goodness of fit index (GFI), adjusted goodness of fit index (AGFI), the root mean square error of approximation (RMSEA), gamma hat (gamma), and the standardized root mean square residual (SRMR) under a variety of modeling conditions. Within these conditions, the effects of different sources of misspecification and different model components (hereby referred to as “nuisance variables”) on index values are inspected. All fit indices are studied in the
population to eliminate any variability in index performance that may result from sampling fluctuation. In conducting this research, the goal is to address the following three questions:

1. To what extent is the relationship between the amount of model misspecification, as defined by the size of the omitted parameter(s), and index value moderated by other nuisance variables (such as loading size)?

2. Does the current research support prior findings suggesting that uniform cutoff values for indices may not be appropriate?

3. Can guidelines for the use of different indices under different model conditions be developed?

By addressing the first question, the hope is to provide a clearer understanding of both index performance and the effects of various nuisance variables on index behavior. Evaluation of index performance will be achieved by comparing index behavior to what might be expected in a given modeling situation.

Also of interest are additional nuisance variables that may affect the relationship between index value and the size of misspecification. In the current study, the nuisance variables examined include loading size, factor correlation size, model complexity, and model balance, where “model complexity” is judged by the total number of indicators, the total number of factors, and the ratio of indicators to factors included in the model and “model balance” is judged by the equal or unequal number of indicators per factor for two-factor models.

Addressing the second question allows for the assessment of commonly applied cutoff values in the literature. The goal is to determine whether the application of these cutoff values across different modeling conditions can be warranted given how indices are affected
by nuisance variables. It is not the goal here to suggest specific cutoff criteria for the indices presented. Rather, commonly applied cutoff values will be evaluated with the goal to determine whether indices behave consistently enough to warrant the use of these values across varying modeling situations.

With respect to the third question, the aim is to put forward a set of guidelines regarding the use of these fit indices in different modeling situations. Index behavior is examined here in a way that allows points of strength and weaknesses to be revealed under different misspecification types and under changing modeling components.

### 1.3 Thesis Structure

I begin with a discussion of the estimation procedure utilized in SEM, followed by an introduction to how model fit is assessed. The remainder of the present chapter introduces the chi-square and several popular indices used to determine model fit, focusing on their performance at the population level. Attention is then shifted to previous literature concerning both the chi-square and fit indices, highlighting both theoretical issues surrounding the use of these measures of fit and on the effects of various components such as estimation method, model size, parameter values, and misspecification type on index behavior.

In Chapter 2, the performances of seven indices defined at the population level are examined for a one-factor CFA model in which misspecification is due to one or several omitted error covariances. Nuisance variables are manipulated in order to determine their effects on index performance. The nuisance variables examined in Chapter 2 include loading size and heterogeneity, the number of omitted error covariances, size of the omitted error covariance(s), and the size of the model.
Chapter 3 presents the performances of indices for a two-factor CFA model in which misspecification is due to one or several omitted error covariances. As in Chapter 2, the nuisance variables manipulated include loading size, the number of omitted error covariances, the size of the omitted error covariance(s), and model size. In addition, the added factor complexity of the two-factor model allows for the manipulation of factor correlation size and model balance as well.

In Chapter 4, index behavior is examined for CFA models with a misspecified latent structure. The nuisance variables manipulated include loading size, factor correlation size, and model size. Finally, Chapter 5 presents an overview and discussion of the findings shown in the previous chapters. The discussion consists of a review of the various nuisance parameters and their effects on index behavior, a more in-depth focus on the performance of each index individually, possible explanations for index behavior, and finally some general guidelines to help aid researchers in their use of these indices.

1.4 Estimation and Model Fit

The primary goal of SEM is to examine how well the causal inferences proposed by a researcher actually match relationships found in sample data. The assessment of the adequacy of a model is done in several steps. First, a model is proposed that represents the supposed relationships amongst variables in the population. The model is fit to sample data and parameter estimates $\hat{\theta}$ are obtained. From these parameter estimates, the model-implied covariance matrix $\Sigma(\hat{\theta})$ is constructed. Model fit is determined by examining the discrepancies between this model-implied covariance matrix and the sample covariance matrix $S$. 
1.4.1 Estimation

It is hypothesized that a population covariance matrix $\Sigma^*$ is generated by $q$ true but unknown parameters under the null. These parameters, written in a $q \times 1$ vector $\theta$, correspond to the particular structure of $\Sigma^*$. A sample covariance matrix $S$, belonging to a sample from the population in question, would converge to $\Sigma^*$ if the sample size were to increase to infinity, at which point the structure of $S$ would be evident (Bentler & Bonett, 1980).

In order to test the null hypothesis $\Sigma^* = \Sigma(\theta)$, which states that the population covariance matrix has the structure implied by the model, estimates of the unknown parameters $\theta$ and matrix $\Sigma^*$ must be calculated under the proposed model. The vector $\hat{\theta}$ contains estimates of the model parameters. $\Sigma(\hat{\theta})$ represents the estimated covariance matrix for the hypothesized model as a function of the estimated model parameters. In an ideal world, one would know $\Sigma^*$ and be able to directly compare its structure to the structure arrived at by the researcher’s proposed model (Bentler & Bonett, 1980). In reality, however, $\Sigma^*$ is never actually known. This makes it impossible to directly test $\Sigma^* = \Sigma(\theta)$ and instead, researchers must compare the hypothesized covariance matrix to a sample matrix $S$.

One of the primary goals in SEM is to arrive at parameter estimates $\hat{\theta}$ such that the hypothesized model’s covariance structure based on these estimates is as similar to the structure of the sample covariance matrix $S$ as possible. This is achieved via the minimization of some discrepancy function $F(\theta)$. The discrepancy function, if given a set of parameters $\theta$, provides an assessment of the difference between the model-implied covariance matrix $\Sigma(\hat{\theta})$ and the sample covariance matrix $S$. The assessment is based on the residuals between these two matrices. According to Anderson and Gerbing (1984), maximum likelihood (ML) has been the predominant estimation procedure. It is used as the standard
default estimator in nearly all major SEM packages and thus will be the only estimation procedure discussed here\(^1\).

The traditional maximum likelihood fit function \(F_{ML}(\theta)\) (hereby written as \(F(\theta)\)) is based on the likelihood ratio and is given by

\[
F(\theta) = \ln|\Sigma(\theta)| - \ln|S| + \text{tr}[\Sigma(\theta)^{-1}] - p,
\]

where \(\Sigma(\theta)\) represents the structure of the covariance matrix implied by the hypothesized model, \(S\) represents the sample covariance matrix, and \(p\) is the number of observed variables (Bollen & Long, 1992). Minimizing this function gives \(F(\hat{\theta})\) (hereby written as \(\hat{F}\)), and the corresponding \(q \times 1\) vector of parameter estimates \(\hat{\theta}\).

The function can be defined in the population as well. If we were to know \(\Sigma^*\), it would replace \(S\) in Equation 1 and the expression of the maximum likelihood fit function in the population would be

\[
F(\theta) = \ln|\Sigma(\theta)| - \ln|\Sigma^*| + \text{tr}[\Sigma^*\Sigma(\theta)^{-1}] - p.
\]

### 1.4.2 Assessing model fit

Once the vector \(\hat{\theta}\) of parameter estimates has been found that minimizes the fit function, the corresponding model-implied covariance matrix \(\Sigma(\hat{\theta})\) can be assessed to determine how well it matches the structure of the sample covariance matrix \(S\). Discussed next are two broad methods of assessing model fit: use of the chi-square tests statistic to assess what is known in the literature as “exact fit,” and use of fit indices to assess “close fit.” These two types of fit are defined in the following sections.

\(^1\)If normality is assumed, all tests under different estimation minimization techniques will converge to a chi-square distribution as \(N\) increases to infinity (Amemiya & Anderson, 1990).
1.4.2.1 The chi-square test statistic

Traditionally, the assessment of model fit has been accomplished via a dichotomous decision process of hypothesis testing. With the null hypothesis $\Sigma^* = \Sigma(\theta)$, the model in question is tested as to whether it is exactly true in the population, and is thus a test of “exact fit” (Fan, Thompson, & Wang, 1999). Testing exact fit involves the use of the chi-square test statistic and its associated p-value. The chi-square assesses the discrepancy between the model-implied covariance matrix and the sample covariance matrix. The test statistic is obtained by multiplying the fit function minimum $\hat{F}$ by $(N - 1)$:

$$T = (N - 1)\hat{F},$$

which asymptotically follows a central chi-square distribution with $\frac{p(p+1)}{2} - q$ degrees of freedom under the assumptions that the model is correct and the data are multivariate normal (MacCallum, Browne, & Sugawara, 1996). This T statistic is used to test the null hypothesis $\Sigma^* = \Sigma(\theta)$.

Larger values of $\hat{F}$ indicate larger residuals between the model-implied covariance matrix and the sample covariance matrix (Amemiya & Anderson, 1990). If the residuals are larger than what would be expected due to sampling fluctuation, the T statistic will be greater than the critical value of the chi-square at the pre-specified alpha level. The null hypothesis that $\Sigma^* = \Sigma(\theta)$ will be rejected, indicating the hypothesized model structure is not true in the population.

If residuals are within sampling fluctuation given N, the resulting T will be small. The null hypothesis stating that the population covariance matrix has the structure $\Sigma(\theta)$ cannot be rejected, and the model is therefore retained.
The ML chi-square is a likelihood ratio test statistic. The likelihood of observing the data under the hypothesized model \( \Sigma(\theta) \) is compared to the likelihood of observing the data under the saturated model. Small values of the likelihood ratio indicate that the data are more likely to occur under the saturated model rather than the hypothesized model. Large values indicate that the data are equally or nearly equally as likely to occur under both the saturated and hypothesized models. Thus, large values imply that the structure behind the hypothesized model, when compared to a model imposing no structure at all, is not so restrictive that it fails to adequately fit the patterns found in the data.

1.4.2.2 Fit indices

Problems arise with the use of the chi-square in both very large and small sample sizes (these problems will be discussed further below). In addition, criticisms have been raised regarding the appropriateness of testing “exact fit” in the context of SEM, particularly in fields where modeling latent factors is common. Since it is unrealistic to assume that a given covariance structure will match that of the population exactly, some (e.g., McDonald & Marsh, 1990; Marsh, Balla, & McDonald, 1988) argue that it is more important to assess the degree of lack of fit rather than determining whether a model fits exactly.

In response to the problems surrounding the use of the chi-square, a number of fit indices have been developed to either replace the use of the chi-square when sampling issues arise or to be used alongside it as an additional way of assessing model fit. Most fit indices have been developed to provide information of how closely a model fits a given data rather than leading the researcher to a binary fit/no fit decision. Fit indices represent goodness of fit along a continuum and are to be interpreted as a gauge of “close fit” rather than of exact fit.
Like the chi-square, fit indices make use of the residuals, defined as \( S - \Sigma(\hat{\theta}) \) in sample settings and \( \Sigma^* - \Sigma(\hat{\theta}) \) in the population. Table 1 provides a list of commonly used fit indices defined both at the sample level and in the population. To derive the population equations, the values of the population minimized discrepancy functions (that is, the \( \hat{P} \) values obtained by minimizing Equation 2) times the sample size were used to replace chi-squares found in the sample definitions of the indices, and then \( N \) was allowed to tend to infinity. The final population expressions do not depend on \( N \), but only on \( \hat{P} \). Because models that are not exactly true yield minimized \( \hat{P} \) values that do not equal zero and worse-fitting models exhibit larger values of \( \hat{P} \) than those whose fit nearly approximates that of the true model (Steiger, Shapiro, & Browne 1985; Bentler, 1990), the value of \( \hat{P} \) can be utilized in the population as a measure of model misspecification.

The current study focuses on the performance of fit indices in the population, and thus only the seven population definitions in Table 1 are studied. Properties of these indices as well as rationale behind their use will be described in more detail below.

1.5 Literature Review

I now turn to the existing literature for a discussion of the issues that surround the use of both the chi-square and fit indices when assessing model fit. I first briefly summarize the literature on the issues affecting the model chi-square, and then turn to the literature on the behavior of the fit indices. A more in-depth discussion of sample and population definitions for fit indices (see Table 1) and the rationale behind them is also provided. While the current research is concerned with index behavior in the population, the majority of existing research on fit indices has been carried out at the sample level. Existing findings on the effects of sample size on fit indices are therefore reviewed briefly.
Table 1

*Names, Sample Definitions, and Population Definitions of Commonly Used Fit Indices*

<table>
<thead>
<tr>
<th>Index Name(s)</th>
<th>Sample Definition</th>
<th>Population Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normed Fit Index (NFI)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bentler-Bonnet Index (BBI)</td>
<td>( \frac{\chi_i^2 - \chi_M^2}{\chi_i^2} )</td>
<td>( 1 - \frac{F_M}{F_i} )</td>
</tr>
<tr>
<td>Bollen’s Fit Index</td>
<td>( \Delta_1 )</td>
<td></td>
</tr>
<tr>
<td>Bollen’s Incremental Fit Index (IFI)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BL89</td>
<td>( \frac{\chi_i^2 - \chi_M^2}{\chi_i^2 - df_i} )</td>
<td>( 1 - \frac{F_M}{F_i} )</td>
</tr>
<tr>
<td>Normed Fit Index 2 (NFI2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \Delta_2 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Normed Fit Index (NNFI)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tucker-Lewis Index (TLI)</td>
<td>( \frac{\chi_i^2}{df_i} - \frac{\chi_M^2}{df_M} )</td>
<td>( 1 - \frac{F_M}{F_i} \times \frac{df_i}{df_M} )</td>
</tr>
<tr>
<td>Bentler-Bonnet Non-Normed Fit Index (BBNFI)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bollen86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative Fit Index (RFI)</td>
<td>( \frac{\chi_i^2}{df_i} - \frac{\chi_M^2}{df_M} )</td>
<td>( 1 - \frac{F_M}{F_i} \times \frac{df_i}{df_M} )</td>
</tr>
</tbody>
</table>
Table 1 (continued)

*Names, Sample Definitions, and Population Definitions of Commonly Used Fit Indices*

<table>
<thead>
<tr>
<th>Index Name(s)</th>
<th>Sample Definition</th>
<th>Population Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goodness of Fit Index (GFI)</td>
<td>$1 - \frac{\chi_i^2}{\chi_M^2}$</td>
<td>$1 - \frac{\text{tr}([\Sigma(\hat{\theta})^{-1}(\Sigma^<em>-\Sigma(\hat{\theta}))^2])}{\text{tr}(\Sigma(\hat{\theta})^{-1}\Sigma^</em>)^2}$</td>
</tr>
<tr>
<td>Adjusted Goodness of Fit Index</td>
<td>$1 - \frac{GFI(\frac{df_M}{df_i})}{1 - \frac{p(p + 1)}{2df} \times (1 - GFI)}$</td>
<td></td>
</tr>
<tr>
<td>(AGFI)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standardized Root Mean Square</td>
<td>$\sqrt{\frac{2}{n(n + 1)} \sum_i \sum_j (s_{ij} - c_{ij})^2}$</td>
<td>$\sqrt{\frac{\sum(R^* - R(\hat{\theta}))^2}{p(p + 1)}}$</td>
</tr>
<tr>
<td>Residual (SRMR)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Root Mean Square Error of</td>
<td>$\frac{\chi^2}{df - 1}$</td>
<td>$\sqrt{\frac{\hat{F}}{df}}$</td>
</tr>
<tr>
<td>Approximation (RMSEA)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gamma</td>
<td>$\frac{p}{p + 2(\frac{\chi^2}{n - 1} - df)}$</td>
<td>$\frac{p}{p + 2df(RMSEA^2)}$</td>
</tr>
</tbody>
</table>

Note. Where $p$ is the number of indicators, $n$ is the sample size, $\chi_i^2$ and $\chi_M^2$ stand for the chi-square values for the independent (baseline) model and the proposed model, respectively; $df_i$ and $df_M$ are the degrees of freedom for the independent model and the proposed model, respectively; $\hat{F}_i$ and $\hat{F}_M$ stand for the minimized fit function for the independent and proposed models, respectively; $R^*$ is the population correlation matrix, $R(\hat{\theta})$ is the model-predicted correlation matrix, and $\hat{F}$ is the minimum of the ML fit function.
1.5.1 Chi-square

1.5.1.1 Sample size

As is true with any statistical test, the power of the model chi-square test is a direct function of sample size. As early as the 1970s, researchers such as Joreskog (1978) and Bentler and Bonett (1980) have noted that unless the model fits perfectly, an increase in sample size will inflate the chi-square value. In large enough samples, a large T statistic can be obtained even when the model is trivially misspecified. While in other settings (such as ANOVA or regression) this power tends to work towards the favor of the researcher, such a feature works against those using SEM where the goal is to retain the null.

The issue of sensitivity to sample size limits the practical use of the chi-square in SEM. While a significant chi-square appropriately indicates that the model does not fit the data, the sensitivity with which it is does so may be impractical for research applications in which it is expected that the hypothesized model will not fit the data exactly (Bearden, Sharma, & Teel, 1982; Gerbing & Anderson, 1992). This can lead to the rejection of models that, while not fitting the data perfectly, fit the data well enough in a practical sense to warrant their use as an appropriate representation of the relationships between the variables in the population (Fornell & Larker, 1981).

Problems with the use of the chi-square can also arise when the sample size is small. The T statistic follows an asymptotic chi-square distribution which may not be well approximated in smaller sample sizes (Bentler & Yuan, 1999). Geweke and Singleton (1980), Hu, Bentler, and Kano, (1992), and Bentler and Yuan (1999), among others, have shown that in small sample sizes the chi-square tends to over-reject the null. This behavior could lead to incorrect conclusions about the adequacy of the model.
1.5.1.2 Normality

The assumptions of multivariate normality must also be taken into consideration when using the chi-square (Curran, Finch, & West, 1996). Many others (e.g., Bentler and Yuan, 1999; Gerbing and Anderson, 1992; Boomsma, 1982, 1983) have shown that violations of multivariate normality result in over-rejection of the true model.

1.5.2 Fit indices

Before reviewing the literature studying the behavior of the fit indices in SEM, a more in-depth introduction to the seven population definitions shown in Table 1 is presented. Throughout the remainder of this thesis, population indices will be expressed by a single index name (e.g., CFI, NNFI, etc.) but are meant to refer to all indices that reduce to that particular population definition. Please refer to Table 1 for the complete list of indices defined at the sample level and their corresponding population definitions.

1.5.2.1 Population definitions

1.5.2.1.1 Class 1

The first class of fit indices studied here is defined at the population level as:

\[ CFI = 1 - \frac{\hat{F}_M}{\hat{F}_I}, \]  

(4)

where \( \hat{F}_M \) is the minimized fit function of the proposed model and \( \hat{F}_I \) is the minimized fit function of a baseline model, most often defined as the model in which all variables are mutually uncorrelated (though the baseline model can be any other model selected by the researcher). It is expected that the value of \( \hat{F}_I \) is large (indicating a poor fit). It is hoped that the value of \( \hat{F}_M \) corresponding to the proposed model is small, indicating good fit. Class 1 indices utilize the information from the T statistics found by fitting the proposed and baseline
models. Thus, they are influenced both by the badness of the baseline model and the goodness of fit of the proposed model (Kenny & McCoach, 2003).

Though the comparative fit index (CFI) is the most commonly used index of this class, an index with the population definition in Equation 4 was initially proposed by Bentler and Bonett (1980) with their introduction of the normed fit index (NFI). Along with CFI and NFI, Bollen’s (1986) BL86 and Bollen’s (1989) incremental fit index (IFI) reduce to the Class 1 population definition. Because CFI is the most widely used index with this population definition, Equation 1 will be referred to as CFI for the remainder of this thesis.

CFI is bound by zero and one, with values of one indicating perfect fit \((F_M = 0)\) and zero indicating poor fit. The generally agreed upon cutoff value for CFI (e.g., Hu & Bentler, 1999; Beauducel & Wittmann, 2005; Hooper, Coughlan, & Mullen, 2008), is .95, with values greater than .95 indicating good model fit.

1.5.2.1.2 Class 2

The second class of fit indices is defined in the population as

\[
\text{NNFI} = 1 - \frac{F_M}{F} \times \frac{df_I}{df_M},
\]

where \(df_I\) and \(df_M\) are the degrees of freedom associated with the baseline (as defined above) and proposed models, respectively. Indices under this class include the non-normed fit index (NNFI) proposed by Bentler and Bonett (1980) and Tucker & Lewis’ (1973) TLI. NNFI is the most widely-used index of this class, and therefore Equation 2 will be referred to as NNFI for the remainder of this thesis.

The inclusion of the ratio of degrees of freedom in Equation 5 causes NNFI to be interpreted differently than CFI (Equation 4). CFI is interpreted as a comparative reduction in noncentrality, with the comparison between the model assuming all variables are
uncorrelated (the baseline model) and a model imposing a certain structure (the hypothesized model). NNFI can be interpreted as the relative reduction in misfit per degree of freedom. That is, indices with the population definition of NNFI involve an adjustment for model parsimony (Bentler, 1990) and are often seen as superior indices to CFI (Mulaik, et al., 1989). However, unlike CFI, NNFI lacks a lower bound and may attain negative values (Hooper, Coughlan, & Mullen, 2008). This could sometimes make interpretation difficult. Like CFI, values approaching one indicate good model fit and the same cutoff criterion of .95 is used.

1.5.2.1.3 Class 3

Unlike Class 1 and Class 2 indices, the remaining fit indices (Classes 3 through 7) depend only on the fit of the hypothesized model. There is no comparison to a baseline model. They evaluate the degree to which the model-implied covariance matrix matches the covariance matrix of the observed data and can be interpreted as a reflection of how well the proposed model fits in comparison to no model at all (Hooper, Coughlan, & Mullen, 2008). Thus, they are said to assess “badness of fit.”

A third population index is defined as:

\[
GFI = 1 - \frac{\text{Tr}[\left(\Sigma(\hat{\theta})^{-1}(\Sigma^* - \Sigma(\hat{\theta}))\right)^2]}{\text{Tr}(\Sigma(\hat{\theta})^{-1}\Sigma^*)^2)}, \tag{6}
\]

where \(\Sigma^*\) represents the population covariance matrix and \(\Sigma(\hat{\theta})\) represents the model-implied covariance matrix. The only sample equation corresponding to this population definition is the goodness-of-fit index (GFI) proposed by Joreskog and Sorbom (1981), and thus the Class 3 index will be referred to as GFI for the remainder of this thesis. GFI calculates the proportion of variance accounted for in the sample covariance matrix by the estimated covariance matrix derived from the proposed model (Anderson & Gerbing, 1988). GFI is
bound by zero and one. As with the previous two indices, an index value of one denotes perfect fit, and the recommended cutoff value is .95.

1.5.2.1.4 Class 4

It has been demonstrated (e.g., Cudeck & Browne, 1983; MacCallum & Hong, 1997) that GFI shows an improvement in fit as additional parameters are included in the model. Joreskog and Sorbom (1981) were aware of this phenomenon and developed an adjustment that compensates for this increase in fit. A population expression of their adjusted goodness-of-fit index (AGFI) is

\[ \text{AGFI} = 1 - \frac{p(p+1)}{2df} \times (1 - GFI), \]  

which includes the GFI value as a component of its calculation. Like NNFI, AGFI adjusts for parsimony by including both \( p \), the number of indicators in the model, and the degrees of freedom. It is interpreted as the proportion of variance in the population covariance matrix accounted for by the estimated population covariance matrix per degree of freedom.

Joreskog and Sorbom state that AGFI has the same properties as GFI. However, unlike GFI, AGFI lacks a lower bound and can take on negative values (Anderson & Gerbing, 1988). An AGFI value of one indicates perfect fit and the cutoff value of .95 is used to indicate models with good fit.

1.5.2.1.5 Class 5

A third index developed by Joreskog and Sorbom (1981) was designed to assess the average magnitude of the residuals between the sample and hypothesized covariance matrices. The population expression of the standardized root mean square residual (SRMR) is

\[ \text{SRMR} = \sqrt{\frac{\sum(R^* - R(0))^2}{p(p+1)}}, \]
where $R^*$ represents the population correlation matrix and $R(\hat{\theta})$ represents the estimated population correlation matrix. In this thesis, SRMR is studied instead of RMR (which involves a comparison between the population covariance matrix and the estimated population covariance matrix). RMR is not bounded and can attain very large values, making interpretation difficult. SRMR, on the other hand, is bounded below by zero and rarely exceeds one. Thus, its range is more easily interpretable and its performance here can be more easily compared to the performances of the other indices in this study.

The fact that SRMR is a function of the residuals classifies it as a measure of “badness of fit” rather than of goodness of fit (Chen, 2007). Thus, a value of zero ($\Sigma^* - \Sigma(\hat{\theta}) = 0$) indicates perfect fit. Models with index values less than .08 are said to be well-fitting (e.g., McDonald, 1989; Beauducel & Wittmann, 2005).

1.5.2.1.6 Class 6

A sixth population definition is based on the minimum of the fit function $F(\theta)$. A population expression of the root mean square error of approximation (RMSEA) can be written as

$$\text{RMSEA} = \sqrt{\frac{\hat{F}}{df}}, \quad (9)$$

where $\hat{F}$ is the minimized fit function value obtained by fitting the model to the population covariance matrix $\Sigma^*$. RMSEA was initially developed by Steiger (1990) and expanded upon by Browne & Cudeck (1992). The index measures the discrepancy between the observed covariance matrix and the model-implied covariance matrix per degrees of freedom. Thus, it includes a built-in penalty for lack of parsimony (Cudeck & Browne, 1983).
RMSEA is bounded below by zero and is regarded as a measure of “badness of fit.” When $\Sigma^* = \Sigma(\hat{\theta})$, $\hat{F} = 0$; thus, a value of zero is achieved when model fit is perfect. The most generally accepted cutoff value of .06 has been proposed by Hu & Bentler (1999).

1.5.2.1.7 Class 7

Finally, a seventh population definition can be expressed as a function of RMSEA as

$$\gamma = \frac{p}{p+2df(RMSEA^2)}.$$  \hspace{1cm} (10)

The only index with this population equation is gamma (also called gamma hat), which was originally proposed by Steiger (1989) and later expressed as a function of RMSEA (most notably by Fan and Sivo, 2007). While RMSEA contains an adjustment for parsimony, it has been shown (e.g., by Kenny & McCoach, 2003) that RMSEA values show an improvement in fit as the number of variables included in the model increases. Gamma addresses this issue by directly including $p$, the number of variables, into its equation. According to Fan and Sivo (2007), this adjustment over and above the adjustment in degrees of freedom made by RMSEA acts to lessen the improvement in fit seen by RMSEA as the size of the model increases. Models with gamma values larger than .95 are said to be well-fitting (e.g., Hu & Bentler, 1999).

1.5.2.2 Omitted indices

We make note of several indices that are not included in the current study. RMSEA, NNFI, and AGFI, presented above, have been developed to take model parsimony into account. It has been shown that indices like NNFI and AGFI lack a lower bound and thus may be difficult to interpret. However, in most scenarios these indices behave similarly to those for which they act as adjustments and thus can generally be interpreted within the same range as their parent indices. Therefore, NNFI and AGFI are included in this study.
Other parsimony-adjusted indices, however, may not be interpretable along any such range. The PNFI and PGFI indices, for example, are two additional parsimony adjustments for NFI and GFI. Unlike NNFI and AGFI, they are not bound by zero and one. In fact, judging the values of PNFI and GFI on the typical zero to one scale is not conducive to understanding their meaning. Because of these interpretation difficulties, indices such as these will not be considered in the current research.

1.5.3 Concerns regarding fit indices

Fit indices have, in general, been developed to overcome some of the problems associated with the use of the chi-square. Despite this, the use of fit indices over time has led researchers to discover that these newer methods are not without their own set of problems. Some of these problems have to do with how fit indices are used and interpreted. Other problems stem from actual applications of indices and the properties that apply to them.

I now summarize several of the pertinent theoretical issues surrounding the use of fit indices in structural equation modeling, focusing on the choice of index and the use of cutoff values in assessing model fit. Following this, relevant studies are summarized that on the behavior of fit indices with respect to different nuisance variables, including model size, estimation method, misspecification type, and parameter size. These summaries are followed by a discussion of what these previous studies have shown and how the current research aims to address some of the remaining issues.

1.5.3.1 Theoretical and methodological concerns

1.5.3.1.1 Choosing an index

Certain theoretical issues arise when fit indices are employed to judge model adequacy. One issue involves choosing what indices to use and report. Popular SEM
programs are capable of providing values for upwards of seven or eight indices, and it is not uncommon for researchers to arrive at conflicting conclusions depending on what indices they choose to examine (Hu & Bentler, 1998).

In addition to this problem, there is disagreement in the literature regarding which indices are best to report and whether all indices are appropriate in all situations. Different indices were developed to assess different aspect of model fit (Sivo, Fan, Witta, & Willse, 2006; Gerbing & Anderson, 1992). Comparative indices, for instance, utilize both the hypothesized model and a baseline model (often the independence model) with the goal to determine which more accurately reproduces the sample covariance structure. Other indices, such as those that rely on the residuals between the hypothesized covariance matrix and the sample covariance matrix, essentially compare the hypothesized model to the saturated model. Still others, like NNFI, include adjustments for model parsimony and thus may behave differently than indices, such as the CFI, that do not.

Researchers unaware of the differences in how model fit is calculated for different indices may be tempted, when provided with seven or eight indices, to report and interpret only those that support their model. Thus, it is important that researchers be aware of the rationale behind different indices so as to make appropriate choices when selecting which indices to report.

1.5.3.1.2 Cutoff values

Another methodological issue highly relevant to the focus of the current research is the use of index cutoff values. Many studies, including recent work by Marsh, Hau, and Wen (2004), Beauducel and Wittmann (2005), Fan and Sivo (2005), and Yuan (2005), have demonstrated that a single cutoff value cannot be used reliably under all measurement and
data conditions. However, there is countless evidence in the literature to suggest that these
cutoffs are in fact being applied across many different modeling situations.

Quite often, little to no explanation is given as to why specific values are being used
as cutoff criterion. For example, there may be agreement among researchers that RMSEA
values greater than .06 indicate poor model fit, but such agreement does not give the value of
.06 any particular meaning, nor does it necessarily help determine appropriate fit in all
situations. A model with an RMSEA value of .062 may not be considered to be well-fitting
using a cutoff of .06, but does this mean that the model is substantially worse than a model
that achieves a value of .058 to the point that the former model should be rejected and the
latter accepted? The lack of rationale behind such cutoff values may lead to situations where
models that adequately fit the data are rejected because their corresponding fit index value
falls on the wrong side of an arbitrarily chosen point.

The lack of a clear rationale behind such cutoff values has caused many to question
the use of cutoff values and of fit indices themselves. Sivo, Fan, Witta, and Willse (2006),
among others, criticize the universal application of cutoff values for several additional
reasons. Particularly, they note that sample size and model type both affect index behavior
and as such, there exists a general lack of comparability across different model types and
sample sizes, even within the same index. This lack of comparability makes the use of a
this view, arguing that since model fit can vary as a function of model type, a universal cutoff
value is unrealistic.

Further issues surrounding the use of cutoff values stem from the reason fit indices
were developed in the first place. Initially, indices were developed to gauge model fit along a
continuum. This provided researchers with an additional perspective on how well a model fit the data. By providing an assessment of the degree of fit of a model, the use of fit indices complemented the use of the binary fit/no fit decision arrived at by use of the chi-square test statistic.

The use of cutoff criteria, however, has led to changes in the interpretation of fit indices. Marsh et al. (2004) state that the strict use of these criteria has transformed fit indices into a type of hypothesis test similar to that of the chi-square. Instead of allowing researchers to examine the approximate fit of a model, the use of cutoff values as strict rules has led many back to the binary decision making process of “fit vs. no fit.” Barrett (2007) also makes note of the change in interpretation and argues that if interpreted this way, fit indices play no significant role in adjudging model fit over and above that of the chi-square. He goes so far as to recommend that certain indices be abandoned, citing interpretational issues and the uncertainties surrounding appropriate cutoff criteria.

While Barrett and his supporters have seen these problems surrounding the use of cutoff values as evidence to question the use of fit indices at all, others, including Marsh, Balla, and McDonald (1988), McDonald and Marsh (1990), and Bentler (2000), support the use of fit indices but criticize the widespread use of cutoff values, stressing that traditional cutoff values are nothing more than rules of thumb based mainly on intuition and limited research.

Even Hu and Bentler (1998, 1999), whose work has focused on developing more appropriate cutoff criteria, argue against using them as “golden rules.” They state that while these criteria can be of use in helping gauge model fit, they should not be sole criterion taken
into consideration. Theoretical and practical reasons should also be given thought when determining the appropriateness of a model, as should common sense.

Given issues surrounding both choosing appropriate indices and interpreting fit indices with respect to cutoff values, can the use of fit indices in structural equation modeling be justified? Those in support of using fit indices and have suggested methods and recommendations in order to help remedy some of the problems surrounding their use.

Research by Hu and Bentler (1998) and recommendations by Bollen and Long (1992) and McDonald and Ho (2002) support the examination of multiple indices to best determine model fit. Hu and Bentler have shown that different indices are sensitive to different components of misspecification. Rather than abandoning the use of these indices because of these differences, they suggest reporting multiple indices to provide as much detail about fit as possible. Bollen and Long (1992) support this idea and add that the examination of additional model components such as the R-squares of the equations and the magnitudes of the coefficients can lend further support to decisions reached by using fit indices.

In a response to criticism surrounding the use of strict cutoff values as strict points at which to determine model fit, many argue that fit indices can still be used, provided they are interpreted as they were meant to be interpreted. Developed to compliment the “fit vs. no fit” decision reached by the chi-square, indices should therefore not be interpreted like the chi-square by imposing cutoff values that support a binary decision (Hu & Bentler, 1998). Fan, Thompson, and Wang (1999) and Millsap (2007), among others, emphasize this point as well. The purpose of a fit index is to allow researchers to examine the degree of misspecification, and thus it is not necessary that indices provide a strict rule by which to claim model acceptance or rejection. Indices are still of use to researchers in determining...
model fit, but researchers must exercise caution when interpreting index values with respect to commonly used cutoff criteria.

1.5.3.2 Sample size, estimation method, and model features

In addition to the more theoretical issues surrounding the use of fit indices, consideration must also be given to how indices behave under different modeling conditions. Specifically, previous research has shown indices to behave differently in different modeling situations with respect to sample size, estimation method, and various model features such as model size, the size of model parameters, and misspecification type. While differences in index performance are expected under different conditions and should therefore not be considered bad, researchers should be fully aware of how indices change and behave with respect to changes across modeling conditions.

The current research addresses the behavior of indices in the population. As such, only a brief overview of the effects of sample size will be summarized here. For the remainder of the thesis, effects of sample size will be discussed only when they are presented in conjunction with other variables that remain of interest at the population level.

1.5.3.2.1 Sample size

Though one of the primary reasons driving the development of fit indices was to develop a test of fit that improved over the chi-square’s sensitivity to sample size, research has shown that the majority of fit indices are not immune to the effects of N. In a large study by Marsh, Balla, and McDonald (1988), the performances of 29 different fit indices were examined to determine which, if any, were independent of sample size. Index performance was evaluated in seven sample size conditions (25, 50, 100, 200, 400, 800, 1600) when fitting a three-factor model to four sets of data. The variation of index value across sample
sizes was used to determine whether an index was independent of sample influences. Of the 29 indices examined, only five were found to be significantly unaffected by sample size. For the majority of indices, substantial variation in index value due to sample size was found. This was despite the fact that the degree of misspecification in the model remained the same in all conditions.

Other studies have confirmed the findings of Marsh, Balla, and McDonald. La Du and Tanaka (1995) found comparative fit indices (those belonging to Class 1 and Class 2 as defined above) to be affected by sample size more so than other types of indices, detecting misspecification more consistently as sample size increased. Their results agreed with previous research by Bearden, Sharma, and Teel (1982), Anderson and Gerbing (1984), and Gerbing and Anderson (1992).

More recent studies (e.g., Fan, Thompson, & Wang, 1999; Sharma, Mukherjee, Kumar, & Dillon, 2005) focusing on indices outside of the “comparative” class of indices have shown GFI to be the index most influenced by the effects of sample size, while RMSEA has been shown to be almost uninfluenced by the effects of sample size once N > 200. Though different studies have been carried out examining fit indices in different modeling situations, the effects of sample size have been documented in almost all cases.

1.5.3.2.2 Estimation method

An additional source of differences among index values across different modeling scenarios may arise from the type of estimation method used. It has been shown that certain indices perform differently depending on the estimation method used. Only a brief summary of studies in this area is included here, as the current research examines behavior only under maximum likelihood (ML) estimation.
Sugawara and MacCallum (1993) found that absolute fit indices (Class 3 through Class 7 indices as defined here) tend to behave more consistently across different estimations than comparative fit indices (Class 1 and Class 2 indices). Their findings were supported by La Du and Tanaka (1995), who examined the impact of estimation method when the goal is model comparison rather than determining the fit of a single model. La Du and Tanaka looked at the behavior of GFI and NFI under both ML and generalized least squares (GLS) estimation for both a true model and a misspecified model that estimated a correlation between two orthogonal factors. GFI was found to be the best behaved, particularly for nonnormal data, across estimation methods. While the effect of different estimation methods is not directly related to the current research, results such as this should be noted, as they suggest that behaviors of certain indices under one estimation method may not generalizable across other methods of estimation, particularly for Class 1 (CFI) and Class 2 (NNFI) indices.

Fan, Thompson, and Wang (1999) found results similar to those of La Du and Tanaka. These researchers examined the behavior of GFI, AGFI, CFI, NNFI, NFI, and RMSEA under ML and GLS estimation. Their design included a true and two misspecified models. The true model was a four-factor, 12-indicator model. The “slightly misspecified” model omitted two factor loadings present in the true model. The “moderately misspecified” model omitted six factor loadings and included one loading not present in the true model. Index values were examined at sample sizes ranging from 50 to 1,000. The authors found large differences in index values across the two estimation procedures for NFI and in general, indices belonging to Class 1 and Class 2 (CFI, NFI and NNFI in Fan, Thompson, and
Wang’s study) behaved much less stably across estimation methods than those belonging to other classes.

1.5.3.2.3 Model size

Fit indices are also affected by model size, as defined both by the number of latent factors and the number of indicators in a given model. Sensitivity to model size can be seen as problematic depending on how the inclusion of more variables affects index value. Kenny and McCoach (2003) argue that the way in which model complexity affects index value could have an impact on how researchers respond to the fit of their models. If an increase in variables leads to a decrease in model fit according to a given index, undesirable practices such as trimming variables or testing submodels rather than original models could be undertaken to achieve a certain degree of fit. On the other hand, an increase in variables may lead to an improvement in fit. If this were the case, it could possibly lead to researchers including variables that should not theoretically belong in a model but are added solely to improve fit.

Several studies have examined the influence of the number of indicators on fit index value. Anderson and Gerbing (1984) focused on the effect of sample size in conjunction with the number of indicators and the number of factors contained in a model. They looked at these effects for GFI, AGFI, RMR, and the chi-square. Sample size varied from 50 to 300 along with three different numbers of indicators (two, three, or four per factor) and three different numbers of factors (two, three, or four). Loadings were either .6 or .9, and factor correlation was set to either .3 or .5.

The authors found that the mean values of GFI, AGFI, and RMR were most strongly influenced by sample size, the number of indicators per factor, and the number of factors in
the model. Particularly, as the sample size increased, GFI and AGFI showed an increase in fit. This effect was slightly moderated by the interaction between the number of indicators per factor and the number of factors in the model. As models grew larger (both in terms of the number of indicators and the number of factors), GFI and AGFI indicated worse fit, while RMR indicated better fit as the number of factors increases.

From this research, one could predict that holding sample size constant, models with more indicators and more complex latent structures will fit worse than simpler, smaller models according to GFI and AGFI. The current study will test this prediction and will examine the effect of model size (both in terms of indicators and factors) on index behavior at the population level (independently of the effects of sample size).

We note also that Anderson and Gerbing did not examine situations where the number of indicators per factor was larger than four. The current study examines index behavior both when the number of indicators per factor is small (three indicators per factor) and very large (up to 47 indicators per factor), to gain insight into the trends of index behavior as the relative sizes of factors change.

The effect of model size on fit index behavior was also examined by Sharma, Mukherjee, Kumar, and Dillon (2005). To study the effects sample size and the number of indicators on the performance of TLI, RMSEA, and GFI with respect to prespecified cutoff values, these researchers constructed four CFA models with different numbers of factors (two, four, six, and eight factors) where the number of indicators per factor was set to four (thus, the two-, four-, six-, and eight-factor models had a total of 8, 16, 24, and 32 indicators total). The authors varied factor loading sizes (.3, .5, .7), factor correlations (.3, .5, .7), and sample sizes (100, 200, 400, 800). Index values were computed for both a true model and a
misspecified model for each of the four levels of factor complexity (misspecification was due to omitted factor correlations). The mean values of the indices were calculated across all manipulations and were used to determine whether index behavior was consistent across the different levels of loading sizes, factor correlation sizes, sample sizes, and number of indicators.

The authors found GFI to be significantly affected by both sample size and the number of indicators present, indicating worse fit as the sample and model size increased. The effect of sample size became more prominent as more indicators were included. While this study did not manipulate the number of factors in the model as did Anderson and Gerbing’s 1984 study, the results for GFI are similar for both studies and suggest that GFI will show worse fit as sample size and the number of indicators in the model both increase, regardless of factor complexity. The current study will examine GFI behavior in the population to determine if the changes in GFI are due solely to an increase in the number of indicators.

The study by Sharma, et al. also revealed that as sample size increased, more misspecified models were accepted as fitting using a cutoff of .05 for RMSEA. This effect was independent of the number of indicators in the model, and suggests that while TLI appeared least affected by sample size of the indices studied, RMSEA exhibits some degree of improvement in fit as the sample size increases. Of interest here, however, are modeling issues that remain influential at the population level. Thus, further discussion with respect to sample size influence will be brief.

Other studies (e.g., Boomsma, 1982; Bearden, Sharma, & Teel, 1982; Anderson & Gerbing, 1984) have found that the number of indicator variables in a model affects index
performance independently of sample size. A study by Chau and Hocevar (1995) examined the effect of the number of indicators in a model on the behaviors of GFI, AGFI, NFI, CFI, and TLI. The initial model in their study consisted of data with seven factors, each with four indicator variables (28 indicators total). In two additional manipulations, the same model was maintained but the number of indicators per factor was reduced down to three (21 indicators total) and then to two (14 indicators total).

The authors found that while all indices showed worse fit for larger models (those with more indicators), NFI and CFI values were relatively stable across different model sizes when compared to GFI and AGFI, and TLI was found to be the most stable of the five indices studied. They note that the results for AGFI are surprising, as the index was specifically proposed as an adjustment to GFI correcting for model parsimony.

We point out that in this study by Chau and Hocevar, no effort was made to control for changes in the ratio of indicators ($p$) to factors ($k$) as the total number of indicators in the model is reduced. That is, as $p$ decreases, the ratio of $p:k$ decreases as well, from 4:1, to 3:1, to 2:1 at model sizes of $p = 28$, 21, and 14, respectively. The current study, in addition to studying index behavior with respect to changes in the number of indicators in a model, will in addition examine whether indices behave differently due to changes in the ratio of $p:k$. A comparison of our results to the results found by Chau and Hocevar will allow us to gain insight as to whether indices are sensitive to changes in the $p:k$ ratio over and above changes in $p$ (or in $k$) alone.

Kenny and McCoach (2003) also examined the effects of the number of indicators on index performance. Their study involved a series of simulations in which models were created with a varying number of indicators. CFI, TLI, and RMSEA were examined in three
types of misspecification error scenarios. The first involved an error related to a minor factor, in which the loadings of a second minor factor were omitted from the fitted one-factor model. The second involved an error involving a two-factor model for which the correlation between the two factors was omitted in the fitted model. The third involved a method error, in which error covariances in a one-factor model were omitted in the fitted model. Each scenario had a degree of specification error equal to a CFI of .92 for an 8-indicator model at N = 200.

The authors found the behavior of CFI differed depending on the source of the misspecification. The addition of more indicators caused a substantial decrease fit according to CFI when the source of misspecification involved a minor factor or involved an omitted factor correlation. The same increase in the number of indicators showed an improvement in fit for CFI when the source of misspecification was due to omitted error covariances in a one-factor model. Similar results were found for TLI.

RMSEA, regardless of the source of the misspecification, showed an improvement in fit as the number of variables in the model increased. This result is consistent with the findings of other studies (e.g., Breivik & Olsson, 2001) and supports previous literature stating that RMSEA shows better fit as models grow larger. The authors reasoned that the decline in RMSEA value (zero indicates perfect fit) is due to the decline in the ratio of the model chi-square to its degrees of freedom, as adding indicators increases the degrees of freedom of the model faster than it increases the model chi-square.

Given the results of this study, we may expect different trends in CFI behavior depending on the source of misspecification, but not for RMSEA. An examination of the equation for the population RMSEA (see Equation 6) reveals that RMSEA values are a function of both the minimized fit function $\hat{F}$ and the degrees of freedom. We explore in the
current study whether this increase in fit as shown by RMSEA is due to changes in $\hat{F}$ or due to the changes in the degrees of freedom associated with increasing model size.

1.5.3.2.4 Parameter values

Research has also examined the effects of parameter values such as loading size and factor covariance size on fit index values. Anderson and Gerbing (1984) found loading size to be influential for several fit indices. The main goal of their study was to determine the effect of sample size on index performance in conjunction with the number of indicators and the number of factors. Along with manipulating sample sizes ($N = 50 – 300$), different numbers of indicators (two, three, or four per factor), and different numbers of factors (two, three, or four factors total), parameter values were manipulated as well. Loading sizes were either set to .6, .9, or were a mix of .6 and .9, and factor correlation was set to either .3 or .5.

The authors found loading size to affect the performance of RMR, with higher loadings corresponding to increased fit. Further findings by Anderson, Gerbing, and Narayanan (1985) support this conclusion. Though not explicitly examined in Anderson and Gerbing’s study, we can assume that these results can be extended to the standardized version of RMR (SRMR). While factor correlation size was found to have no practical effect on any index studied, we suspect this was the case because of the relatively small change in the size of the factor correlation (index performance was evaluated only at factor correlations of .3 and .5). The current study further investigates the effects of factor correlation on index performance by examining a larger range of factor correlation sizes.

In a study of the GFI, Shevlin and Miles (1998) found loading size to have an influence on GFI performance. They studied both a correctly specified one-factor model and two misspecified models. The “approximate” model included three correlated errors not
present in the data. The “misspecified” model contained a second factor and a factor correlation of .7. GFI behavior was then examined at five sample sizes (N = 50, 100, 200, 400, or 800) and at three loadings sizes (.3, .5, and .7).

The study revealed that GFI was influenced by an interaction between sample size and loading size. Previous studies (e.g., Anderson & Gerbing, 1984; MacCallum & Hong, 1997; Sharma, et al. (2005)) have shown that GFI indicates better fit as sample size increases. Shevlin and Miles, however, found this relationship only for their true and “approximate” models. For the “misspecified” model, GFI showed an improvement in fit with larger sample sizes only when loadings were moderate (.5) or high (.7). The authors argue that this behavior is a positive trait of GFI. It demonstrates that, for more severely misspecified models, Type 1 error does not strictly increase with sample size. It also may suggest a relationship between GFI and loading size that is independent of sample size. By examining the behavior of GFI and other indices defined at the population level, the current study investigates whether loading size, affects index performance independently of sample size.

Further exploration of the role of parameter values on index behavior was performed by Miles and Shevlin (2007) who showed that, controlling for the level of misspecification, index performance can change due to loading size alone. In the second half of a two-part study, the authors fit a one-factor model to two-factor data with loadings of .8 and a factor correlation of .5. They found that the chi-square and all the indices in their study, including RMSEA, CFI, NNFI, and RMR, showed the model as fitting the data poorly. However, when the loadings were reduced to .5, the chi-square and RMSEA indicated a well-fitting model. The comparative indices (CFI and NNFI) still showed a poor fit. These results are somewhat discouraging, as they suggest that at smaller loadings, RMSEA may not be sensitive enough
to detect a misspecification as large as an omitted factor. The authors argue that such findings support the use of comparative fit indices such as CFI or NNFI alongside the chi-square to gain a better understanding of model fit in situations where the chi-square may not reveal an actual misspecification.

The scenario presented by Miles and Shevlin is perhaps the most similar to the scenario carried out in the current study to explore index’ ability to detect cases of latent structure misspecification. Given the results of their study, we may expect to see differences in RMSEA’s ability to detect a misspecified latent structure as a function of the size of the loadings.

1.5.3.2.5 Source of misspecification

While sensitivity to modeling components such as sample size and estimation method may be seen as a weakness for a fit index, it is desirable for these indices to be sensitive to the nature and degree of misfit between a proposed model and the data. For example, if the size of a factor correlation in the data is .7, an ideal fit index would denote worse fit for a model omitting this correlation than if the correlation in the data were only .2. Similarly, an ideal index would appropriately reflect worse fit for a model omitting two error covariances that exist in the data than for a model only omitting one.

However, it has been shown that indices do not always accurately reflect the degree of misspecification. Hu and Bentler (1998) examined the behavior of 15 indices with respect to the severity of model misspecification. They wanted to determine which indices, if any, tended to accurately reflect the degree of misspecification. The authors examined index performance in three separate three-factor CFA models: a true model, one involving omitted factor covariances, and a third involving omitted factor cross-loadings. Different estimation
methods (ML, GLS, and ADF) and sample sizes (N = 150 – 5000) were examined as well. Index behavior was judged based on the proportion of variance accounted for solely by model misspecification as calculated by a series of ANOVAs.

Results from the study showed that in the case of misspecification involving omitted factor covariances, SRMR performed best in terms of a large proportion of its variance being accounted for by differences in the degree of model misspecification as judged by the size of the \( \eta^2 \) s. TLI, BL89, CFI, and RMSEA performed nearly as well as SRMR. With respect to misspecification involving omitted cross-loadings, large proportions of the variances of TLI, BL89, CFI, and RMSEA were accounted for by misspecification. However, sample size effects were large for NFI, BL86, GFI, and AGFI in both cases. SRMR was found to be the least similar to other indices.

Overall, the authors found all indices apart from SRMR to be more sensitive to misspecifications due to omitted cross-loadings than those due to omitted factor covariances. They found TLI, BL89, CFI, and RMSEA to be sensitive to model misspecification (as quantified by the estimated statistical power for rejecting the misspecified model under a given sample size), and relatively insensitive to other factors, including estimation method and sample size. They recommend the use of these indices and suggest avoiding the use of the comparative indices NFI and BL86 (a Class 1 and Class 2 index, respectively). They also propose a two-index presentation in results reporting, coupling SRMR with one of the other recommended indices. This would allow researchers to present two indices that are most sensitive to two different types of misspecification.

Fan and Sivo (2005) sought to expand upon the research done by Hu and Bentler. They note that the original study fails to either quantify or control the severity of
misspecification across the two misspecification scenarios. They argue that before any conclusions can be drawn regarding index behavior, any confounding effects due to differences in misspecification severity must be addressed.

To determine the degree of misspecification present in the two scenarios in Hu and Bentler’s study, the authors computed a chi-square the degrees of freedom for each scenario by fitting the two misspecified models to population covariance matrices. They treated the obtained $\chi^2 - df$ values as estimates of noncentrality by assessing the degree to which these values differed from zero, with larger values indicating worse fit. Comparing the two $\chi^2 - df$ from the different models, they found differences in the values between the two misspecified models used in Hu and Bentler’s original study. This suggested that the results obtained from the original study may have been partly due to differences in the severity of misspecification between the two models as measured by an estimate of noncentrality.

To determine if this difference in misspecification severity was the cause of Hu and Bentler’s results, Fan and Sivo controlled for the severity of model misspecification in their study. This was accomplished by adjusting the population model parameters in Hu and Bentler’s original designs so that the two models would have comparable degrees of misspecification as measured by the estimated noncentrality parameter.

After controlling for the severity of misspecification, Fan and Sivo found index behavior patterns similar to those found in the original study. SRMR showed lower correlations with all other indices examined. TLI, BL89, RNI, CFI, and RMSEA all appeared more sensitive to omitted cross-loadings than omitted factor covariances. However, the authors make note that the difference in performance of SRMR may be due to the way in which the omitted factor covariance scenario was constructed. The misspecification resulted
from estimating a correlation between two factors as 1 when in the factors were orthogonal in
the population. Doing so allows for one or two misspecified parameters to result in a large
number of zeros in the model-implied covariance matrix. Of all the indices examined, SRMR
is most sensitive to this phenomenon, as it is most directly influenced by the discrepancy
between the model implied covariance matrix and that of the sample. The authors note that
such a result may not be generalizable beyond their study and do not support Hu and
Bentler’s two-indicator approach to reporting fit results.

The manipulations performed in the current study will allow for further investigation
as to whether SRMR’s performance as seen in these two studies can be generalized to the
point where Hu and Bentler’s two-index strategy may be considered a good procedure to
follow. Misspecification examined in the current study arises from different sources (omitted
error covariances and omitted latent factors) than those examined by Hu and Bentler. The
performance of SRMR in models involving these different sources of misspecification may
help to lend support to the suggestion that SRMR is sensitive to different types of
misspecification than the majority of other indices.

An additional study by Fan and Sivo (2007) sought to further examine index behavior
with respect to the degree of model misspecification. They argue that the usefulness of cutoff
values is dependent on what model features indices are sensitive to. Indices should be
sensitive to the severity of model misspecification but should not be sensitive to different
types of models containing the same degree of misspecification error. Index value should not
differ, for example, between a CFA model with six indicators and a structural model (a
model containing no latent variables) with twelve indicators, assuming the models contain
the same degree of misspecification.
The authors wished to examine what indices were sensitive to different model types, assuming all models in question had the same degree of misspecification, defined as having comparable statistical power for rejecting the misspecified models. To do so, they examined index behavior in two different CFA models. One model contained misspecification due to omitted cross-loadings. The other contained misspecification due to both omitted cross-loadings and omitted factor covariances.

Misspecification in each model was examined at three levels—no misspecification (the true models), a single instance of misspecification (e.g., one omitted factor covariance), and two instances of misspecification (e.g., two omitted cross-loadings). The degree of misspecification was held constant across model types for each level (we note here that the current study will take this form of misspecification to the extreme by examining situations where upwards of 10 error covariances are omitted in a fitted model).

Of the twelve indices studied by Fan and Sivo, NFI and SRMR were shown to be sensitive to model type, with 20% or more of their variation attributable to this factor. Gamma and RMSEA exhibited more desirable behavior. The majority of variation in these indices was due to model misspecification, while very little was due to model type and sample size. Gamma and RMSEA also appropriately accepted the true model and exhibited more power to reject increasingly misspecified models.

Finally, the authors also included an additional condition to the study by comparing index behavior in the original models to behavior in smaller but otherwise similar models. They found that in smaller models RMSEA was no longer insensitive to different model types. Only gamma remained unchanged across model sizes. Overall, the authors concluded that absolute indices (Classes 3 through 7) outperform those under the general class of
comparative indices (Classes 1 and 2) and were more exclusively sensitive to model misspecification and had greater power to correctly reject false models. These findings agree with research focused on model size and suggest, as previous studies do, that the number of indicators and factors plays a role in the behavior of RMSEA.

1.5.3.3 Summary

Given the results of previous studies, we may expect certain behaviors to be characteristic of the seven classes of indices defined and discussed above.

Previous research points to CFI and NNFI as being more stable across different model sizes when compared to indices such as GFI, RMSEA, and SRMR. However, this stability may in part depend on the type of misspecification, and the performance of CFI and NNFI may not be consistent across different estimation methods. These two indices have also been shown to be relatively unaffected by loading size, and the current study will aim to show whether these indices are relatively unaffected by other parameter sizes (such as factor correlation) as well.

Unlike CFI and NNFI, GFI is not independent of model size and shows worse fit for larger models, both for models with more indicators and for models with increasingly complex latent structures. While some studies have shown AGFI to perform more consistently across different model sizes, others have shown this index to behave like GFI, indicating worse fit as the size of the model increases. Such findings are problematic for AGFI, since the index was developed as an adjustment to GFI to compensate for the decrease in fit shown for larger models.

Since some research points to AGFI as performing similarly to GFI while other research suggests it adjusts for larger models as it was developed to do, it is unclear in what
situations we would expect GFI and AGFI to behave similarly and in what situations we would expect them to behave differently. Ideally, however, we would expect AGFI to fail to exhibit the decrease in fit shown by GFI for larger models.

Based on previous literature, we may expect RMSEA performance to be dependent on loading size, with higher loadings corresponding to increased fit. RMSEA has not been shown to be affected by model size, however, which suggests that this index may be ideal for comparing the relative fit of different models.

Finally, previous research has shown that SRMR may be more sensitive to different types of misspecification than the majority of other indices, which gives evidence to support the two index strategy of reporting as proposed by Hu and Bentler (1998). Thus, we may expect to arrive at different conclusions regarding the fit of a model if the misspecification in question is one for which SRMR is more sensitive and the other indices are not, or vice versa.

1.6 Methods

The goals of the current research are to study the behavior of seven fit indices defined in the population as a function of the amount of model misspecification (as defined by the size of the omitted parameter(s)) under different model scenarios. To limit the scope of the study, only CFA models are considered. Listed here again are the three main research questions of interest:

1. To what extent is the relationship between the amount of model misspecification, as defined by the size of the omitted parameter(s), and index value moderated by nuisance variables?

2. Does the current research support prior findings suggesting that uniform cutoff
values for indices may not be appropriate?

3. Can guidelines for the use of different indices under different model conditions be developed?

To address the first question, indices are evaluated for one- and two-factor CFA models involving one of two sources of misspecification. The first source of misspecification involves error covariances that are present in the true model but omitted in the fitted model (the latent structure is correctly specified). While other types of misspecification such as omitted cross-loadings and omitted factor correlations are possible in the context of CFA models, we restrict the first type of misspecification to that involving only omitted error covariances. This is done to allow for more focus on one type of misspecification that is relevant in both one-factor and two-factor models.

The second source of misspecification involves a misspecification in the latent structure of a model. Misspecifications of this sort can be considered more serious than those of the first, since misspecification is not due to an omitted pathway but rather to a failure to model the same number of latent factors as there are in the true model.

Within these two misspecification types, fit indices are evaluated as functions of different nuisance variables, including loading size, loading homogeneity, the size of the factor correlation when two or more factors are present, model complexity, and model balance. Model complexity is judged by the total number of indicators, the total number of factors, and the ratio of indicators to factors included in the model. Model balance is judged by the equal or unequal number of indicators per factor for two-factor models.

The effect of the degree of misspecification is the main variable of interest and is also evaluated. When the source of misspecification involves omitted error covariances, the
degree of misspecification is defined as either the size of the omitted error covariance or the total number of error covariances omitted in the model. When the source of misspecification involves the latent structure of a model, the degree of misspecification is defined as the difference between the number of latent factors in the fitted model and the number present in the true model.

To help address the second question, the results of this study are plotted using continuous curves. Most previous research has focused on examining index performance at only a few select values. Continuous curves show index value as a continuous function of whatever other variable (e.g., the size of an omitted error covariance) is of interest. This method of presentation was utilized by Savalei (2010) for research examining RMSEA performance under different modeling scenarios. Continuous curves allow for a clearer representation of index performance under the model components of interest.

In addition to the curves, flat lines indicating commonly used cutoff values are plotted for each index as well. This is done so that index behavior can be examined against these values as misspecification size and nuisance variables change. The most agreed upon cutoff value for CFI, NNFI, GFI, AGFI, and gamma is .95. Models with index values above .95 are accepted and said to fit the data well. Models with RMSEA values less than .06 and SRMR values less than .08 are also accepted as being well-fitting models. Index performance with respect to these values is examined in the conditions listed above.

All computations were performed using R. Misspecified models were created at varying values of the nuisance parameters in question (e.g., a model with a single error covariance was created for which loading sizes varied from .4 to .9). The covariance matrices of these models were then tested against the covariance matrix of data generated from the
“true” model, or the model without the misspecification (e.g., the model in which no error covariances are present). The ML fit function was used to obtain the minimized $\hat{F}$ values used as the measure of model misspecification. Several randomly chosen cases for each model scenario were verified using EQS 6.1. In cases where convergence problems were present, the outlier values were replaced by the average of the surrounding values.

Results are given in Chapters 2 through 4. To keep results organized by model and misspecification type, results are presented first for one-factor models involving error covariance misspecifications (Chapter 2), then for two-factor models involving error covariance misspecifications (Chapter 3), and finally for models involving a misspecified latent structure (Chapter 4).
Chapter 2: Results for One-Factor Models

The first scenarios examined are those involving a misspecification for a simple one-factor CFA model. The covariance structure for this model is given by $\Sigma = \lambda \lambda' + \Psi$, where $\lambda$ is a $p \times 1$ vector of factor loadings and $\Psi$ is the covariance matrix of the residuals, taken to be diagonal. It is assumed that the latent structure is not misspecified in any of these scenarios. Thus, the only possible source of misspecification is one or more omitted error covariances, meaning $\Psi$ is not diagonal in the population.

For one-factor models, the behavior of the seven index classes were examined as functions of loading size (and loading heterogeneity), the number of omitted error covariances, the size of the omitted error covariance(s), and the size of the model. While discussion involving all seven indices is present in the text, the similarities in behavior between CFI and NNFI, GFI and AGFI, and RMSEA and gammar led us to include only the plots for CFI, GFI, RMSEA, and SRMR in the figures for this section.

2.1 Effects of Loading Size

The first two figures presented involve manipulations of a one-factor model with eight indicator variables. They examine the effect of a single omitted error covariance on index values when the misspecification occurs between the first two indicator variables.

2.1.1 Homogeneous loadings

The plots of Figure 1 show the relationship between index value (plotted on the y-axis) and the size of a single omitted error covariance (on the x-axis). The six colored curves correspond to six loading sizes, with red, orange, green, blue, violet, and black corresponding to loadings of .4, .5, .6, .7, .8, and .9, respectively. All eight indicators in the model have the same loading size. These curves end at different points because the maximum value of the
omitted error covariance is a function of the loading size $1 - \lambda^2$. Values exceeding this lead to a nonpositive definite $\Sigma$ matrix corresponding to the true model. For all plots, a horizontal line is drawn to represent the commonly used cutoff criterion for each index.

Figure 1 makes it clear that the size of the loadings has a large effect on the values of RMSEA, gamma, GFI, and AGFI. These indices behave appropriately in the sense that they show worse fit as the size of the omitted error covariance increases. This is the case for all loading sizes, though the effect is more dramatic for higher loadings (.8, .9) than for lower loadings (.4, .5). When loadings are .9, RMSEA increases above the conventional cutoff value of .06 when the omitted error covariance is larger than .06. However, when loadings are .4, the omitted error covariance has to be larger than .28 for RMSEA to increase above .06. This behavior may indicate that these indices may be too sensitive or not sensitive enough to misspecification, depending on the size of the loadings and suggests that perhaps a universal cutoff value may not be appropriate across all loading sizes.

SRMR demonstrates a pattern similar to that of the indices above, showing worse fit as the size of the omitted error covariance increases. The only difference is its relative insensitivity to loading size for omitted error covariances less than .15. As the omitted error covariance increases past this value, there is an increased sensitivity to model misspecification as the size of the loadings increases.

Both CFI and NNFI show an unexpected, non-monotone curvilinear relationship between index value and the size of the omitted error covariance. These indices denote show worst fit for moderately-sized omitted error covariances (between about .2 and .6, depending on loading size) and best fit for small omitted error covariances and large omitted error covariances. These patterns go against what is expected—that indices should show worse fit
Figure 1. The plots of population fit index values vs. omitted error covariance for a 1-factor model with 8 indicators, separately for six values of factor loadings. Colors red, orange, green, blue, violet, and black correspond to loadings of .4, .5, .6, .7, .8, and .9, respectively. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
Figure 1 (continued). The plots of population fit index values vs. omitted error covariance for a 1-factor model with 8 indicators, separately for six values of factor loadings. Colors red, orange, green, blue, violet, and black correspond to loadings of .4, .5, .6, .7, .8, and .9, respectively. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
as misspecification increases—and thus are problematic for interpretation. CFI and NNFI indicate that a model with a smaller omitted error covariance can fit much worse than a model with a substantially larger misspecification. Such behavior suggests that these indices may not be appropriate to use when the source of misspecification is thought to arise from an omitted error covariance.

2.1.2 Heterogeneous loadings

The effect of loading heterogeneity on index value is examined next. Figure 2 replicates the same scenario as that in Figure 1, but adds two more conditions in which loading sizes are the same on average but vary across indicators. For all conditions, red, orange, green, blue, and violet curves correspond to (average) loadings of .4, .5, .6, .7, and .8. Loadings of .9 are omitted for readability. Solid lines are a reprint of Figure 1. In the first additional condition (dashed lines), six of the eight loadings depart .1 from the average loading size. In the second condition (dotted lines), six of the eight loadings depart .15 from the average.

Figure 2 shows that the effect of loading heterogeneity is not consistent. When loadings are large, GFI, AGFI, RMSEA, and gamma show worse fit when loadings are heterogeneous than when they are homogeneous. When loadings are small, heterogeneity improves fit over the homogeneous loadings case. As the departure from the average loading size increases, these differences are stronger. The differences grow larger amongst index values for the homogeneous and two heterogeneous cases as the size of the omitted error covariance increases. However, these differences are not so apparent around the respective cutoff values of these indices. This suggests that while loading heterogeneity affects index
Figure 2. The plots of population fit index values vs. omitted error covariance for a 1-factor model with 8 indicators, separately for five values of factor loadings. Colors red, orange, green, blue, and violet correspond to loadings of .4, .5, .6, .7, and .8, respectively. Solid lines correspond to equal loadings, dashed lines correspond to unequal loadings of +/- .1, dotted lines correspond to unequal loadings of +/- .15. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
Figure 2 (continued). The plots of population fit index values vs. omitted error covariance for a 1-factor model with 8 indicators, separately for five values of factor loadings. Colors red, orange, green, blue, and violet correspond to loadings of .4, .5, .6, .7, and .8, respectively. Solid lines correspond to equal loadings, dashed lines correspond to unequal loadings of +/- .1, dotted lines correspond to unequal loadings of +/- .15. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
values, it does not do so to the point that models with different degrees of loading heterogeneity will be rejected at substantially different degrees of misspecification.

SRMR is least affected by loading heterogeneity. For smaller values of omitted error covariance (less than .3), there is little to no difference in SRMR values between the homogeneous and heterogeneous scenarios. As the omitted error covariance increases above about .3, SRMR denotes worse fit for heterogeneous loadings than for homogeneous loadings.

The inflection points found in Figure 1 for CFI and NNFI are affected by loading heterogeneity. When loadings are small, lowest CFI and NNFI values occur at smaller values of omitted error covariance for heterogeneous loadings than for homogeneous loadings. When loadings are large, this pattern is reversed. This discrepancy is smaller for smaller loadings around the cutoff of .95, but for models with larger loadings overall, this discrepancy may lead to different decisions regarding model appropriateness depending on whether loadings are homogeneous or not.

### 2.1.3 Omitted error correlation

The non-monotone behavior of CFI and NNFI in the previous two figures prompted further exploration of the relationship between index value and the size of misspecification. Figure 3 examines the same relationship as shown in Figure 1, but on the correlation metric rather than the covariance metric. The maximum size of the omitted error correlation is one, and is not a function of loading size, allowing for a clearer examination of index behavior. Figure 3 shows a clear relationship between the size of the loadings and the point of inflection found for CFI and NNFI. The dip in index value occurs at a smaller omitted error correlations for smaller loading sizes. When loadings are .4, CFI is at its smallest value
Figure 3. The plots of population fit index values vs. omitted error correlation for a 1-factor model with 8 indicators, separately for six values of factor loadings. Colors red, orange, green, blue, violet, and black correspond to loadings of .4, .5, .6, .7, .8, and .9, respectively. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
Figure 3 (continued). The plots of population fit index values vs. omitted error correlation for a 1-factor model with 8 indicators, separately for six values of factor loadings. Colors red, orange, green, blue, violet, and black correspond to loadings of .4, .5, .6, .7, .8, and .9, respectively. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
(indicating worst fit) when the omitted error correlation is .56. When loadings are .9, CFI is smallest when the omitted error correlation is .96. While it was evident in Figure 1 that the size of the loadings had an effect on the inflection points of these indices, this relationship is made more obvious in Figure 3.

The effects of loading size are clearer for GFI, AGFI, and SRMR as well. Smaller loadings (.4, .5) correspond to a more rapid decrease in fit for larger values of omitted error correlation. There is little effect of loading size on the values of RMSEA and gamma when the omitted error correlation is small or moderate (less than about .4). The smaller effect of loading size in the correlation metric is an important finding, as it suggests that cutoff values for RMSEA and gamma can actually be interpreted across different loading sizes as long as the misspecification involves parameters measured on the correlation metric.

2.1.4 Continuous loadings

We wanted to gain another perspective on how indices behaved with respect to loading size. Figure 4 examines the same relationship as that in Figure 1, but plots index value (on the y-axis) against the value of continuous factor loadings (along the x-axis). Index behavior is examined at omitted error covariance sizes of .01, .05, .15, and .25, with colored curves red, orange, green, and blue corresponding to these values. As documented both by Browne and MacCallum (2002) and Savalei (2010), RMSEA values higher than .06 are produced for even small sizes of omitted error covariance when loadings are large. Figure 4 allows us to examine the behavior of RMSEA and the other indices for small values of omitted error covariance as loading size continuously increases.

The solid lines in Figure 4 represent the case in which all indicators have the same loading size. We examined the effects of individual loading sizes as well, since it is not
uncommon for researchers to be faced with one or two indicators that load much more
heavily onto factors than the other indicators in the model. It is also possible to encounter the
case where a significant number of variables load highly onto a factor while the others have
small to moderate loadings. Boundary solutions where an estimated loading takes on the
value of one (or .99) occur in practice (Savalei & Bentler, 2006). Because of this,
understanding the effects of having one or more really high loadings on index behavior is of
value. To examine these types of scenarios, Figure 4 includes a case in which the true model
contains a single high loading of .99 (dashed lines). In addition, another case was included in
which half (four) of the loadings are equal to .99 in the true model (dotted lines). In both
cases, the other loadings were varied continuously and the omitted error covariance occurred
between variables that did not have loadings of .99.

The solid lines in Figure 4 show that gamma, GFI, AGFI, and RMSEA behave
similarly to one another and demonstrate the same behavior as Savalei (2010) found for
RMSEA when all loadings were kept equal. Regardless of the size of the omitted error
covariance, index values rapidly change from showing good or moderate fit to showing
extremely poor fit as the loadings grow large (larger than .7). RMSEA values reach as high
as .25 and GFI values fall as low as .64 when loadings are .84 and the omitted error
covariance is .25. The pattern appears for all four sizes of omitted error covariance, but is
most dramatic for larger misspecifications. This behavior appears to be a function of the
loadings reaching their maximum value (this value being a function of the size of the omitted
error covariance and expressed as $\sqrt{1 - \text{residual covariance}}$).

The inclusion of a single loading of .99 (dashed lines) has very little effect on gamma,
GFI, AGFI, and RMSEA patterns when compared to the case in which there are no high
Figure 4. The plots of population fit index values vs. loading size for a 1-factor model with 8 indicators, separately for four values of a single omitted error covariance. Colors red, orange, green, and blue correspond to omitted error covariances of .01, .05, .15, and .25, respectively. Solid lines correspond to equal loadings, dashed lines correspond to the inclusion of one super-high loading (.99), dotted lines correspond to the inclusion of half of the variables having super-high loadings (.99). NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
Figure 4 (continued). The plots of population fit index values vs. loading size for a 1-factor model with 8 indicators, separately for four values of a single omitted error covariance. Colors red, orange, green, and blue correspond to omitted error covariances of .01, .05, .15, and .25, respectively. Solid lines correspond to equal loadings, dashed lines correspond to the inclusion of one super-high loading ($\lambda = .99$), dotted lines correspond to the inclusion of half of the variables having super-high loadings ($\lambda = .99$). NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
loadings present (solid lines). When half of the loadings are set to .99 (dotted lines), there
again is little change in the patterns of these four indices compared to the case involving no
high loadings. The exception is when the other half of the loadings are small (less than .1).
When this is the case, index values for larger omitted error covariances do not show perfect
fit when the other loadings are zero.

With no high loadings included (solid lines), SRMR behaves similarly to the indices
mentioned above. It indicates a stronger differentiation between the different sizes of omitted
error covariance when the loadings are low (around .1, .2). Values level off until the loadings
begin to increase to their maximum values. As loadings reach their maximum values, only
the largest omitted error covariance examined here (.25) is prone to the rapid change in index
value seen gamma, GFI, AGFI, and RMSEA.

SRMR also displays the strongest change when high loadings are included. When
only one loading is set to .99, index values increase rapidly from zero as the size of the other
loadings increase. As the other loadings approach sizes of .7, SRMR values begin to decrease
again. This is almost opposite of what occurs when there is no high loading present (in which
case, index values either remain constant or increase when the size of the omitted error
covariance is largest). In the case in which half of the loadings are .99, index values fail to
change across all other loading sizes and begin, like RMSEA, at higher values even when the
size of the other loadings are zero.

A nonlinear relationship between index value and the size of the loadings is present
both for CFI and NNFI in all three scenarios in Figure 4. These indices decrease rapidly for
small loadings, then begin to increase as loading size increases. Once loadings reach .6 or .7
(depending on the size of the omitted error covariance), index values again begin to drop.
This pattern is most dramatic for the highest value of omitted error covariance examined (.25). As in previous figures, the behaviors of CFI and NNFI suggest that researchers would have a difficult time appropriately judging fit without knowledge of these patterns. For example, when the size of the omitted error covariance is .15 and there are no super high loadings present (represented by the solid green line), CFI values are initially high for small loading sizes (indicating good fit), then drop as low as .85 when the size of the loadings increases to .2. As loadings continue to increase in size, CFI increases as well, indicating better fit. Once loadings increase past .8, values again fall below the commonly accepted cutoff value. The inclusion of a single high loading causes index values to drop more quickly and more dramatically for low loading sizes (0 to .2). That is, a single loading of .99 seems to exaggerate the curvilinear effect for low loadings, but reduces it when all loadings are higher. When half of the loadings are set to .99, this pattern disappears completely, and CFI and NNFI indicate good fit for all loading sizes until the loadings begin to reach their maximum values. At this point, index values drop dramatically.

### 2.1.5 Increasing number of large loadings

Figure 4 makes it clear that the behaviors of some indices are affected by the number of high loadings present. To gain a better understanding of this relationship, Figure 5 presents the case where index values (y-axis) are plotted against an increasing number of loadings of .99 (along the x-axis). This is done for a one-factor model with 12 indicators, where a single omitted error covariance occurs between indicators that do not have loadings of .99. Thus, since there are 12 indicators total, the largest number of loadings set to .99 is ten. Index value is examined at four values of omitted error covariance, with colored curves red, orange,
green, and blue corresponding to omitted error covariances of .01, .05, .15, and .25, respectively. All other loadings are set to .4.

We expected that all indices would show an improvement in fit in as the number of indicators with high loadings increased. This expectation was based partly on the fact that the remaining loadings in Figure 5 were set to .4. Because the omitted error covariance in this manipulation occurs between indicators that do not have high loadings, the indicators involving the misspecification would always have loadings of .4, while the loadings of the indicators that were independent of the misspecification would one by one increase to .99. This would mask the severity of the misspecification and lead to an improvement in overall fit.

However, this was not the pattern found for GFI, AGFI, RMSEA, gamma, and SRMR. Instead, Figure 5 reveals that these indices show a slight decrease in fit when going from the case with no super-high loadings to a case with a single loading of .99 included. These indices are insensitive to the inclusion of further high loadings, as there is very little change in index value past the inclusion of a single loading of .99. This may initially seem contradictory to Figure 4, which does show differentiation in index value for the three scenarios presented. For Figure 5, however, all other loadings are set to .4. It can be seen in Figure 4 that when loadings are of this value, there is very little difference in index value between the two scenarios involving loadings of .99.

CFI and NNFI show the pattern we expected. They exhibit an increasingly better fit as the number of loadings set at .99 increases. The improvement in fit is most rapid between the case involving no high loadings to the case involving a single loading of .99. Fit improves
Figure 5. The plots of population fit index values vs. the number of super-high loadings ($\lambda = .99$, from 0 to 10 super high loadings) for a 1-factor model with 12 indicators, done separately for four values of a single omitted error covariance. Colors red, orange, green, and blue correspond to omitted error covariance values of .01, .05, .15, and .25, respectively. All other loadings are set to .4. The omitted error covariance does not occur between variables with high loadings. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
Figure 5 (continued). The plots of population fit index values vs. the number of super-high loadings ($\lambda = .99$, from 0 to 10 super high loadings) for a 1-factor model with 12 indicators, done separately for four values of a single omitted error covariance. Colors red, orange, green, and blue correspond to omitted error covariance values of .01, .05, .15, and .25, respectively. All other loadings are set to .4. The omitted error covariance does not occur between variables with high loadings. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
much more gradually after three or more super-high loadings are included. This pattern is most dramatic for the highest level of omitted error covariance (.25).

These results suggest the ability of CFI and NNFI to detect a misspecification involving an omitted error covariance is decreased when indicators not involved with the misspecification have high loadings. For GFI, AGFI, RMSEA, gamma, and SRMR, on the other hand, the size of the loadings of indicators not involved in the misspecification has little effect on index sensitivity.

2.2 Effects of an Increasing Degree of Misspecification

2.2.1 Second omitted error covariance

In a model containing a large number of indicators, it is likely that there may be more than one error covariance present. Figure 6 presents a scenario in which there are two error covariances omitted in a one-factor model with eight indicators. Index value is plotted against the size of the second omitted error covariance (along the x-axis). The first omitted error covariance (between variables 1 and 2) is fixed to .2. This value was chosen because we wished to examine as many loading sizes as possible while still maintaining a reasonably sized omitted error covariance. An omitted covariance of .2 is only impossible with loadings of .9, so it was selected as an appropriate value as it allows for comparison of the five other loading sizes examined thus far in this paper. The five colored curves represent different loading sizes, with red, orange, green, blue, and violet corresponding to loadings of .4, .5, .6, .7, and .8, respectively. The first and second covariances either share a variable (when the second omitted error covariance is between variables 2 and 3, solid lines) or do not share a variable (when the second omitted error covariance is between variables 3 and 4, dashed lines).
Figure 6. The plots of population fit index values vs. a second omitted error covariance for a 1-factor model with 8 indicators, done separately for five values of factor loadings. Colors red, orange, green, blue, and violet correspond to loadings of .4, .5, .6, .7, and .8, respectively. The value of the first omitted error covariance is set to .2. Solid lines correspond to the situation in which the two error covariances share a variable; dashed lines correspond to the situation in which the two error covariances do not share a variable.

NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
Figure 6 (continued). The plots of population fit index values vs. a second omitted error covariance for a 1-factor model with 8 indicators, done separately for five values of factor loadings. Colors red, orange, green, blue, and violet correspond to loadings of .4, .5, .6, .7, and .8, respectively. The value of the first omitted error covariance is set to .2. Solid lines correspond to the situation in which the two error covariances share a variable; dashed lines correspond to the situation in which the two error covariances do not share a variable. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
Particularly when the second omitted error covariance is small (less than about .2), RMSEA, gamma, GFI, and AGFI show very little differentiation between the scenarios in which the first and second omitted error covariances share/do not share a variable, though slightly better fit is shown when the two omitted error covariances share a variable. As the size of the second omitted error covariance increases, there is stronger differentiation between the two cases. When loadings are small, index values show worse fit when the two omitted error covariances share a variable than when they do not. At larger loadings, worse fit is indicated when the omitted error covariances do not share a variable. This pattern is present until the second error covariance reaches its largest values for each loading size, at which point worse fit is shown when the two omitted error covariances share a variable.

For second omitted error covariance values less than about .25, SRMR does not differentiate between either different loading sizes or cases where the first and second omitted error covariances do and do not share a variable. As the size of the second omitted error covariances increases, the index is better able to differentiate between the shared/not shared variables cases. That is, better fit is shown for the case in which the two omitted error covariances share a variable. SRMR is also much less sensitive to detecting a second omitted error covariance when compared to the other indices examined. The smallest second omitted error covariance that produces SRMR values above the cutoff value of .08 is about .3. All other indices, depending on the size of the loadings, are sensitive enough to the first omitted error covariance (set at .2) that they indicate poor fit even when the second omitted error covariance is zero. This suggests, like previous figures, that SRMR is not as sensitive in detecting misspecifications involving omitted error covariances as the other indices are.
CFI and NNFI values show a nonlinear relationship with the size of the second omitted error covariance when the two covariances do not share a variable. In such cases, the same pattern is seen for the index values as was seen when there was a single omitted error covariance (Figure 1). Index values drop from above the accepted cutoff value of .95 as the second omitted error covariance values grow to a moderate size (approximately .3). They then begin to show better fit as the second omitted error covariance continues to increase. This phenomenon is particularly evident for small loadings. When the first and second omitted error covariances do share a variable, we finally see CFI and NNFI behaving linearly (at least for loadings greater than .4), indicating worse fit as the size of the second omitted error covariance increases.

2.2.2 Increasing number of omitted error covariances

Figure 7 presents further exploration of the relationship between index value and the number of omitted parameters. Index value is plotted against an increasing number of omitted error covariances (1 to 10) for a one-factor model with 20 indicators. The number of omitted error covariances is measured as a categorical variable, but the neighboring points in the figure have been connected for readability and to better capture the trends of the curves. Omitted error covariances occur between variables that do not share omitted covariances with other variables. That is, variables 1 and 2 share an omitted covariance, variables 3 and 4 share an omitted covariance, variables 5 and 6 share an omitted error covariance, etc. The size of the omitted error covariances is set to .2. Again, this value was chosen because it is large enough to be considered an omission that could be of concern in a model, but it is not so large that it cannot be attained at larger loadings. We also examined index behavior when the omitted error covariance was fixed to .4 and when it was fixed to .6. The same general
Figure 7. The plots of fit index values vs. the number of omitted error covariances (1 – 10) for a 1-factor model with 20 indicators, done separately for five values of factor loadings. Colors red, orange, green, blue, and violet correspond to loadings of .4, .5, .6, .7, and .8, respectively. The size of the omitted error covariances is set to .2. Neighboring points are connected for readability. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
Figure 7 (continued). The plots of fit index values vs. the number of omitted error covariances (1 – 10) for a 1-factor model with 20 indicators, done separately for five values of factor loadings. Colors red, orange, green, blue, and violet correspond to loadings of .4, .5, .6, .7, and .8, respectively. The size of the omitted error covariances is set to .2. Neighboring points are connected for readability. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
patterns of behavior as described below were found at these two values and these results are not presented. The five colored curves represent different loading sizes, with red, orange, green, blue, and violet corresponding to loadings of .4, .5, .6, .7, and .8, respectively.

All indices appropriately show a decrease in fit as more error covariances are present in the underlying data while the fitted model remains the same. Furthermore, this decrease in fit is linear in nature. All indices show approximately the same decrease in fit per each additional omitted error covariance, regardless of loading size. GFI, AGFI, GAMMA, and RMSEA values are all heavily influenced by loadings at all degrees of misspecification, with lower loadings corresponding to better fit overall. For example, when four error covariances are omitted and loadings are .4, GFI is about .98, above the acceptable cutoff value of .95. However, when the loadings are .8 for the same scenario, GFI falls to about .87.

While CFI and NNFI values are also affected by loading size, they are not affected in the same manner as the above mentioned indices. There is a progressive worsening of fit as the loading sizes increase for GFI, AGFI, GAMMA, and RMSEA. For example, when there are five omitted error covariances, GFI values of .97, .96, .95, .92, and .86 correspond to loadings of .4, .5, .6, .7, and .8, respectively. CFI and NNFI values are not related to loading size in the same manner. For these indices, poorest fit is indicated when loadings are .4 and .8 and best fit is indicated when loadings are .6 or .7. This pattern reflects that found for both CFI and NNFI in Figure 1 when the single omitted error covariance included in that manipulation is .2 (the value at which all error covariances are set in this scenario).

For SRMR, there appears to be virtually no effect of loading size on index value as the number of omitted error covariances increases. SRMR values at all loadings are so similar that the lines corresponding to the different loading sizes appear as one line in Figure 7. This
lends further support to SRMR being the index most independent of loading size of the indices studied here, at least when the size of the omitted error covariances are small (.2 in this Figure). SRMR also appears to be the least sensitive index to the number of omitted error covariances. Even when there are ten error covariances present in the data but omitted in the fitted model, SRMR fails to increase above the commonly used cutoff value of .08.

2.3 Model Size

Finally, it was of interest to examine the effects of the number of indicator variables included in the model. Figure 8 plots index value (y-axis) against an increasing number of indicator variables ($p = 4, 6, 8, 10, 12, 14, 16, 18,$ and 20, x-axis) for a one-factor model. The number of indicators is a categorical variable, but the neighboring points in the plots have been connected for readability. The colored curves correspond to different sizes of omitted error covariance, with red, orange, green, blue, and violet corresponding to omitted error covariances of .1, .2, .3, .4, and .5, respectively. Loadings are arbitrarily fixed to .4 (solid lines) or .7 (dashed lines).

Our expectations regarding model size were based largely on findings described in the literature. It has been shown that RMSEA indicates an improvement in fit as more indicators are added to the model (e.g., Breivik & Olsson, 2001; Kenny & McCoach, 2003), as the addition of more indicators “dilutes” the misspecification and masks its effects. However, recent research by Savalei (2010) has revealed a nonlinear relationship between RMSEA and the number of indicators in a model when misspecification is due to an omitted error covariance. We expected to find the same behavior for RMSEA here. Further, we predicted a different pattern for gamma, as it transforms the RMSEA to include a correction for the number of variables included in the model (Fan & Sivo, 2007). Joreskog and Sorbom
Figure 8. The plots of population fit index values vs. the number of indicators (p = 2, 4, 6, 8, 10, 12, 14, 16, 18, and 20) for a 1-factor model, done separately for five values of a single omitted error covariance. Colors red, orange, green, blue, and violet correspond to omitted error covariances of .1, .2, .3, .4, and .5, respectively. Loadings are .4 (solid lines) or .7 (dashed lines). Neighboring points are connected for readability. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
Figure 8 (continued). The plots of population fit index values vs. the number of indicators (p = 2, 4, 6, 8, 10, 12, 14, 16, 18, and 20) for a 1-factor model, done separately for five values of a single omitted error covariance. Colors red, orange, green, blue, and violet correspond to omitted error covariances of .1, .2, .3, .4, and .5, respectively. Loadings are .4 (solid lines) or .7 (dashed lines). Neighboring points are connected for readability. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
(1984), developers of GFI, acknowledged that GFI tended to increase (show better fit) as more variables were added to the model. Studies done since the development of GFI (e.g., Cudeck & Browne, 1983; MacCallum & Hong, 1997) have shown this to be the case. Joreskog and Sorbom developed the AGFI, and adjustment to GFI, that would not exhibit this behavior. We expected GFI to show better fit as the number of indicators increased, but had no prior predictions regarding the behavior of AGFI. Due to lack of prior research, no predictions are made for the remaining indices.

As seen in Figure 8, all indices demonstrate a nonlinear relationship between index value and the number of indicators included in the model. Apart from RMSEA, all indices showing good fit when $p = 4$ (particularly when loadings are .4), show worsening fit as the number of indicators initially increases, and then, once a certain model size has been reached, begin to gradually show an improvement in fit. This inflection point is dependent both on the size of the loadings and the size of the omitted error covariance. For example, when loading size is .7 and the omitted error covariance is .5, GFI values show worst fit when the model has 10 indicators. However, when the loading size is .4 and the omitted error covariance is .5, worst fit is indicated when there are 12 indicators.

While this pattern can also be seen in RMSEA for loadings of .4 combined with large omitted error covariance sizes, it is not nearly as pronounced as it is for the other indices. The nonlinear relationship is not seen for RMSEA when loadings are .7, for which fit improves steadily as more indicators are included in the model. SRMR again appears to be least affected by loading size, particularly when there are more indicators included in the model, and is least sensitive to an omitted error covariance overall. Only when the omitted error covariance is .4 or .5 do SRMR values increase past the commonly used cutoff value of .08.
2.4 Summary and Discussion

The figures in this section have explored the relationship between the population CFI, NNFI, GFI, AGFI, RMSEA, gamma, and SRMR indices and model misspecification in the context of one-factor CFA models. For all figures presented in Chapter 2, the behaviors of CFI and NNFI were highly similar. Because of this, only the behavior of CFI will be discussed in this section. GFI and AGFI also behaved very similarly in all figures, so only the results of GFI will be summarized. While RMSEA and gamma are interpreted differently in terms of what values indicate good model fit (higher gamma values indicate better fit, while lower RMSEA values indicate better fit), the behaviors of these indices were also similar enough across the different figures that we felt it appropriate to discuss only the results of RMSEA here. Thus, the four indices discussed in this section are CFI, GFI, RMSEA, and SRMR. The conclusions reached for CFI, GFI, and RMSEA can be applied to NNFI, AGFI, and gamma, respectively. The effects of loading size, the number of omitted error covariances, the size of the omitted error covariance(s), and the size of the model were examined to determine their effect on index values.

With respect to accurately reflecting an increasing degree of misspecification, GFI, RMSEA, and SRMR behave desirably. As the size of a single omitted error covariance increases, these indices show an increasingly worse fit for the model (Figure 1). CFI, on the other hand, displays a disturbing nonlinear relationship between index value and the size of the omitted error covariance. Not only can two considerably different omitted error covariances yield the same CFI value, but the pattern for this index shows that, regardless of loading size, better fit is shown for both small and large omitted error covariances than for moderate sizes of omitted error covariance.
Such a pattern is problematic, as it indicates that a model omitting a larger covariance can actually fit better than a model omitting a smaller covariance according to CFI. This pattern could easily lead researchers to draw the wrong conclusions about their model, either by leading them to accept a model as well-fitting when it omits an error covariance as large as .8, or by leading them to reject a model even though the size of the omitted error covariance is moderate. The presence of this disturbing curvilinear relationship between index value and the size of an omitted error covariance suggests that CFI should not be relied upon when it is expected that misspecification may be due to an omitted error covariance.

The size of the factor loadings was found to be influential for GFI, RMSEA, and SRMR, both when loadings are homogeneous (Figure 1) and when they are heterogeneous (Figure 2). SRMR is not influenced by loading size when the size of the omitted error covariance is small, but (like GFI and RMSEA) becomes more sensitive in detecting an omitted error covariance when loadings are large (.8, .9) as opposed to when they are of a moderate size (.4, .5). The sensitivity of these indices can be viewed as a good thing, as it suggests that higher loadings are beneficial when detecting misspecification involving an omitted error covariance. However, the large differences in sensitivity amongst the different loading sizes themselves may suggest that a universal cutoff value is not appropriate. For example, loadings of .4 and .7 yield RMSEA values of .02 and .1 respectively in Figure 1. Using the conventional cutoff value of .06, a model with loadings of .4 would be accepted as fitting the data while the same model with loadings of .7 would be rejected. Therefore, we recommend that researchers take note of the size of the estimated factor loadings in their models, perhaps by examining the average estimated loading as an indicator of the approximate size of the true loading.
CFI values are not consistently influenced by loading size as the size of the omitted error covariance increases. Plotting index value against an omitted error correlation (Figure 3) reveals that CFI denotes worse fit at smaller sizes of omitted error correlation for smaller loading sizes. Such a pattern could be viewed as another reason to not use this index in the context of misspecifications involving an omitted error covariance.

The exploration of index behavior at small values of omitted error covariance (Figure 4) revealed further evidence regarding the influence of loading size on index value. Findings by Browne and MacCallum (2002) and Savalei (2010) were replicated for RMSEA and newly established for the GFI. As the previous research has shown, high loadings lead to these indices becoming extremely sensitive to even very small omitted error covariances (e.g., .01, .05). This pattern is found for SRMR only for the largest omitted error covariance value studied (.25), indicating that SRMR is not as influenced by loading size as RMSEA and GFI are. The behavior of CFI with respect to small misspecifications lends further evidence to support abandoning the use of this index in contexts where misspecification is thought to arise from an omitted error covariance. CFI indicates poor fit when the size of the omitted error covariance is small, both when loadings are large (above .8) and when they are quite small (below .2), with best fit being shown when loadings are of a moderate size. Again, this finding suggests that interpretational problems may arise when using this index.

Further exploration showed that GFI and RMSEA are minimally influenced by the presence of variables with extremely high loadings, in the case where there is one variable with a high loading (Figure 4) or multiple (Figure 5). This is contrary to what was expected, as it was shown that even mild loading heterogeneity (Figure 2) had an effect on these indices for larger values of omitted error covariance. The inclusion of more and more
indicators with high loadings increased fit for CFI, a pattern we expected, as it was shown that when the size of the omitted error covariance is small, CFI denotes best fit when loadings are large.

The effects of increased misspecification were also explored in the form of additional omitted error covariances (Figures 6 and 7). With the addition of a second omitted error covariance, GFI, RMSEA, and SRMR show the expected and appropriate behavior of indicating worse fit as the size of the second omitted error covariance grows larger. Again, the effect of loading size is evident for GFI and RMSEA, as these indices are more sensitive to misspecification when loadings are large.

CFI values show a non-monotone relationship with the size of the second omitted error covariance (Figure 6). This relationship is present at all loading sizes when the two omitted error covariances do not share a variable. When they have a variable in common, however, the curvilinear pattern is only apparent only for small loadings (.4, .5). CFI behaves linearly for larger loadings, showing worse fit as the size of the second omitted error covariance increases. When loadings are small and the omitted error covariance is small to moderate, GFI, RMSEA, and SRMR indicate worse model fit if the two omitted error covariances do not share a variable. When loadings are large and the second omitted error covariance is large as well, worse fit is shown when the two omitted error covariances have a variable in common. The discrepancies in index values between the two scenarios (sharing/not sharing a variable) are small, but increase as the size of the second omitted error covariance increases.

This finding was unanticipated, but suggests that there may be a relationship between index value and the relative “isolation” of the misspecification for all indices examined.
Indices are most sensitive to misspecification if there is a combination of large loadings and a shared variable between the omitted error covariances, or a combination of small loadings and the omitted error covariances having no variable in common. If there exists a misspecification in which the omitted error covariances are “concentrated” around a variable (that is, they have a variable in common), indices are more sensitive to this when loadings are large, suggesting that they are sensitive to an isolated misspecification that involves a problematic variable.

All indices also show an appropriate decrease in fit as the number of error covariances present in the underlying data increases while the fitted model remains the same (Figure 7). This pattern is almost linear, particularly for smaller loading sizes. Knowledge of this behavior could be useful to researchers looking to isolate the sources of misspecification in their models. Knowing that every additional omitted error covariance brings about the same decline in fit for these indices, a researcher could examine index values after making changes to a model to determine whether additional sources of misspecification exist above and beyond those involving omitted error covariances.

Figure 7 also contains the strongest evidence that SRMR is relatively insensitive to misspecification involving omitted error covariances when compared to the other indices. Even when the degree of misspecification is severe (e.g., the fitted model omits 10 error covariances present in the underlying data structure), SRMR fails to indicate poor model fit. This suggests that when it is thought that the most likely sources of misspecification are due to one or more omitted error covariances (e.g., a typical scale validation scenario where the measured construct is presumed to be unidimensional but items could share specific variance), SRMR should not be exclusively relied upon. It also suggests that when
researchers are confronted with large SRMR values, it indicates that something is likely very 
wrong with their model in terms of the severity of misspecification arising from something 
other than an omitted error covariance (as SRMR values do not generally increase past the 
cutoff value of .08 when misspecification is due to omitted error covariances).

The size of the model as measured by the number of indicators was also shown to 
have an effect on indices’ ability to detect misspecification (Figure 8). Perhaps the most 
interesting finding given the previous literature is that all indices studied display a nonlinear 
relationship between index value and the number of indicators included in the model. AGFI 
and gamma, developed as adjustments to GFI and RMSEA respectively to compensate for 
behavior problems associated with model size, behave almost exactly as their parent indices 
do. As more indicators are added to a one-factor model involving a single omitted error 
covariance, indices in general show an initial worsening of fit, then begin to show an 
improvement in fit once a certain model size is reached.

This nonlinear behavior suggests that there may be a model size at which indices are 
maximally sensitive to misspecification, though this size depends on loading size as well as 
the severity of the misspecification. This is an unexpected finding and one that can be 
problematic for adequately judging model fit. If a one-factor model has an omitted error 
covariance of .5, according to GFI the model fits well if there are fewer than eight indicators 
or more than 14 indicators in the model. If the model includes 10 or 12 indicators, GFI 
indicates poor fit. Thus, by increasing or decreasing the size of a model by a few indicators, 
researchers can reach different conclusions regarding the fit of their model when using these 
indices. This information could be useful for researchers who have larger or smaller versions
of tests or scales available to use. Further exploration involving effects of model size on index value are carried out in later figures.
Chapter 3: Results for Two-Factor Models

The following scenarios involve a misspecification for a two-factor CFA model. Under a two-factor model, the covariance structure is given by $\Sigma = \Lambda \Phi \Lambda' + \Psi$, where $\Lambda$ is a $p \times k$ matrix of factor loadings, $\Phi$ is a $k \times k$ matrix of factor correlations, and $\Psi$ is the covariance matrix of residuals, taken to be diagonal. As in the one-factor model case, it is assumed that the latent structure is not misspecified. While both omitted error covariances ($\Psi$ is not diagonal in the population) and omitted factor loadings (the elements of $\Lambda$ that are zero in the fitted model are not zero in the population) are sources of misspecification under the two-factor model, explorations here are restricted to misspecifications only involving omitted error covariances. Previous research by Savalei (2010) found that when the source of misspecification involved an omitted cross-loading, RMSEA showed nonlinear behavior with respect to the number of the omitted cross-loadings. Because the present research involves the examination of more fit indices than just RMSEA and because strange behaviors from these indices have already been seen in one-factor model situations, attention was focused on index behavior in the context of misspecification arising from omitted error covariances.

As in the one-factor case, the results of NNFI, AGFI, and gamma were highly similar to those found for CFI, GFI, and RMSEA, respectively. For clarity, only the results of CFI, GFI, RMSEA, and SRMR will be discussed in-text and figures will include plots for CFI, GFI, RMSEA, and SRMR. For two-factor models, the behavior of the seven classes of fit indices will be examined as functions of loading size, the location of the omitted error covariance with respect to the two factors, the size of the factor correlation, model size, and model balance.
3.1 Parameter Size

The first five figures involve a two-factor model with an equal number of indicators per factor (the total number of indicators varies from figure to figure, as described in more detail below). These figures examine the effects of loading size, factor correlation size, and the location of the misspecification on index performance. Unless otherwise specified, the misspecification is a single omitted error covariance.

3.1.1 Location of omitted error covariance

The increased factor complexity of the two-factor models allows for additional forms of misspecification. Specifically, an omitted error covariance can be exclusive to variables loading onto one factor or it can involve variables from different factors. Figure 9 presents the same scenario as Figure 1, but for a two-factor model, and varying the location of the omitted error covariance. Index value (on the y-axis) is plotted against the size of a single omitted error covariance (x-axis) for a two-factor model with eight indicators. The omitted error covariance either occurs between variables loading onto the same factor (solid lines) or between variables loading onto different factors (dashed lines). The red, orange, green, blue, violet, and black curves correspond to loadings of .4, .5, .6, .7, .8, and .9, respectively. The correlation between the two factors is held at .4.

The curves in Figure 9 are similar in shape to the corresponding curves for the one-factor model scenario (Figure 1). That is, the increase in factorial complexity does not change the pattern of the relationships between index values and the size of the omitted error covariance. However, when the number of factors is two, all indices show a better fit overall than in the one-factor scenario. For example, in the one-factor case when loadings are .4 (red lines plotted in Figure 1), GFI falls below .95 when the omitted error covariance is .49. In the
Figure 9. Plots of fit index values vs. omitted error covariance for a 2-factor model for six values of factor loadings (.4 - .9). Colors red, orange, green, blue, violet, and black correspond to loadings of .4, .5, .6, .7, .8, and .9, respectively. Solid lines correspond to a model in which the omitted covariance occurs between indicators of the same factor, dashed lines correspond to a model in which the omitted covariance occurs between indicators of different factors. Total number of indicators is 8, factor correlation is .4. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
Figure 9 (continued). Plots of fit index values vs. omitted error covariance for a 2-factor model for six values of factor loadings (.4 -.9). Colors red, orange, green, blue, violet, and black correspond to loadings of .4, .5, .6, .7, .8, and .9, respectively. Solid lines correspond to a model in which the omitted covariance occurs between indicators of the same factor, dashed lines correspond to a model in which the omitted covariance occurs between indicators of different factors. Total number of indicators is 8, factor correlation is .4. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
two-factor case (solid red lines plotted in Figure 9), GFI fails to fall below .95 when the omitted error covariance occurs between indicators of the same factor, regardless of the loading size. This means that an omitted error covariance between indicators of the same factor will never be detected by the GFI (using traditional cutoffs) in the case of a 2-factor model with four indicators per factor.

All indices show worse fit when the omitted error covariance occurs between variables loading onto the same factors (solid lines) than when the omitted error covariance occurs between variables loading onto different factors (dashed lines). Particularly noticeable is the increased severity of the nonlinear relationship shown by CFI when the omitted error covariance occurs between indicators of different factors. It is also interesting to note that SRMR is less sensitive to the size of the loadings when the omitted error covariance occurs between indicators of different factors when compared to when the misspecification involves indicators of the same factor and is, overall, less sensitive in detecting the presence of an omitted error covariance than the other indices studied. This was seen in the one-factor model case as well.

3.1.2 Factor correlation

The addition of another factor in the model allows us to examine the possible effects of the size of the factor correlation on the relationship between fit indices and the size of the omitted error covariance. Figure 10 presents index value (on the y-axis) against the size of a single omitted error covariance (x-axis) occurring between variables that load onto the same factor (solid lines) or different factors (dashed lines) for a two-factor model with eight indicators. Loading size is fixed at .7 (other loading sizes were explored and it was found that smaller loading sizes resulted in even weaker dependence on factor correlation). The colored
Figure 10. Plots of fit index values vs. omitted error covariance for a 2-factor model for six values of factor correlation. Colors red, orange, green, blue, violet, and black correspond to factor correlations of .1, .2, .3, .4, .5, and .6, respectively. Solid lines correspond to a model in which the omitted covariance occurs between indicators of the same factor, dashed lines correspond to a model in which the omitted covariance occurs between indicators of different factors. Total number of indicators is 8. Loadings are .7. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
Figure 10 (continued). Plots of fit index values vs. omitted error covariance for a 2-factor model for six values of factor correlation. Colors red, orange, green, blue, violet, and black correspond to factor correlations of .1, .2, .3, .4, .5, and .6, respectively. Solid lines correspond to a model in which the omitted covariance occurs between indicators of the same factor, dashed lines correspond to a model in which the omitted covariance occurs between indicators of different factors. Total number of indicators is 8. Loadings are .7. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
curves correspond to different factor correlation sizes, with red, orange, green, blue, violet, and black corresponding to factor correlations of .1, .2, .3, .4, .5, and .6, respectively.

Figure 10 reveals that for all indices, larger factor correlations correspond to worse fit when compared to smaller factor correlations. This effect is small, however, particularly when the omitted error covariance occurs between variables loading onto different factors. CFI, for example, shows almost no difference in value between a model with a factor correlation of .1 and a model with a factor correlation of .6, regardless of the size of an omitted error covariance occurring between variables loading onto different factors (dashed lines).

What is most apparent in Figure 10 is the effect of the location of the omitted error covariance. Apart from SRMR, all indices show a much more rapid decrease in fit as the size of the omitted error covariance increases when it occurs between indicators of different factors than when it occurs between indicators of the same factor. For example, regardless of the size of the factor correlation or the size of the omitted error covariance, GFI values fail to fall below the commonly-accepted cutoff value of .95 when the omitted error covariance occurs between indicators of the same factor. When indicators of different factors are involved, however, GFI drops below .95 once the omitted error covariance becomes larger than about .29.

The exception to this pattern is SRMR. While worse fit is shown when the omitted error covariance occurs between indicators of different factors, SRMR values are not substantially different than when the omitted error covariance occurs between indicators of the same factor. For example, when factor correlation is .6 and the omitted error covariance is .5, SRMR is .072 when the omitted error covariance occurs between indicators of the same
factor and .081 when the omitted error covariance occurs between indicators of different factors.

3.2 Second Omitted Error Covariance

The following three figures examine index behavior with respect to a second omitted error covariance. Because there is the additional nuisance variable of factor correlation affecting index behavior in the two-factor case in addition to the location of the omitted error covariances in the one-factor case, we preface the following three figures with a brief description of the general model used in Figures 11, 12, and 13 as well as an explanation of the nuisance variables of interest here.

Figures 11, 12, and 13 all involve a two-factor model with eight indicators total. The first four indicators (1, 2, 3, and 4) load onto factor 1. The last four indicators (5, 6, 7, and 8) load onto factor 2. The correlation between the two factors is .4. In the case of a two-factor model, the influence of misspecification due to two (or more) omitted error covariances can depend on two things: the isolation of the misspecification and the location of the misspecification.

As in the one factor case, index behavior is expected to change depending on whether the two omitted error covariances share an indicator variable. We refer to this nuisance variable as one relating to the isolation of the misspecification. A more isolated misspecification occurs when both omitted error covariances share an indicator (e.g., the first omitted error covariance occurs between indicators 1 and 2, the second occurs between indicators 2 and 3). A less isolated misspecification occurs when both omitted error covariances do not share a indicator (e.g., the first omitted error covariance occurs between indicators 1 and 2, the second occurs between indicators 3 and 4).
An additional component that comes into play in the two-factor model is whether the omitted error covariance occurs between indicators of the same or different factors. We define this nuisance variable as relating to the location of the misspecification. A misspecification located on only one factor involves omitted error covariances that all load onto the same factor (e.g., the first omitted error covariance occurs between indicators 1 and 2, the second occurs between indicators 3 and 4). A misspecification located on both factors involves one (or both) omitted error covariances occurring between variables loading onto different factors (e.g., the first omitted error covariance occurs between indicators 1 and 2, the second omitted error covariance occurs between indicators 4 and 5).

The effects of the combination of the isolation and location of misspecification involving two omitted error covariances are explored in Figures 11, 12, and 13. Figure 11 involves the case where misspecification is located within one factor. Figure 12 involves the case where the first omitted error covariance involves variables of the same factor the second involves indicators loading onto both factors. Figure 13 involves the case where both omitted error covariances involve indicators loading onto both factors. For each figure, the isolation of misspecification (whether the two omitted error covariances share or do not share an indicator) is explored as well. Please refer to the three Illustrations presented on the following pages for models corresponding to Figures 11, 12, and 13 for further clarification.

Before discussing the results found in Figures 11, 12, and 13, we first present several expectations regarding index behavior in these figures. Because indices are assessments of global model fit, it is expected that the more isolated a misspecification is (the more indicator variables are shared between the two omitted error covariances), the less the misspecification affects global fit. Thus, it is anticipated that indices will show better fit when the two omitted
Illustration 1. Model diagram for Figure 11.

Solid lines

Dashed lines
Illustration 2. Model diagram for Figure 12.

**Solid lines**

![Solid lines diagram]

**Dashed lines**

![Dashed lines diagram]
Illustration 3. Model diagram for Figure 13.
error covariances share a variable when compared to situations where they don’t. We also anticipate that when both the loadings and the size of the second omitted error covariance are small, indices will show best fit when the two omitted error covariances share a variable. This was seen in Figure 6, which examined the effect of the omitted error covariances sharing/not sharing a variable in a one-factor model. When both the loadings and the size of the second omitted error covariance are large, we anticipate indices will show best fit when the two omitted error covariances do not share a variable.

Figures 11, 12, and 13 are now discussed in detail. Figure 11 presents the case in which there are two omitted error covariances in a two-factor model with 8 indicators. Both omitted error covariances occur between indicators loading onto the first factor. Solid lines correspond to the case where the two omitted error covariances share an indicator (the first omitted error covariance involves indicators 1 and 2, the second involves indicators 2 and 3). Dashed lines correspond to the case where the two omitted error covariances do not share an indicator but still occur between indicators of the same factor (the first omitted error covariance involves indicators 1 and 2, the second involves indicators 3 and 4). Index value (on the y-axis) is plotted against the size of the second omitted error covariance (on the x-axis). This is done for five values of factor loadings, with colored curves red, orange, green, blue, and violet corresponding to loadings of .4, .5, .6, .7, and .8, respectively. The first omitted error covariance is set to .2.

As was anticipated, the patterns in Figure 11 are generally similar to those in Figure 6. The nonlinear relationship between the size of the second omitted error covariance and the values of CFI is still present when the omitted error covariances do not share an indicator,
Figure 11. Plots of fit index values vs. a second omitted error covariance for a 2-factor model with 8 indicators. The first omitted error covariance is set to .2. Done when the omitted error covariances occur within one factor. Colors red, orange, green, blue, and violet correspond to factor loadings of .4, .5, .6, .7, and .8, respectively. Solid lines correspond to the situation in which the two omitted error covariances share a variable, dashed lines correspond to the situation in which the two omitted error covariances do not share a variable. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
Figure 11 (continued). Plots of fit index values vs. a second omitted error covariance for a 2-factor model with 8 indicators. The first omitted error covariance is set to .2. Done when the omitted error covariances occur within one factor. Colors red, orange, green, blue, and violet correspond to factor loadings of .4, .5, .6, .7, and .8, respectively. Solid lines correspond to the situation in which the two omitted error covariances share a variable, dashed lines correspond to the situation in which the two omitted error covariances do not share a variable. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
and GFI and RMSEA indicate worse fit as the size of the second omitted error covariance increases, as they did in the one-factor model case (Figure 6). There are larger differences in index values between the case where the omitted error covariances share an indicator (solid lines) and the case where they do not (dashed lines). For example, while GFI values for the two different curves (solid and dashed) remained similar in value to each other as the size of the second omitted error covariance increased in Figure 6, GFI becomes more sensitive to misspecification and diverge quite rapidly when the two omitted error covariances do not share a variable in Figure 11. This pattern may suggest that these indices (GFI and RMSEA) are more sensitive to misspecification when the misspecification is spread out across more indicators (that is, when multiple omitted error covariances do not share an indicator).

Unlike in the one-factor model scenario presented in Figure 6, Figure 11 shows SRMR to be much more sensitive to misspecification when the two omitted error covariances share an indicator than when they do not. In Figure 6, SRMR values for the case involving shared indicators (solid lines) and the case involving unshared indicators (dashed lines) were fairly similar, regardless of the size of the second omitted error covariance. In Figure 11, the values quickly diverge as the second omitted error covariance increases. When loadings are .7 and the omitted error covariances share an indicator (solid blue curve), SRMR is as high as .21 when the second omitted error covariance is .46. At the same loading size, SRMR is only .08 when the second omitted error covariance is .46 and the omitted error covariances do not share an indicator (dashed blue curve).

Figure 12 involves a similar scenario to that studied in Figure 11, but the misspecification is not located within one single factor. In Figure 12, the first omitted error covariance occurs between indicators of the same factor while the second omitted error
Figure 12. Plots of fit index values vs. a second omitted error covariance for a 2-factor model with 8 indicators. The first omitted error covariance is set to .2. Done when the omitted error covariances involve indicators from both factors. Colors red, orange, green, blue, and violet correspond to factor loadings of .4, .5, .6, .7, and .8, respectively. Solid lines correspond to the situation in which the two omitted error covariances share a variable; dashed lines correspond to the situation in which the two omitted error covariances do not share a variable. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
Figure 12 (continued). Plots of fit index values vs. a second omitted error covariance for a 2-factor model with 8 indicators. The first omitted error covariance is set to .2. Done when the omitted error covariances involve indicators from both factors. Colors red, orange, green, blue, and violet correspond to factor loadings of .4, .5, .6, .7, and .8, respectively. Solid lines correspond to the situation in which the two omitted error covariances share a variable; dashed lines correspond to the situation in which the two omitted error covariances do not share a variable. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
covariance occurs between indicators of different factors. Solid lines correspond to the case where the two omitted error covariances share an indicator (the first omitted error covariance involves indicators 3 and 4, the second involves indicators 4 and 5). Dashed lines correspond to the case where the two omitted error covariances do not share an indicator (the first omitted error covariance involves indicators 1 and 2, the second involves indicators 4 and 5). Index value (on the y-axis) is plotted against the size of the second omitted error covariance (on the x-axis). This is done for five values of factor loadings, with colored curves red, orange, green, blue, and violet corresponding to loadings of .4, .5, .6, .7, and .8, respectively. The first omitted error covariance is set to .2.

In Figure 12, the same general nonlinear patterns are seen for CFI as were seen in Figure 11. This may suggest that the location of misspecification (that is, whether it involves indicators of one or both factors) does not influence the pattern of CFI, at least when misspecification is due to multiple omitted error covariances. However, index values on the whole (for CFI as well as the other indices) show worse fit in Figure 12 than in Figure 11, suggesting that indices are generally more sensitive to misspecification if it is “spread out” across factors rather than if it is isolated within one.

Compared to Figure 11, there are larger discrepancies in Figure 12 between the case where the omitted error covariances share an indicator (solid lines) and the case where they do not (dashed lines) for GFI and RMSEA. When the misspecification was isolated to indicators loading onto only one of the two factors (Figure 11), these indices were more sensitive to misspecification when the two omitted error covariances did not share an indicator. In Figure 12, where the misspecification involves indicators from both factors, GFI and RMSEA are more sensitive to misspecification when the omitted error covariances do
share an indicator. This might suggest that the sensitivity of GFI and RMSEA to the number of indicators involved in the misspecification is altered by the location of those indicators with respect to the factors.

SRMR shows less differentiation in Figure 12 between the case where the two omitted error covariances share an indicator (solid lines) and case where they do not (dashed lines) when compared to Figure 11. This indicates that, like RMSEA and GFI, the location of the misspecification with respect to the factors might play a role in the sensitivity of SRMR to the number of indicators involved in the misspecification.

Finally, Figure 13 presents the last misspecification involving two omitted error covariances. While Figure 11 involved two omitted error covariances located within one factor and Figure 12 involved two omitted error covariances where one was located within one factor and the other involved indicators of both factors, Figure 13 presents the case where both omitted error covariances involve indicators from both factors. In Figure 13, solid lines correspond to the case where the two omitted error covariances share an indicator (the first omitted error covariance involves indicators 4 and 5, the second involves indicators 4 and 6). Dashed lines correspond to the case where the two omitted error covariances do not share an indicator (the first omitted error covariance involves indicators 4 and 5, the second involves indicators 3 and 6). Index value (on the y-axis) is plotted against the size of the second omitted error covariance (on the x-axis). This is done for five values of factor loadings, with colored curves red, orange, green, blue, and violet corresponding to loadings of .4, .5, .6, .7, and .8, respectively. The first omitted error covariance is set to .2.

A comparison between Figures 12 and 13 shows that all indices show worse fit when misspecification is spread out across factors rather than if it is isolated within one. That is,
Figure 13. Plots of fit index values vs. a second omitted error covariance for a 2-factor model with 8 indicators. The first omitted error covariance is set to .2. Done when the omitted error covariances involve indicators from both factors. Colors red, orange, green, blue, and violet correspond to factor loadings of .4, .5, .6, .7, and .8, respectively. Solid lines correspond to the situation in which the two omitted error covariances share a variable in one factor; dashed lines correspond to the situation in which the two omitted error covariances do not share a variable. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
Figure 13 (continued). Plots of fit index values vs. a second omitted error covariance for a 2-factor model with 8 indicators. The first omitted error covariance is set to .2. Done when the omitted error covariances involve indicators from both factors. Colors red, orange, green, blue, and violet correspond to factor loadings of .4, .5, .6, .7, and .8, respectively. Solid lines correspond to the situation in which the two omitted error covariances share a variable; dashed lines correspond to the situation in which the two omitted error covariances do not share a variable. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
worse fit is shown in Figure 13, when both omitted error covariances involve indicators loading onto both factors, than in Figure 12, where only one of the omitted error covariances involved indicators loading onto both factors. Comparing Figures 11, 12, and 13, then, there is a trend of decreasing fit as the misspecification becomes more “spread out” across both factors.

In general, the same patterns observed for all indices in Figure 13 were similar to those in Figure 12. An interesting pattern to note across all three Figures (11, 12, and 13), however, is that worse fit is shown in Figures 11 and 13 for cases where the two omitted error covariances did not share an indicator than when the two omitted error covariances did share an indicator. In Figure 12, however, this pattern is reversed, and worse fit is shown when the two omitted error covariance share an indicator. This is an interesting finding and will be discussed further in the conclusions below.

3.3 Model Size

We also examined the effect of model size on fit indices in the context of a two-factor model. Figure 14 presents the effect of the number of indicators on index value, as Figure 8 did for the 1-factor model. Index value (on the y-axis) is plotted against an increasing number of indicators ($p = 4, 6, 8, 10, 12, 14, 16, 18, \text{ and } 20$). Only even numbers of indicators are included to avoid having an unequal number of indicators per factor (this scenario is explored later in Figures 15 and 16). The number of indicators is a categorical variable, but adjacent points are connected in Figure 14 for readability. The size of the single omitted error covariance is now represented by the different colored curves, with red, orange, green, blue, and violet corresponding to omitted error covariance sizes of .1, .2, .3, .4, and .5,
Figure 14. Plots of fit index values vs. an increasing number of indicators ($p = 2, 4, 6, 8, 10, 12, 14, 16, 18, \text{ and } 20$) for a 2-factor model. Done for five values of omitted error covariance (.1 to .5). Colors red, orange, green, blue, and violet correspond to omitted error covariances of .1, .2, .3, .4, and .5, respectively. Neighboring points are connected for readability. Solid lines correspond to factor loadings of .4, dashed lines correspond to factor loadings of .7. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
Figure 14 (continued). Plots of fit index values vs. an increasing number of indicators \((p = 2, 4, 6, 8, 10, 12, 14, 16, 18, \text{ and } 20)\) for a 2-factor model. Done for five values of omitted error covariance (.1 to .5). Colors red, orange, green, blue, and violet correspond to omitted error covariances of .1, .2, .3, .4, and .5, respectively. Neighboring points are connected for readability. Solid lines correspond to factor loadings of .4, dashed lines correspond to factor loadings of .7. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
respectively. Loadings are arbitrarily fixed to .4 (solid lines) or .7 (dashed lines). Factor correlation is set to .1.

Compared to the one-factor model examined in Figure 8, there is additional model complexity brought on by the inclusion of a second factor in the model for Figure 14. This added complexity alters index behavior when compared to Figure 8. The non-monotone relationship between the size of the model and CFI values is less dramatic when there are two factors in the model (Figure 14) as opposed to one (Figure 8). That is, the inflection point at which these indices cease showing worse and begin to show better fit occurs for larger model sizes when two factors are present in the model. The non-monotone relationship actually disappears for the largest value of omitted error covariance (.5) and index values consistently show worsening fit as the size of the model increases. The non-monotone relationship between index value and the size of the model is also greatly diminished for GFI in Figure 14 (when compared to patterns found in Figure 8), but only when loadings are small (.4). Both RMSEA and SRMR show a non-monotone relationship between index values and the size of the model. The pattern is present both when loadings are .4 and when loadings are .7. As the number of indicators increases, RMSEA and SRMR initially show a worsening of fit. Depending on the size of the loadings and the size of the omitted error covariance, both indices begin to show an improvement in fit after reaching a certain model size. For example, when the omitted error covariance is .3 and the loadings are .7, RMSEA shows an increasingly worse fit until there are 10 indicators in the model. At this point, RMSEA values begin to show an improvement in fit as the model grows to a total of 20 indicators.

Because the equation for the population RMSEA includes dividing the minimized fit function by the degrees of freedom (see Equation 9), we also looked at the plot of the square
root of the minimized fit function against an increasing number of indicators. It was found that, while RMSEA exhibits a non-monotone relationship with the size of the model, the fit function itself increases as more indicators are included in the model, regardless of loading size or the size of the omitted error covariance, and then begins to level off at larger model sizes. This pattern suggests that nonlinear effect shown in Figure 14 for RMSEA (a worsening of fit as model size increases, then an improvement in fit after a certain model size is reached) is due to dividing the minimized fit function by the degrees of freedom and is not due to any nonlinear behavior in the fit function itself.

The degrees of freedom change with model size. The degrees of freedom are calculated as \( \frac{p(p+1)}{2} - q \). Thus, for models with \( p = 4, 6, 8, 10, 12, 14, 16, 18, \) and \( 20 \) (the model sizes explored in Figure 14), the corresponding degrees of freedom are \( df = 1, 8, 19, 34, 53, 76, 103, 134, \) and \( 169 \). Plotting the number of variables against the degrees of freedom revealed that as model size increases, the degrees of freedom increase in an almost exponential pattern. By plotting the square roots of the degrees of freedom, a linear pattern is seen, with \( \sqrt{df} \) values increasing steadily as the size of the model increases. Comparing this plot to that of \( \sqrt{F} \), or the numerator of the RMSEA equation, allows for an explanation as to why the nonlinear patterns may be present in Figure 14. Depending on loading size and factor correlation size, values of \( \sqrt{F} \) increase and then begin to level off at a certain model size. However, \( \sqrt{df} \) values continue to increase as model size increases. The change in ratio of \( \sqrt{F} \) to \( \sqrt{df} \) that occurs when \( \sqrt{F} \) values begin to level off corresponds to the inflection points seen in Figure 14. Thus, it could be said that the difference in rate of change between
$\sqrt{F}$ and $\sqrt{dF}$ as model size increases is responsible for RMSEA’s nonlinear pattern. This ratio does not explain, however, the nonlinear pattern observed for the other indices.

Finally, Figure 14 shows that for both RMSEA an SRMR, the smaller the size of the omitted error covariance, the smaller the number of indicators is that corresponds to the inflection point, or the point at which indices cease showing a worsening of fit and begin to show an improvement of fit as model size increases. While these two indices display similar patterns, it should be noted that only when the size of the omitted error covariance is greatest (.5) and the loadings are .7 do SRMR values increase above the cutoff value of .08.

### 3.4 Model Balance

We now turn to model balance. We define a balanced model as one having an equal number of indicators per factor (e.g., a balanced two-factor model with $p = 16$ has eight indicators per factor). Studies in the literature that examine index behavior in the context of multi-factor models involve balanced models almost exclusively. In the rare case when an imbalanced model is used, the imbalance usually involves only one additional indicator loading onto a factor while the second factor has one fewer indicators. In such cases, the unequal number of indicators per factor is noted but its effects are never examined.

In the following four figures, Figures 15, 16, 17, and 18, the effects of model imbalance on index behavior are explored in two-factor models with a single omitted error covariance. The true models for these Figures have an unequal number of indicators per factor, and the fitted models accurately model this imbalance. That is, misspecification arises only from the omitted error covariance, not from the fitted models incorrectly representing the imbalance in the true model.
Before discussing the individual figures in detail, a brief discussion of what is of interest for each figure is provided. To examine the effect of model imbalance on index behavior in the simplest case, Figure 15 examines a balanced and five increasingly imbalanced two-factor models when the omitted error covariance occurs between indices loading onto the larger factor (the factor with more indicators). To explore whether the location of the misspecification also plays a role on index value in the context of imbalanced models, Figure 16 expands on Figure 15 and includes the case where the misspecification occurs between indicators loading onto the smaller factor. In Figure 17, the effects of model size are explored by holding one factor at a constant size and varying the size of the other. Finally, in Figure 18, the ratio of the size of the two factors is explored as to whether it has an effect on index value.

The initial explorations of the effects of model imbalance are done with few preconceived assumptions regarding how indices will behave. However, since all indices showed increased sensitivity to misspecification when two omitted error covariances were isolated to indicators of a single factor (Figure 11), it may be expected that indices will show better fit for more imbalanced models, as the single omitted error covariance here becomes more “isolated” to one of the factors the more the model becomes imbalanced.

The results of Figures 15, 16, 17, and 18 are now discussed in this and the following two sections. Figure 15 presents index value (on the y-axis) against the value of an omitted error covariance (x-axis) for a two factor model with a total of 24 indicators, but different number of indicators per factor. The colored curves correspond to a balanced true model and true models with varying degrees of imbalance. Red corresponds to a balanced model with 12 indicators per factor. Orange corresponds to a model with 11 and 13 indicators per factor,
Figure 15. Plots of fit index values vs. omitted error covariance for a two-factor model with 24 indicators total. Colors red, orange, green, blue, violet, and black correspond to a balanced model (12 indicators per factor) and five increasingly imbalanced models (11 and 13, 10 and 14, 9 and 15, 8 and 16, and 7 and 17 indicators per factor), respectively. Solid lines correspond to loadings of .4, dashed lines correspond to loadings of .7. Factor correlation is set at .1. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
Figure 15 (continued). Plots of fit index values vs. omitted error covariance for a two-factor model with 24 indicators total. Colors red, orange, green, blue, violet, and black correspond to a balanced model (12 indicators per factor) and five increasingly imbalanced models (11 and 13, 10 and 14, 9 and 15, 8 and 16, and 7 and 17 indicators per factor), respectively. Solid lines correspond to loadings of .4, dashed lines correspond to loadings of .7. Factor correlation is set at .1. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
green corresponds to a model with 10 and 14 indicators per factor, blue corresponds to a model with 9 and 15 indicators per factor, purple corresponds to a model with 8 and 16 indicators per factor, and black corresponds to a model with 7 and 17 indicators per factor. Factor correlation is set to .1 and loadings are either .4 (solid lines) or .7 (dashed lines)\(^2\). For all models, the single omitted error covariance occurs between indicators that load onto the factor with the greater amount of indicators.

Figure 15 reveals that model imbalance has an effect on all indices examined when the omitted error covariance occurs between indicators loading on the larger factor. In all cases when loadings are .4, indices show best fit for the most severely imbalanced model (violet curves) and increasingly poor fit as the model becomes more balanced. These effects increase as the omitted error covariance increases. For example, when the omitted error covariance is .3, there is very little difference in GFI values for the different degrees of model balance (the different colored curves). When the omitted error covariance increases to .6, the effect of model balance could actually affect a researcher’s decision regarding the adequacy of a model’s fit. When loadings are .4 and the omitted error covariance is .6, the GFI for the balanced model (solid red curve) is .94. At the same omitted error covariance size and loading size, a model with a moderate degree of imbalance (10 and 14 indicators per factor, solid green curve) has a GFI of .96. Utilizing the commonly accepted cutoff criterion of .95, a researcher would reject the balanced model as fitting poorly but accept the imbalanced model, even though the size of the misspecification is the same for both.

\[\text{2Other values of factor correlation were explored and we found no substantial difference in index behavior to warrant including the results of this manipulation in this thesis.}\]
When loadings are .7, the relationship between model balance and index value is not consistent across indices. CFI and RMSEA show the same relationship for loadings of .7 as the one for loadings of .4: best fit is shown for the most imbalanced model, and worst fit is shown for the balanced model. For GFI and SRMR, this relationship does not hold when loadings are .7. It appears that worst fit is denoted for the least imbalanced model (11 and 13 indicators per factor) and improves as the model becomes more imbalanced. The balanced model is shown to have better fit than most of the imbalanced models. This phenomenon occurs at the larger values of omitted error covariance for loadings of .7.

### 3.4.1 Location of misspecification

Explored next are the effects of the location of the omitted error covariance on index behavior in the context of imbalanced models. In Figure 15, the omitted error covariance was located between variables loading onto the “larger” factor. That is, the misspecification was always isolated to the factor with the most indicators. Figure 16 reprints Figure 15 at loadings of .4 (represented by solid lines) and includes the same misspecification for the case in which the omitted error covariance occurs between variables that load onto the smaller factor (represented by dashed lines). For instance, the solid violet curve is the case in which the single omitted error covariance occurs between variables loading onto the factor with \( p = 17 \), and the dashed violet curve is the case in which the single omitted error covariance occurs between variables loading onto the factor with \( p = 7 \). The balanced model in this Figure is represented by a bold red line, as the location of the misspecification does not matter when the factors have an equal number of indicators.

Figure 16 makes it clear that the location of the misspecification has an effect on all index values, though the effect is only readily apparent when the omitted error covariance
Figure 16. Plots of fit index values vs. omitted error covariance for a two-factor model with 24 indicators total. Red, orange, green, blue, violet, and black lines correspond to a balanced model (12 indicators per factor) and five increasingly imbalanced models (11 and 13, 10 and 14, 9 and 15, 8 and 16, and 7 and 17 indicators per factor), respectively. Solid lines correspond to the misspecification occurring within the large factor, dashed lines correspond to the misspecification occurring within the small factor. Asterisks denote the factor with the misspecification. Loadings are set to .4, factor correlation is .1. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
Figure 16 (continued). Plots of fit index values vs. omitted error covariance for a two-factor model with 24 indicators total. Red, orange, green, blue, violet, and black lines correspond to a balanced model (12 indicators per factor) and five increasingly imbalanced models (11 and 13, 10 and 14, 9 and 15, 8 and 16, and 7 and 17 indicators per factor), respectively. Solid lines correspond to the misspecification occurring within the large factor, dashed lines correspond to the misspecification occurring within the small factor. Asterisks denote the factor with the misspecification. Loadings are set to .4, factor correlation is .1. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
increases above a moderate size (about .5). When the omitted error covariance occurs between variables loading onto the larger factor, fit improves as the model goes from balanced (red curve) to most severely imbalanced (solid violet curve). However, when the omitted error covariance occurs between variables that load onto the smaller factor, fit worsens as the model goes from balanced to the most severely imbalanced (dashed violet curve). The location of the omitted error covariance can play a rather dramatic role in index value for an imbalanced model. For example, consider a two-factor model with 10 indicators loading onto one factor and 14 indicators loading onto the other factor. If an omitted error covariance of .6 occurred between variables loading onto the larger factor, the GFI for the model would be .96 (solid green curve). If, in the same model, the omitted error covariance occurred between variables loading onto the smaller factor, the GFI for the model would be .94 (dashed green curve). This pattern is evident for all indices regardless of loading size and suggests that perhaps the differences in fit value found between the balanced and imbalanced models has more to do with the size of the factor containing the misspecification than with the model balance itself.

3.4.2 Factor size and misspecification

Figure 16 revealed that in an imbalanced model, the larger the factor containing an omitted error covariance, the less sensitive indices are to misspecification. To further explore this phenomenon and its relationship to model size, the case where the size of the factor containing the omitted error covariance is held constant while the size of the model as a whole changes is examined next.

Figure 17 plots index value (on the y-axis) against an increasing number of indicators (x-axis) for an imbalanced two-factor model. The size of the first factor is held at \( p = 6 \) for all
Figure 17. Plots of fit index values vs. an imbalanced two-factor model with an increasing number of indicators. The size of the first factor is held at $p = 6$, such that models with $p = 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23,$ and $24$ have a second factor of size $p = 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17,$ and $18$, respectively. Colors red, orange, green, blue, and violet correspond to omitted error covariances of $.4, .5, .6, .7,$ and $.8$, respectively. Factor correlation is held at $.1$, loadings are held at $.4$. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
Figure 17 (continued). Plots of fit index values vs. an imbalanced two-factor model with an increasing number of indicators. The size of the first factor is held at $p = 6$, such that models with $p = 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, \text{ and } 24$ have a second factor of size $p = 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, \text{ and } 18$, respectively. Colors red, orange, green, blue, and violet correspond to omitted error covariances of $.4, .5, .6, .7, \text{ and } .8$, respectively. Factor correlation is held at $.1$, loadings are held at $.4$. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
model sizes. Therefore, models with 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, and 24 indicators total have a first factor with six indicators and a second factor with 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, and 18 indicators, respectively. The colored curves represent five different sizes of omitted error covariance, with colors red, orange, green, blue, and violet corresponding to omitted error covariances of .4, .5, .6, .7, and .8, respectively. The omitted error covariance occurs between indicators loading onto the first factor. Factor correlation is held at .1 and loadings are held at .4.

Figure 17 makes it clear that the addition of more indicators to an imbalanced model improves fit when the size of the factor containing the omitted error covariance remains constant. While the size of the omitted error covariance has a stronger effect on CFI values than on the values of GFI, SRMR, and RMSEA, the overall trend of improvement in fit is the same for all indices. This is the same pattern as that found for balanced two-factor models (Figure 14). Though nonlinear behavior was seen as the size of the model increased, the trends for most of the curves in Figure 14 showed fit improving as the number of indicators increased.

The results from Figure 17 suggest that larger imbalanced models will have better fit than smaller models with the same degree of misspecification, even when the size of the factor containing the misspecification remains the same. This finding is reasonable. As more indicators are added to a model but the size of the factor containing the omitted error covariance remains the same, the size of the factor without the misspecification grows larger than that containing the misspecification and the severity of the misspecification is diluted.

However, it is not clear from Figure 17 whether this diluting effect is due solely to the increasing size of the model on its own or if the relative size of the two factors, one
containing the misspecification and one not, has an effect over and above that of model size. In Figure 16, it was revealed that holding model size constant while increasing the size of the factor containing the misspecification led to an improvement in fit.

To gain better understanding of this relationship between the size of the factors in an imbalanced model and fit index behavior, Figure 18 examines the relationship between index value and the size of the factor containing an omitted error covariance, as was done in Figure 16. In Figure 18, however, the size of the factor containing the misspecification is plotted along the x-axis, allowing for a better comparison between the behavior of indices with respect to model size alone (Figure 17) and the behavior of indices with respect to the size of the factor containing an omitted error covariance.

Figure 18 plots index value (y-axis) against the size of the factor containing an omitted error covariance (x-axis) for a two-factor model with 50 indicators total. The size of the factor containing the misspecification increases from 3 to 47. Thus, the size of the factor not containing the misspecification decreases from 47 to 3. A factor with three indicators was chosen as the most extreme case of imbalance due to the fact that research has shown that factors need at least three indicators loading onto them to be identified. The colored curves represent five sizes of omitted error covariance, with red, orange, green, blue, and violet corresponding to omitted error covariances of .4, .5, .6, .7, and .8, respectively. Factor correlation is held at .1 and loadings are held at .4.

Both CFI and RMSEA show a steady decrease in fit as the size of the factor containing the omitted error covariance increases from three indicators to 19. Once the factor reaches this size, there is little change in index value overall as the size of the factor increases to having 47 of the 50 indicators loading onto it. CFI begins to show improvement in fit as
Figure 18. Plots of fit index values vs. an imbalanced two-factor model with an increasingly large factor involving variables with an omitted error covariance. Number of indicators is 50. Colors red, orange, green, blue, and violet correspond to omitted error covariances of .4, .5, .6, .7, and .8, respectively. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
Figure 18 (continued). Plots of fit index values vs. an imbalanced two-factor model with an increasingly large factor involving variables with an omitted error covariance. Number of indicators is 24. Colors red, orange, green, blue, and violet correspond to omitted error covariances of .4, .5, .6, .7, and .8, respectively. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
the factor containing the misspecification becomes much larger than the factor without the misspecification, but this improvement is very slight.

GFI and SRMR exhibit the same initial decrease in fit as the size of the factor containing the omitted error covariance increases. Depending on the size of the omitted error covariance, index values show worst fit once the factor containing the misspecification reaches a certain number of indicators, show an improvement in fit once one more indicator loads onto that factor, and then level off, suggesting that after a certain number of indicators, the size of the factor containing the misspecification does not influence index value. For example, when the size of the omitted error covariance is .7, SRMR increases to .034 when the size of the factor containing the misspecification is 17, decreases to .018 when the size of the factor containing the misspecification is 18, and remains at about this size as more and more indicators load onto the factor containing the misspecification.

Comparing Figures 16, 17, and 18 allows us to better understand the relationship between the relative size of the factor containing the misspecification and the factor without the misspecification and index behavior. If the size of the factor with the omitted error covariance stays the same as the model size increases overall, all indices show an improvement in fit (Figure 17). Holding $p$ constant (Figures 16 and 18) reveals that the ratio of the differently-sized factors to each other affects index value up until a certain ratio is reached, after which index values are no longer affected by differences in the size of the two factors. Therefore, change in index value in imbalanced models is not due solely to an increase in the size of the factor without the misspecification but rather the change in the ratio of the two differently-sized factors.
3.5 Summary and Discussion

The figures in this section have explored the relationship between the population CFI, NNFI, GFI, AGFI, RMSEA, gamma, and SRMR indices and model misspecification in the context of two-factor CFA models. The effects of six modeling components (loading size, the size of an omitted error covariance, the location of the omitted error covariance, factor correlation size, model size, and model balance) were examined to determine their effect on index values.

As was the case for the one-factor models, the behaviors of CFI and NNFI were highly similar, as were the behaviors of GFI and AGFI, and RMSEA and gamma. The four indices discussed in this section, then, are CFI, GFI, RMSEA, and SRMR. The conclusions reached for CFI, GFI, and RMSEA can be applied to NNFI, AGFI, and gamma, respectively.

The addition of a second factor did not alter the patterns of any index’ behavior when compared to behavior in a one-factor model (Figure 1). GFI, RMSEA, and SRMR show worse fit as the size of the omitted error covariance increases and exhibit increased sensitivity to misspecification when loadings are large compared to when they are small (Figure 10). The non-monotone relationship between CFI and the size of the omitted error covariance is still present.

In addition to the nuisance variables explored in the one-factor model scenarios (Chapter 2), the additional influence of factor correlation was explored in this Chapter. The size of the factor correlation was shown to have a minimal impact on index value when misspecification was due to an omitted error covariance, particularly for CFI (Figure 11) and particularly when compared to the effects of loading size on index value. Though the effect
was small, all indices appear more sensitive to misspecification when the size of the factor correlation is larger.

As in the one-factor case, the size of the loadings exert a strong influence on index behavior regardless of other model features such as the value of the factor correlation or the location of the omitted error covariance (Figure 10). Sensitivity to misspecification is greater for all indices, however, when the omitted error covariance involves variables loading onto different factors than when all variables involved in the misspecification load onto the same factor (Figure 10). This is an interesting finding and suggests that these indices are most sensitive to misspecification at a “global” level, when misspecification involves variables from multiple factors. When an omitted error covariance occurs between indicators of the same factor, the effect the misspecification has on the other factor loadings is largely limited to the factor loadings of the factor containing the misspecification. However, when the omitted error covariance occurs between indicators loading onto different factors, the effect of the misspecification is likely spread out to affect the loadings of both factors and is thus more detectable by the fit indices.

This increased sensitivity seen when the omitted error covariance occurs between indicators of different factors can be seen as a good characteristic of fit indices, as a misspecification that involves more than one factor is most likely more problematic than a misspecification unique to a single factor. However, it is also a behavior of which researchers should be aware. A model with an omitted error covariance as large as .7, for example, will be deemed well-fitting by the GFI using the common cutoff value of .95 when the error covariance occurs between indicators of the same factor. If the same size error covariance occurs between indicators of different factors, however, the model will be rejected by the
GFI. Thus, lower values of fit indices may be indicative of the location of the misspecification as well as its size.

As in the one-factor case, the effects of an additional omitted error covariance on index values were explored for a two-factor model (Figures 11, 12, and 13). Because of the additional complexity brought on by the additional factor, we were able to examine both the effects of the omitted error covariances sharing or not sharing an indicator (referred to above as the effects of misspecification isolation) and the effects of the omitted error covariances occurring between indicators from the same or different factors (referred to above as the effects of misspecification location).

Figures 11, 12, and 13 involved the examination of three different degrees of misspecification location: a model in which both omitted error covariances occurred between indicators loading onto the same factor (Figure 11), a model in which one of the omitted error covariances occurred between indicators of the same variable but the other omitted error covariance involved indicators from both variables (Figure 12), and a model in which both omitted error covariances involved indicators from both variables (Figure 13). At each degree of misspecification location (that is, for each of the three figures), the effects of misspecification isolation were examined as well.

Consistent with our expectations, when misspecification was least isolated (when the omitted error covariances did not share an indicator) and as more of the misspecification involved indicators from both factors (that is, as location shifted from one factor to both factors, from Figure 11 to Figure 12 to Figure 13), all indices showed a decrease in fit. One curious pattern observed, however, pertained to the effects of misspecification isolation. When both omitted error covariances involved indicators all loading onto one factor (Figure
11) or when both involve indicators from different factors (Figure 13), fit was worse when
the omitted error covariances do not share a variable. Only when one omitted error
covariance remained isolated within a factor and the second involved indicators from both
factors (Figure 12) did all indices show better fit when the two omitted error covariances did
not share a variable.

To gain further insight into the differences among Figures 11, 12, and 13, an
examination of the residual matrices ($\Sigma^* - \Sigma(\hat{\theta})$) at select values were examined for all three
Figures. Both omitted error covariances were fixed to .2 and loadings were fixed to .6. The
residual matrices for Figures 11, 12, and 13 were obtained and compared.

It was found that regardless of the location of the misspecification (whether the
omitted error covariances involved indicators from one or both factors), residuals were found
to be more spread out across both factors in cases where the omitted error covariances did not
share an indicator (when misspecification was least isolated, represented by the dashed lines
in Figures 11, 12, and 13).

At the same time, omitted error covariances not sharing a variable led to lower
average residuals for Figures 11 and 13 but led to higher average residuals for Figure 12.
These findings suggest that indices are sensitive to the isolation of the misspecification—the
more indicators involved in the misspecification, the worse the model will be said to fit
(according to all indices) when compared to cases where omitted error covariances might
share a variable (the solid lines in Figures 11, 12, and 13).

The effects of model size were also examined, and as in the one-factor case, the
number of indicator variables included in the model was found to have an effect on index
value (Figure 14). Results were consistent with the case involving a one-factor model (Figure
for both CFI and GFI, though the non-monotone relationship between index value and the size of the model was reduced. The additional factor complexity completely eliminated the non-monotone relationship for the largest sizes of omitted error covariance examined and for smaller loadings for GFI. The reduction of the non-monotone trend suggests that for CFI and GFI, the complexity of the factor structure is influential on index value. The increased factor complexity also influenced RMSEA values when compared to the one-factor model case.

The addition of a second factor in Figure 14 caused RMSEA to exhibit an initial decrease in fit as the size of the model increases and then an increase in fit once a certain model size is reached (this inflection point is dependent on loading size and the size of the omitted error covariance), a pattern shown by Savalei (2010). In the one-factor case (Figure 8), RMSEA showed an improvement in fit as the number of indicators increased. The findings for the two-factor case suggests that there is an optimal model size at which RMSEA will indicate best fit, depending on the size of the misspecification and loading size.

Finally, the last model component explored was model balance. There has been no prior research investigating whether an unequal number of indicators per factor has an effect on fit index value. Our first exploration found that all indices showed an improvement in fit as the true model became more imbalanced (Figure 15). Further exploration revealed that best fit was shown for the most severely imbalanced model when the omitted error covariance occurred between indicators loading onto the larger factor. Worst fit, on the other hand, was shown for the most severely imbalanced model when the omitted error covariance occurred between indicators loading onto the smaller factor (Figure 16).

The behavior found in Figure 16 suggests that it may be the size of the factor containing the misspecification, rather than the equality of the number of indicators per
factor, that plays a role in index value. As a model grows more and more imbalanced, one factor becomes larger than the other. If an omitted error covariance occurs between variables loading onto the larger factor, the size of the misspecification will be more “diluted” than if it were to occur between variables of a smaller factor. Conversely, if the omitted error covariance occurs between variables loading onto the smaller factor, the size of the misspecification will be “amplified” within the smaller factor, which will be reflected by the fit index as a more severe misspecification. Further exploration revealed that relative size of the factors (as measured by the number of indicator variables) affects index value up until a certain ratio is reached (Figure 18). The relative size of the factors to each other influenced index value over and above the increase in the size of the factor without the misspecification (Figure 17). This suggests that in the case of imbalanced models, changes in fit due to the number of indicator variables in the model cannot be explained just by the total number of indicators present in the entire model (that is, just the size of $p$) but involves a more complex relationship between the sizes of the two factors to each other.
Chapter 4: Results for Misspecified Latent Structure Models

The final scenarios examined are those involving misspecification of the latent structure. Misspecifications of this form can be said to be more serious than those involving an omitted error covariance since they involve a discrepancy between the number of factors in the fitted model and the number of factors underlying the data.

To gain a clear understanding of index behavior in the case of a misspecified latent structure, the figures in this chapter are broken into two general sets. Figures 19 through 22 examine the case where a one-factor model is fit to true models with either two or three factors. Within these figures, the effects of factor correlation size, inter-item correlation size, model size (in terms of the number of indicator variables), and the balance of the true model are examined. These figures offer an in-depth exploration of the effects of these nuisance variables in the presence of latent misspecification.

The second set of figures, Figures 23 through 26, examines the effects of an increasingly severe latent misspecification as measured by the discrepancy between the number of factors in the fitted model and the number in the true model. In these figures, a one-factor model is fitted to true models that have an underlying structure of two, three, four, five, six, seven, or eight factors. Within this set of figures, the effects of factor correlation, inter-item correlation, the number of indicators, and the ratio of indicators to factors \((p:k)\), are examined. Figures 23 through 26 also provide an opportunity to examine the effects of model size on index value, both with respect to the number of indicators total and the ratio of indicators to factors.

Unlike in Chapters 2 and 3, differences in behavior were found between CFI and NNFI, GFI and AGFI, and RMSEA and gamma for several of the figures presented in this
chapter. In these cases, each index will be discussed individually and a plot for each index will be provided. Otherwise, only the results of CFI, GFI, RMSEA, and SRMR will be discussed.

4.1 One-Factor Model Fit to Two- and Three-Factor Data

Figure 19 presents the case in which a one-factor model with 12 indicators is fit to true models with either two underlying factors (solid lines) or three underlying factors (dashed lines). Index value (on the y-axis) is plotted against the size of the factor correlation. For models with more than two factors, all factor correlations are set to be equal. The six colored curves correspond to six loading sizes with red, orange, green, blue, violet, and black corresponding to loadings of .4, .5, .6, .7, .8, and .9, respectively. For all indices, the curves corresponding to loadings of .9 (black curves) have been smoothed to improve readability without altering their general shape\(^3\).

Factor correlation was chosen to vary continuously in Figure 19 since it is the most likely nuisance variable to exert additional influence on index value above the influence of the misspecification itself. As factor correlation increases to one, both the two- and three-factor models can be better represented by the one-factor fitted model. Thus it is expected that all indices will show better fit as the factor correlation increases. In addition to this expectation, it is hoped that indices will be able to reflect the difference in the degree of misspecification between the case where a one-factor model is fit to a true model with two factors and the case where a one-factor model is fit to a true model with three factors. Ideally, indices will show worse fit when the true model has three factors compared to when the true model has two factors.

\(^3\) Curve smoothing was done in R using the package smooth.Pspline, which creates a natural polynomial smooth of the input data.
Figure 19. Plots of fit index values vs. an increasing factor correlation (0 – 1) when a 1-factor model is fit to 2-factor data (solid lines) and a 1-factor model is fit to 3-factor data (dashed lines). Number of indicators is 12. Colors red, orange, green, blue, violet, and black correspond to loadings of .4, .5, .6, .7, .8, and .9, respectively. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
Figure 19 (continued). Plots of fit index values vs. an increasing factor correlation (0 – 1) when a 1-factor model is fit to 2-factor data (solid lines) and a 1-factor model is fit to 3-factor data (dashed lines). Number of indicators is 12. Colors red, orange, green, blue, violet, and black correspond to loadings of .4, .5, .6, .7, .8, and .9, respectively. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
Figure 19 reveals that as the factor correlation increases to one, all indices show an improvement in fit when a one-factor model is fit to both the two- and three-factor models and show perfect fit for both cases when the factor correlation is one (as expected). Figure 19 also shows CFI for the first time exhibiting monotone behavior, indicating better fit as the size of the factor correlation increases. This is in contrast to the index’s behavior in cases where misspecification was due to omitted error covariances, where non-monotone behavior was apparent in almost all figures.

It appears, too, that CFI is the only index for which the predictions regarding the degree of misspecification hold. CFI shows worse fit in the case when a one-factor model is fit to the three-factor model (dashed lines) than when a one-factor model is fit to the two-factor model (solid lines). This behavior suggests that CFI appropriately reflects the difference in the degree of misspecification between the two cases.

Unlike CFI, however, Figure 19 shows RMSEA to be unable to distinguish between the case where a one-factor model is fit to the two-factor model and the case where a one-factor model is fit to the three-factor model. Regardless of loading size, RMSEA shows extremely similar values for the two-factor and three-factor models (solid and dashed lines, respectively). This result replicates findings by Savalei (2010). The inability of RMSEA to distinguish a difference in fit when a one-factor model is fit to two models with different factor complexities is problematic for interpretation. It may suggest that RMSEA is not the best index to use when misspecification is likely due to an incorrectly modeled latent structure.

The behaviors of GFI and SRMR are even more problematic. Figure 19 shows that like CFI, these indices reflect the different degree of fit between a one-factor model fit to the
two-factor model and a one-factor model fit to the three-factor model. Unlike CFI, however, GFI and SRMR show better fit when a one-factor model is fit to the three-factor model (dashed lines) than when a one-factor model is fit to the two-factor model (solid lines). Such behavior is even more problematic for interpretation than RMSEA’s inability to distinguish between the two cases. Further investigation into the performance of these indices as the latent structure of the true model grows increasingly more complex is carried out in Figure 23.

CFI also appears to be more sensitive to latent misspecification than GFI, RMSEA, or SRMR. For example, Figure 19 shows that when loadings are .4, SRMR values never increase above the commonly accepted cutoff value of .08, meaning a one-factor model would always be accepted as fitting two-and three-factor models at loadings of .4. At the same loading size, a factor correlation of .7 is necessary before the one-factor model is accepted as fitting either two-factor model or three-factor model according to CFI. The fact that larger factor correlations are necessary for CFI to indicate a well-fitting model than for GFI, RMSEA, or SRMR suggests, particularly when loadings are small, that CFI may be better able to detect misspecifications involving latent structure than these other indices.

In addition to revealing whether the degree of misspecification (as measured by the discrepancy between the number of factors present in the true model and the number present in the fitted model) is reflected by fit indices, Figure 19 also shows that index performance is influenced by loading size. While all indices show better fit as the factor correlation increases, the size of the factor correlation necessary for indices to show the one-factor model as fitting the data well (using conventional cutoff criteria) is different depending on loading size.
The effect of loading size is particularly apparent for GFI, RMSEA, and SRMR. For example, when a one-factor model is fit to a true model with a two-factor structure, the factor correlation has to be smaller than .21 for GFI to indicate poor fit using the commonly accepted cutoff of .95 when loadings are .4 (solid red curve). However, when loadings are .9, GFI indicates poor fit even when the factor correlation is as high as .96 (solid black curve). The fact that model rejection using the conventional cutoff value occurs at such different sizes of factor correlation may suggest that a universal cutoff value may not be appropriate for these indices in this type of misspecification case.

4.1.1 Model size

Model size, as defined by the total number of indicator variables $p$ in the model, was shown to have an effect on index value with respect to detecting an omitted error covariance. Previous figures (Figures 8 and 14) revealed a nonlinear relationship between index size and the number of indicators included in the model. Despite this nonlinear trend, an eventual decrease in fit was seen as $p$ continued to increase.

In the context of misspecification due to omitted error covariances, this eventual decrease in fit suggested that as more indicators were added to a model, the effect of the misspecification was “diluted.” That is, the effect the omitted error covariance had on the residual matrix $\Sigma^* - \Sigma(\hat{\theta})$ appeared less severe as the overall matrix grew with the addition of more variables.

For similar reasons, we expect in the case of latent misspecification to find the opposite pattern as that found in cases involving omitted error covariances. That is, as more variables are included in the model, indices will show worse fit. A decrease in fit is expected because in the case of latent misspecification, the influence of the misspecification is not
isolated to two variables as it is in the omitted error covariance case but rather involves the entire structure of the model, particularly the relationships between the indicators and the factors. The covariance matrix of a one-factor model has a different structure than that of a two-factor model, for example. If a one-factor model were fit to a true model with two factors, there would be discrepancies between the structures of the two covariance matrices that, if the discrepancies were large enough, could lead to the rejection of the fitted model as an accurate representation of the structure of the true data. As more indicators are added to the model, it would seem reasonable to expect that any differences in structure between a fitted model and a true model would be “amplified,” as there would be more elements in the covariance matrices that could differ between the one-factor covariance structure and the two-factor covariance structure. Therefore, it is expected that as \( p \) increases, indices will show worse fit in cases involving latent misspecification.

To examine the effects of the number of indicators on index performance the context of latent structure misspecifications, Figure 20 presents index value (on the y-axis) against an increasing number of indicators \( (p = 6, 12, 18, 24, 30, 36) \) when a one-factor model is fit to a true model with two factors (solid lines) and when a one-factor model is fit to a true model with three factors (dashed lines). The number of indicators in a model is a categorical variable, but neighboring points are connected in Figure 20 to improve readability. The true models have an equal number of indicators per factor, such that the two-factor model has 3, 6, 9, 12, 15, and 18 indicators per factor for the different levels of \( p \) and the three-factor model has 2, 4, 6, 8, 10, and 12 indicators per factor for each level of \( p \). The six colored curves correspond to different loading sizes, with red, orange, green, blue, violet, and black
Figure 20. Plots of fit index values vs. an increasing number of indicators when a 1-factor model is fit to 2-factor data (solid lines) and a 1-factor model is fit to 3-factor data (dashed lines). Colors red, orange, green, blue, violet, and black correspond to loadings of .4, .5, .6, .7, .8, and .9, respectively. Neighboring points are connected for readability. Factor correlation is .1.
Figure 20 (continued). Plots of fit index values vs. an increasing number of indicators when a 1-factor model is fit to 2-factor data (solid lines) and a 1-factor model is fit to 3-factor data (dashed lines). Colors red, orange, green, blue, violet, and black correspond to loadings of .4, .5, .6, .7, .8, and .9, respectively. Neighboring points are connected for readability. Factor correlation is .1.
Figure 20 (continued). Plots of fit index values vs. an increasing number of indicators when a 1-factor model is fit to 2-factor data (solid lines) and a 1-factor model is fit to 3-factor data (dashed lines). Colors red, orange, green, blue, violet, and black correspond to loadings of .4, .5, .6, .7, .8, and .9, respectively. Neighboring points are connected for readability. Factor correlation is .1.
Figure 20 (continued). Plots of fit index values vs. an increasing number of indicators when a 1-factor model is fit to 2-factor data (solid lines) and a 1-factor model is fit to 3-factor data (dashed lines). Colors red, orange, green, blue, violet, and black correspond to loadings of .4, .5, .6, .7, .8, and .9, respectively. Neighboring points are connected for readability. Factor correlation is .1.
corresponding to loadings of .4, .5, .6, .7, .8, and .9, respectively. All factor correlations were set to .1.

An examination of Figure 20 reveals that some indices are more affected by model size than others. NNFI and RMSEA appear to be most affected and indicate poor fit when \( p = 6 \), show a rapid improvement in fit when \( p \) is increased to 12, and continue to show an improvement in fit as the number of indicators increases. This is the pattern that was expected. However, regardless of loading size, NNFI values never increase above .5 and RMSEA values never fall below the commonly used cutoff value of .06 (except when loadings are .4 and \( p = 24 \)). This suggests that while an increase in the size of the model improves fit for NNFI and RMSEA, both indices are so highly sensitive to latent misspecification that any change in model size will most likely not affect a researcher`s decision to accept or reject a model using conventional cutoff values. This same sensitivity to latent misspecification is found here for CFI and SRMR as well. Changes in model size have little effect on the values of these two indices, as is evident by the fact that the colored lines stay relatively flat as model size increases.

GFI and gamma show a decrease in fit in Figure 20 as the number of indicators included in the model increases, though this decrease is slight and suggests that in the context of latent misspecification, model size has little to no effect on these indices. GFI and SRMR indicate poor fit for all values of \( p \) and all loading sizes, except when loadings are .4 and \( p = 6 \). The behavior of AGFI is inconsistent across loading sizes. When loadings are small (.4, .5), AGFI shows a decrease in fit as the size of the model increases. When loadings are large (.8, .9), AGFI shows an increase in fit as the size of the model increases.
With respect to indicating a difference in fit between the two-factor model (solid lines) and the three-factor model (dashed lines), Figure 20 shows that RMSEA remains unable to detect the difference in the degree of misspecification between the two cases regardless of model size. CFI and NNFI indicate worse fit when a one-factor model is fit to a three-factor true model than a two-factor true model, though again, values for both scenarios are very low (less than .5). GFI, gamma, and SRMR show better fit when a one-factor model is fit to a three-factor true model than when a one-factor model is fit to a two-factor true model. These patterns were also seen in Figure 19.

As a final comment on this figure, we note that this is the first of the manipulations in this thesis for which NNFI, AGFI, and gamma behave differently than CFI, GFI, and RMSEA, respectively. While an explanation for this behavior could be that the differences are due to the inclusion of the degrees of freedom in the equation for NNFI and the inclusion of both the degrees of freedom and the number of indicators \( p \) in the equation for AGFI and gamma, the fact that these three indices did not perform differently than CFI, GFI, and RMSEA in previous figures involving the manipulation of model size suggests that there may be something in particular that is different between misspecifications involving omitted error covariances and misspecifications involving the latent structure of a model that causes NNFI, AGFI, and gamma to behave differently in one type of misspecification but not the other.

Figure 20 is limited in the sense that factor correlation is set to .1, and therefore the influence of factor correlation is not as model size changes. Figure 21 explores whether the patterns observed in Figure 20 remain the same at different factor correlation sizes. Figure 21 plots index value (y-axis) against an increasing number of indicators (\( p = 6, 12, 18, 24, \))
Figure 21. Plots of fit index values vs. an increasing number of indicators when a 1-factor model is fit to 2-factor data (solid lines) and a 1-factor model is fit to 3-factor data (dashed lines). Colors red, orange, green, blue, violet, and black correspond to factor correlations of .1, .2, .3, .4, .5, and .6, respectively. Neighboring points are connected for readability. Loadings are .7. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
Figure 21 (continued). Plots of fit index values vs. an increasing number of indicators when a 1-factor model is fit to 2-factor data (solid lines) and a 1-factor model is fit to 3-factor data (dashed lines). Colors red, orange, green, blue, violet, and black correspond to factor correlations of .1, .2, .3, .4, .5, and .6, respectively. Neighboring points are connected for readability. Loadings are .7. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
when a one-factor model is fit to a true model with two factors (solid lines) and when a one-factor model is fit to a true model with three factors (dashed lines). The number of indicators is a categorical variable, but neighboring points in Figure 21 are connected to improve readability. The six colored curves now correspond to different factor correlation sizes, with red, orange, green, blue, violet, and black corresponding to factor correlations of .1, .2, .3, .4, .5, and .6, respectively. All loadings are set to .7.

Figure 21 reveals that regardless of model size, all indices show better fit for both the 2-factor and 3-factor models for larger sizes of factor correlation. As seen in Figure 19, this result is expected, as the more correlated the factors in the true models become, the more accurate a single-factor model representation is. Figure 21 also reveals further evidence supporting CFI as an index that is very sensitive to misspecifications involving the latent structure. Even at the largest model examined here ($p = 30$) and for the largest value of factor correlation (.6), CFI is .77, well below the commonly used cutoff value of .95. GFI and RMSEA also fail to indicate good fit using commonly-applied cutoff values, but do show an improvement in fit as model size increases.

4.1.2 Model balance

Next we examine the effect of model balance on index sensitivity. It has already been seen that model balance has an effect on index value when misspecification is due to an omitted error covariance. Figures 15 and 16 revealed that the size of the factor containing the misspecification was a key component in the fit of imbalanced models compared to the fit of a balanced model, with fit improving as the size of the factor containing the omitted error covariance increased.
In the case of latent misspecification, the source of the misspecification can be seen as a more “global” form of misspecification that affects the model as a whole. As such, it can be expected that the size of the factors will have an influence on index performance.

Consider two true models: one is a two-factor model that has eight indicators per factor. The other is a two-factor model that has three indicators loading onto one factor and 13 indicators loading on to the other factor. If a one-factor model were fit to each of these true models, we would expect better fit for the imbalanced true model than the balanced model. This is because the structure of the covariance matrix of the imbalanced model resembles the structure of a one-factor covariance matrix more than the covariance matrix of the balanced model resembles the structure of a one-factor covariance matrix. For the imbalanced true model, only three of the 16 indicators fail to load onto the single factor proposed to exist by the fitted model. For the balanced true model, eight of the 16 indicators fail to load onto the single proposed factor. Using this reasoning, we expect to see best fit in the case where the true model is most severely imbalanced and worst fit when the true model has an equal number of indicators per factor.

The effects of model balance on index behavior are explored in Figure 22. A one-factor model is fit to a balanced two-factor true model and five increasingly imbalanced true models, all with 16 indicators total. Index value (y-axis) is plotted against an increasing factor correlation (x-axis). The colored curves represent six levels of balance in the structure of the true model (not the structure of the fitted model, as the model being fit is always a one-factor model in this Figure). Red corresponds to a balanced true model (eight indicators per factor), orange corresponds to a true model with nine and seven indicators per factor, green corresponds to a true model with ten and six indicators per factor, blue corresponds to a true
Figure 22. Plots of fit index values vs. an increasing factor correlation (0 – 1) when a 1-factor model with 16 indictors is fit to 2-factor data. Lines red, orange, green, blue, violet, and black correspond to balanced data (8 indicators per factor) and five increasingly imbalanced data (7 and 9 indicators per factor, 6 and 10 indicators per factor, 5 and 11 indicators per factor, 4 and 12 indicators per factor, and 3 and 13 indicators per factor), respectively. Solid lines correspond to loadings of .4, dashed lines correspond to loadings of .7. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
Figure 22 (continued). Plots of fit index values vs. an increasing factor correlation (0 – 1) when a 1-factor model with 16 indictors is fit to 2-factor data. Lines red, orange, green, blue, violet, and black correspond to balanced data (8 indicators per factor) and five increasingly imbalanced data (7 and 9 indicators per factor, 6 and 10 indicators per factor, 5 and 11 indicators per factor, 4 and 12 indicators per factor, and 3 and 13 indicators per factor), respectively. Solid lines correspond to loadings of .4, dashed lines correspond to loadings of .7. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
model with 11 and five indicators per factor, violet corresponds to a true model with 12 and four indicators per factor, and black corresponds to a true model with 13 and three indicators per factor. Loadings are arbitrarily fixed to .4 (solid lines) or .7 (dashed lines).

Figure 22 makes it clear that all indices show worse fit for the one-factor model fit to the balanced true model (red curves) and best fit for the one-factor model fit to the most severely imbalanced true model (black curves). For example, when loadings are .7 and the factor correlation is .4, CFI is .65 when a one-factor model is fit to the balanced true model (red dashed curve). At the same loading size and factor correlation size, CFI is .92 in the most imbalanced true model case, where 13 indicators load onto one factor and three indicators load onto the other (black dashed curve). These results support prior expectations and suggest that in the context of models involving latent misspecification, the more imbalanced the true model is, the less sensitivity all indices have in detecting misspecification.

4.2 Increasing Number of Latent Factors

Attention is now turned to scenarios involving true models with increasingly complex latent structures. Figures 19 through 22 have provided insight regarding index performance when a one-factor model is fit to true models with either two or three underlying factors. It is of interest to determine whether the patterns found in these previous figures generalize to cases in which the true model has a higher factor complexity.

The next two figures extend the results from Figure 19, in which a one-factor model was fit to true models with either two or three factors, to cases where the true models have even higher factor complexities. In Figure 23, a one-factor model is fit to true models that have an underlying factor structure of two, three, four, five, six, seven, or eight factors. In
Figure 24, a one-factor model is again fit to true models with underlying factor structures of two, three, four, five, six, seven, or eight factors, but in this case, the average inter-item correlation is held constant across the different factor complexities. Recall that in Figure 19, RMSEA appeared unable to detect the difference in the amount of misspecification between a one-factor model fit to two-factor data and a one-factor model fit to three-factor data. As noted by Steiger (2000) and later by Savalei (2010), a possible explanation for RMSEA’s behavior is that an increase in the number of latent factors present in the true model lowers the average item inter-correlation. An example of this can best be seen with uncorrelated factors. Suppose a two-factor model in which the factors were uncorrelated had 12 indicators, each with loadings of .4. If this two-factor model were true in the data, there would be 30 inter-item correlations of .16 (.4²) and 36 inter-item correlations of zero. If a three-factor model were instead true, there would be only 18 inter-item correlations of size .16 and 48 inter-item correlations of zero. Thus, the average inter-item correlation is lower the more factors are included in the model.

Steiger (2000) has shown that fit indices such as the RMSEA are sensitive to the average inter-item correlation, and are higher when this correlation is high. Thus, holding the average inter-item correlation constant across the different factor complexities will allow insight as to whether the behavior of RMSEA as seen in Figure 19 is due to its sensitivity to the inter-item correlation. The procedure detailing how the inter-item correlations are held constant in Figure 24 will be described in detail below.

We first examine the effects of true models with increasingly complex latent structures on index behavior. Figure 23 presents the case where a one-factor model with 24 indicators is fit to true models with 2, 3, 4, 5, 6, 7, and 8 latent variables. Index value (y-axis)
Figure 23. Plots of fit index values vs. an increasing number of latent factors in the underlying model (2 – 8). Done for six values of factor loadings. Colors red, orange, green, blue, violet, and black correspond to loadings of .4, .5, .6, .7, .8, and .9, respectively. Solid lines correspond to a factor correlation of .4, dashed lines correspond to a factor correlation of .1. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
Figure 23 (continued). Plots of fit index values vs. an increasing number of latent factors in the underlying model (2 – 8). Done for six values of factor loadings. Colors red, orange, green, blue, violet, and black correspond to loadings of .4, .5, .6, .7, .8, and .9, respectively. Solid lines correspond to a factor correlation of .4, dashed lines correspond to a factor correlation of .1. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
is plotted against the number of latent factors in the true model (2 – 8, x-axis). The number of factors is a categorical variable, but neighboring points are connected in Figure 23 for readability. The six colored curves correspond to six loading sizes, with red, orange, green, blue, violet, and black corresponding to loadings of .4, .5, .6, .7, .8, and .9, respectively. All factor correlations are fixed either to .4 (solid lines) or .1 (dashed lines).

CFI was the only index in Figure 19 to behave as was hoped, showing worse fit when a one-factor model was fit to a three-factor model than when a one-factor model was fit to a two-factor model. Figure 23, however, reveals some troubling behavior as the number of latent factors increases. When factor correlation is .1, CFI performs as in Figure 19 and shows a decrease in fit as a one-factor model is fit to true models with an increasing number of underlying latent factors. However, when factor correlation is .4, CFI shows an initial decrease in fit as the number of latent factors increases, but then begins to show a slight improvement in fit to the point where, at smaller loadings, a one-factor model fit to eight-factor data fits just as well as a one-factor model fit to two-factor data. This suggests that CFI is the only index of those studied for which the size of the factor correlation has an effect on whether index behavior can be considered appropriate. While this overall pattern is not the one we expected based on the results of Figure 19, we note that regardless of the pattern, CFI values never increase above .8. This lends further support to this index being extremely sensitive to misspecification at the latent level.

In Figure 19, it was revealed that while GFI and SRMR were sensitive to the difference in degree of fit between the two-factor and three-factor models when a one-factor model was fit, they showed better fit when a one-factor model is fit to a true model with three factors than when a one-factor model is fit to a true model with two factors. The results from
Figure 23 show that this behavior generalizes to more complex factor structures. Regardless of loading size and factor correlation size, GFI and SRMR show an increase in fit when a one-factor model is fit to true models with an increasing number of latent factors. For larger loading sizes (.7, .8, .9), this improvement in fit does not affect model acceptance/rejection at the commonly used cutoff values. Even though there is an improvement in fit from when the number of latent factors is three to when the number of latent factors is eight, the indices still show poor enough fit that models will not be accepted using these cutoff criteria.

However, for smaller loading sizes (.4, .5), relying on these indices could lead to problems regarding appropriate model acceptance/rejection. For example, when loadings are .5 and factor correlation is .4, a one-factor model fit to a true model with three factors would be rejected using the commonly accepted cutoff value of .08 for SRMR. However, at the same loading and factor correlation sizes, a one-factor model fit to eight-factor data would be accepted.

Figure 19 showed RMSEA to be almost completely insensitive to differences in fit between the case where a one-factor model was fit to a true model with two underlying factors and the case where a one-factor model was fit to a true model with three underlying factors. Figure 23 shows that this behavior generalizes to true models with higher factor complexity. Of all the indices, RMSEA appears least sensitive to the increase in misspecification as a one-factor model is fit to true models with an increasing number of latent factors. Particularly when loadings are small, RMSEA values remain almost constant as the number of latent factors in the true model increases. At larger loadings, it actually shows an improvement in fit as the misspecification grows larger, though like GFI and
SRMR, RMSEA values are large enough such that the improvement in fit does not affect the model acceptance/rejection decision at the commonly used cutoff value.

In Figure 23, there were substantial differences in the underlying factor structures for different true models. Therefore there were substantial differences in the inter-item correlations for the different true models as well. For example, when $p = 24$, loadings are .9, and all factor correlations are equal to .4, the average inter-item correlation for the true model with two underlying factors is .556, while the average inter-item correlation for the true model with eight underlying factors is .366.

Given these substantial differences in inter-item correlations and given Steiger’s (2000) findings regarding RMSEA’s sensitivity to the size of the inter-item correlations, we sought to determine whether index behavior in Figure 23 could be explained by the decrease in the average inter-item correlation as more latent factors were present in the true model.

To determine if the patterns in Figure 23 were due at least in part to the differences in the average inter-item correlation, an additional figure was constructed in which the average inter-item correlation was held constant as the number of latent factors in the true model increased. Figure 24 presents the case where a one-factor model with 24 indicators is fit to true models with an increasing number of latent factors when the average inter-item correlation is held constant across all degrees of factor complexity ($k = 2, 3, 4, 5, 6, 7, \text{ or } 8$). To accomplish this, loadings for the seven different factor complexities were altered such that the average inter-item correlation for each level of complexity was the same. Index value (y-axis) is plotted against the different degrees of factor complexity (x-axis). Factor complexity is a categorical variable, but neighboring points are connected in Figure 24 for readability. The six colored curves correspond to six loading sizes. The sizes of these
Figure 24. Plots of fit index values vs. an increasing number of latent factors in the underlying model (2 – 8) when the average correlation is held constant. The number of indicators is 24. Done for six values of factor loadings, with colors red, orange, green, blue, violet, and black correspond to loadings of .4, .5, .6, .7, .8, and .9, respectively. Solid lines correspond to a factor correlation of .4, dashed lines correspond to a factor correlation of .1. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
Figure 24 (continued). Plots of fit index values vs. an increasing number of latent factors in the underlying model (2 – 8) when the average correlation is held constant. The number of indicators is 24. Done for six values of factor loadings, with colors red, orange, green, blue, violet, and black correspond to loadings of .4, .5, .6, .7, .8, and .9, respectively. Solid lines correspond to a factor correlation of .4, dashed lines correspond to a factor correlation of .1. NNFI, AGFI, and gamma plots omitted due to similarities with CFI, GFI, and RMSEA.
loadings change depending on \( k \), and thus it is not possible to list in Figure 24 each specific value the colored curves correspond to. Instead, red, orange, green, blue, violet, and black are said to correspond to increasingly large loading sizes, with red representing the smallest loading for each level of \( k \) and black representing the largest loading for each level of \( k \). Figure 24 retains the \(.4 - .9\) labeling to preserve consistency with other figures. Factor correlation is fixed to \(.4\) (solid lines) or \(.1\) (dashed lines).

If index behavior is affected by the decrease in the average inter-item correlation as the latent structure of the true model data becomes more complex, we expect to see different patterns arising in Figure 24 from those found in Figure 23, in which the average inter-item correlation was not held constant. Indeed, Figure 24 makes it clear that the change in the average inter-item correlation as \( k \) increases does effect index behavior. These effects are most apparent for RMSEA. When the average inter-item correlation is held constant as the number of latent factors in the true model increases, RMSEA shows worse fit for higher levels of factor complexity, regardless of the size of the factor correlation and while the loadings change to allow for the average inter-item correlation to be held constant across the different degrees of factor correlation.

GFI and SRMR also show a decrease in fit in Figure 24 as the number of latent factors present in the true model increases, but only when the factor correlation is \(.1\). When factor correlation is \(.4\), these indices show an improvement in fit as more latent factors are present, as they did when the average inter-item correlation was not held constant (Figure 23). Thus, while the size of the average inter-item correlation plays a role in the behaviors of these two indices, holding it constant only affects index behavior when the factors are only slightly correlated. As factor correlation increases, GFI and SRMR are less able to detect an
increase in misspecification, regardless of whether the average inter-item correlation is held constant or left to change.

CFI shows a decrease in fit in Figure 24 as the factor complexity of the true model increases, regardless of the size of the factor correlation. When the average inter-item correlation was not held constant (Figure 23), CFI values either remained fairly constant (when factor correlation was .1) or began to indicate an improvement in fit (when factor correlation was .4) as the number of latent factors in the true model increased. Thus, it appears that CFI is influenced by the size of the average inter-item correlation. Again, however, it should be noted that CFI values are never higher than .6 in Figure 24, indicating (as other figures have) that CFI is highly sensitive to misspecification involving the latent structure, regardless of other nuisance variables.

4.2.1 Indicator to factor ratio balance

The number of indicators in Figures 23 and 24 was held constant across all degrees of factor complexity, leading to fewer indicators per factor as the number of factors in the true model increased. The following two figures, Figure 25 and Figure 26, examine the relationship between the number of indicators and the number of factors present in the data and how the ratio of indicators to factors (p:k) affects index ability to detect the increase in misspecification as a one-factor model is fit to true models with an increasing number of latent factors.

Figure 25 will present the case in which a one-factor model is fit to true models with an increasing number of latent factors, when the number of indicators per factor is held constant across levels of k. Doing so allows us to examine the effects of p:k being held constant across different values of k. If p:k affects index behavior above and beyond the
affects of increasing $k$ only, different patterns should be expected in Figure 25 than in Figure 23 (in which $k$ increased but $p$ did not).

Figure 26 will present the case where both $p$ and $k$ increase but $p:k$ does not remain constant across levels of $k$. In Figure 26, for every one factor increase in the true model, one more indicator per factor is included. For example, a true model with four factors has a total of 20 indicators ($p:k = 5:1$), while a true model with five factors has a total of 30 indicators ($p:k = 6$). Thus, Figure 26 represents the case where both $p$ and $k$ change, but do so at a rate such that the ratio $p:k$ is not constant. While it may be considered unrealistic to examine models in which there are 72, 56, or even 42 indicators (values of $p$ when $k = 8, 7,$ and 6, respectively), the purpose of the manipulation in Figure 26 is to simply compare the patterns found in this Figure with those found for Figure 25 (in which both $p$ and $k$ change at a rate such that $p:k$ is constant) and Figure 23 (where $k$ changes but $p$ does not).

If the patterns in Figure 26 are similar to or more exaggerated than those in Figure 25, it could be said that index behavior is due not to the ratio of indicators per factor being held constant but rather to the general increase in model size as both $p$ and $k$ increase. If Figure 26 presents different patterns than those found in Figure 25, it could be said that the ratio $p:k$ influences index behavior over and above the influence of the increasing model size necessary to keep $p:k$ constant as the true model’s factor complexity increases.

We begin by exploring the effects of holding $p:k$ constant across different levels of $k$. Figure 25 presents the case in which a one-factor model is fit to true models with an increasing number of latent factors when the number of indicators per factor is held constant. Index value (y-axis) is plotted against true models with an increasingly complex latent structure (two, three, four, five, six, seven, or eight factors, x-axis). The number of factors is
Figure 25. Plots of fit index values vs. an increasing number of latent factors in the underlying model (2 – 8) when the same number of indicators per factor is held constant across the different numbers of factors. (3 indicators per factor, leading to 6, 9, 12, 15, 18, 21, and 24 indicators total for 2, 3, 4, 5, 6, 7, and 8 factors, respectively). Done for six values of factor loadings, with colors red, orange, green, blue, violet, and black correspond to loadings of .4, .5, .6, .7, .8, and .9, respectively. Solid lines correspond to a factor correlation of .1, dashed lines correspond to a factor correlation of .4. AGFI plot omitted due to similarities with GFI.
Figure 25 (continued). Plots of fit index values vs. an increasing number of latent factors in the underlying model (2 – 8) when the same number of indicators per factor is held constant across the different numbers of factors. (3 indicators per factor, leading to 6, 9, 12, 15, 18, 21, and 24 indicators total for 2, 3, 4, 5, 6, 7, and 8 factors, respectively). Done for six values of factor loadings, with colors red, orange, green, blue, violet, and black correspond to loadings of .4, .5, .6, .7, .8, and .9, respectively. Solid lines correspond to a factor correlation of .1, dashed lines correspond to a factor correlation of .4. AGFI plot omitted due to similarities with GFI.
a categorical variable, but neighboring points are connected for readability. For each level of $k$, each factor has three indicator variables. There are, then, 6, 9, 12, 15, 18, 21, and 24 indicators total for models with 2, 3, 4, 5, 6, 7, and 8 underlying factors, respectively. The six colored curves correspond to six loading sizes, with red, orange, green, blue, violet, and black corresponding to loadings of .4, .5, .6, .7, .8, and .9, respectively. All factor correlations are fixed either to .1 (solid lines) or .4 (dashed lines).

It is clear in Figure 25 that the ratio of indicators to factors has an effect on the patterns and values of NNFI, GFI, RMSEA, and gamma when compared to Figure 23, where $k$ varied but $p$ did not. Only CFI values and patterns appear significantly unchanged, and in both Figures 23 and 25, CFI values remain relatively constant as $k$ increases for factor correlations of .4, and show a gradual worsening of fit when factor correlations are .1.

When $k$ varied but $p$ did not (Figure 23), GFI, RMSEA, and gamma showed an improvement in fit as the size of the latent misspecification increased. NNFI showed an improvement in fit when factor correlation was .4 but a worsening of fit when factor correlation was .1. In Figure 25, when $p:k$ is held constant as $k$ increases, both GFI and gamma lose their ability to detect a change in the degree of misspecification as the number of latent factors present in the data increases. NNFI shows an improvement for both factor correlations of .1 and .4 as the number of underlying factors increases. RMSEA also shows a greater increase in fit as the number of underlying factors increases when the ratio of indicators per factor is held constant when compared to the case where the number of indicators total in the model was held constant.

Holding $p:k$ constant as $k$ increases results in an increase in overall model size as the number of latent factors in the data increases. Therefore, it could be argued that the results
found in Figure 25 are due solely to an increase in model size as the number of factors underlying the data increased, not due to holding the ratio of indicators per factor constant. To determine if this was the case, a further manipulation was done in which one more indicator per factor was included for every one factor increase in the true model. Figure 26 plots index value (y-axis) against a one-factor model fit to true models with increasingly complex latent structures (two, three, four, five, six, seven, or eight factors, x-axis). The number of factors is not treated as continuous, but neighboring points are connected in Figure 26 for readability. For each additional latent factor present in the true model, there is one additional indicator per factor, such that data with 2, 3, 4, 5, 6, 7, and 8 underlying factors have a total of 6, 12, 20, 30, 42, 56, and 72 indicators, respectively. The six colored curves correspond to six loading sizes, with red, orange, green, blue, violet, and black corresponding to loadings of .4, .5, .6, .7, .8, and .9, respectively. Factor correlation is fixed to .1 (solid lines) or .4 (dashed lines).

Based on comparisons between Figures 23, 25 and 26, we can see that the $p:k$ ratio affects the indices in different ways. To determine if the ratio of $p:k$ has an influence on index values over and above an increase in $k$, we compare Figures 23 and 25 first.

For CFI, any changes in $p:k$ appear to have little influence over its behavior. CFI behaved similarly when $p$ was held constant and $k$ increased (Figure 23) and when $p:k$ was held constant as $k$ increased (Figure 25). NNFI, GFI, RMSEA, gamma, and SRMR all exhibited changes in performance between Figure 23 and Figure 25. The differences seen for NNFI, however, have few practical implications for the index. That is, while the patterns of these indices differed slightly between the two Figures, their absolute values remained
Figure 26. Plots of fit index values vs. an increasing number of latent factors in the underlying model (2 – 8) when one more indicator per factor is included per every increase in the number of latent factors (such that the models with 2, 3, 4, 5, 6, 7, and 8 latent factors have a total of 6, 12, 20, 30, 42, 56, and 72 indicators, respectively). Done for six values of factor loadings, with colors red, orange, green, blue, violet, and black correspond to loadings of .4, .5, .6, .7, .8, and .9, respectively. Solid lines correspond to a factor correlation of .1, dashed lines correspond to a factor correlation of .4. AGFI plot omitted due to similarities with GFI.
Figure 26 (continued). Plots of fit index values vs. an increasing number of latent factors in the underlying model (2 – 8) when one more indicator per factor is included per every increase in the number of latent factors (such that the models with 2, 3, 4, 5, 6, 7, and 8 latent factors have a total of 6, 12, 20, 30, 42, 56, and 72 indicators, respectively). Done for six values of factor loadings, with colors red, orange, green, blue, violet, and black correspond to loadings of .4, .5, .6, .7, .8, and .9, respectively. Solid lines correspond to a factor correlation of .1, dashed lines correspond to a factor correlation of .4. AGFI plot omitted due to similarities with GFI.
Figure 26 (continued). Plots of fit index values vs. an increasing number of latent factors in the underlying model (2 – 8) when one more indicator per factor is included per every increase in the number of latent factors (such that the models with 2, 3, 4, 5, 6, 7, and 8 latent factors have a total of 6, 12, 20, 30, 42, 56, and 72 indicators, respectively). Done for six values of factor loadings, with colors red, orange, green, blue, violet, and black correspond to loadings of .4, .5, .6, .7, .8, and .9, respectively. Solid lines correspond to a factor correlation of .1, dashed lines correspond to a factor correlation of .4. AGFI plot omitted due to similarities with GFI.
similar enough to indicate that any changes in the p:k ratio will have little effect on the
behavior of NNFI in the context of a latent misspecification.

Holding p:k constant as k increases causes GFI and AGFI to lose the ability to detect
the change in the degree of misspecification as the number of latent factors in the true model
increases. While this difference does not matter for higher loadings, when loadings are low,
GFI and AGFI indicate good fit regardless of if the true model has two factors or the true
model has eight factors. This result is consistent with previous findings which show that GFI
and AGFI show an increase in fit for greater numbers of indicators.

When p:k is held constant, SRMR shows better fit overall, particularly when loadings
are .4. Differences in SRMR’s values between Figures 25 and 28 are small when the factor
correlations are .1. but when factor correlations are .4, SRMR shows better fit overall when
p:k is held constant.

RMSEA shows the most dramatic change in both absolute value and in pattern
between the case where p:k is held constant and the case where k increases but p does not. In
Figure 23, the index shows little ability to detect the increasing degree of misspecification
when a one-factor model is fit to true models containing an increasing number of latent
factors. In Figure 25, while RMSEA shows worse fit overall when compared to Figure 23, it
also shows an increase in fit as the number of latent factors increases. Thus, when the ratio of
p:k is held constant, RMSEA is more sensitive to changes in fit, but shows an improvement
in fit where ideally we would want an index to show a worsening of fit

Finally, we look at the comparison between Figures 25 (p:k is constant across all
levels of k) and 26 (p and k both increase, but p:k is not constant). This comparison addresses
the question of whether pattern differences found between Figures 25 and 26 are due to the
ratio $p:k$ being held constant or due to the general overall increase in model size as both $p$ and $k$ increase.

A comparison of Figures 25 and 26 reveals GFI to be the only index showing a difference in behavior between the two scenarios. When $p:k$ is held constant (Figure 25), GFI is insensitive to the increasing degree of misspecification as a one-factor model is fit true models with an increasingly complex latent structure. When there is an additional indicator per factor for every increase in the number of latent factors (Figure 26), we finally see GFI indicating worse fit as the size of the misspecification increases. All other indices behave similarly in Figures 25 and 26. These results suggest that the difference in performance seen in NNFI, GFI, RMSEA, gamma, and SRMR between Figures 25 and 26 is not necessarily related to the ratio of indicators per factor but is instead related to the general overall increase in model size as more indicators are added for data involving more latent factors. The behavior of GFI, on the other hand, may suggest that this index is more sensitive to the ratio of $p:k$, which doesn’t change in Figure 25 but changes both in Figure 23 and Figure 26.

4.3 Summary and Discussion

The figures in this section have explored the relationship between population CFI, NNFI, GFI, AGFI, RMSEA, gamma, and SRMR indices and model misspecification in the context of CFA models with a latent misspecification. Four nuisance variables (loading size, factor correlation size, model size, and model balance), along with the degree of latent misspecification (as measured by the discrepancy between the number of latent factors present in the true model and the number of latent factors in the fitted model), were examined to determine their effects on index values.
In most cases, the behaviors of CFI and NNFI were highly similar, as were the behaviors of GFI and AGFI and gamma and RMSEA. The focus of this section, then, is on CFI, GFI, RMSEA, and SRMR. NNFI, AGFI, and gamma are mentioned only when their behaviors are substantially different. Otherwise, the conclusions reached for CFI, GFI, and RMSEA can be applied to NNFI, AGFI, and gamma, respectively.

It can be argued that a misspecification at the latent level is more severe than a misspecification involving an omitted pathway. Ideally, then, indices would be more sensitive to the more severe latent misspecifications than those arising from omitted error covariances. Comparing index values found in this chapter to those from Chapter 2 and Chapter 3 (in which misspecification was due to one or more omitted error covariances), it was found that all indices showed worse fit in general when misspecification was due to an incorrectly modeled latent structure than when misspecification was due to omitted error covariances. Particularly apparent for CFI, latent misspecification led to model rejection at commonly used cutoff values more often than did misspecification due to omitted error covariances, which is an encouraging finding and suggests that the indices examined in this study are, in general, able to differentiate between major misspecifications and minor misspecifications.

In addition to allowing an assessment of index behavior with respect to the type (latent vs. omitted pathway) of misspecification, the figures presented in Chapter 4 also revealed changes in index performance due to the severity of the latent misspecification and due to several nuisance variables (particularly parameter size, model balance and model size).
Ideally, indices would be sensitive to the severity of the latent misspecification (as defined by the discrepancy between the number of factors present in the true model and the number of factors in the fitted model) and would indicate worse fit as this discrepancy between the true and fitted models increased. However, this was not the case for all indices. Contrary to the behavior we had hoped to find, GFI and SRMR showed an improvement in fit as a one-factor model was fit to true models with an increasing number of latent factors, regardless of factor correlation size (Figure 23). RMSEA behaved similarly, though values leveled off (for smaller loading sizes) as more latent factors were present in the true model. Only CFI performed ideally, showing a decrease in fit as the number of latent factors in the true model increased, though this behavior was only found at small factor correlation sizes. These results suggest that CFI may be the best index in terms of sensitivity to misspecified latent structures, but its sensitivity is dependent on the degree to which the factors in the true model are correlated. The behavior exhibited by CFI, however, still suggests a redeeming quality of the index. While unexpected and problematic nonlinear behavior has been seen for CFI in almost all the previous scenarios involving an omitted error covariance, its behavior with respect to appropriately detecting the degree of latent misspecification (at least in models with slightly correlated factors) may help to justify relying on this index in situations where misspecification is suspected to arise from an incorrect number of latent factors in the fitted model.

Additional model features and nuisance variables were shown to play a role in index performance. It was revealed that the ability of fit indices to detect an increasingly severe degree of misspecification is dependent on the size of the inter-item correlations of a model (Figure 24). Holding the average inter-item correlation constant across different levels of
factor complexity, RMSEA no longer showed an improvement in fit but instead showed a decrease of fit as a one-factor model was fit to true models with increasingly complex latent structures. This supports findings by Steiger (2000) who showed RMSEA to be sensitive to the size of the inter-item correlation. CFI showed a worsening of fit as the factor complexity of the data increased, but only at small sizes of factor correlation. This lends further support to the claim that CFI’s ability to appropriately detect latent misspecification is affected by the degree to which the factors in the data are correlated. Similar results were found for GFI and SRMR. Both indices showed an improvement in fit as factor complexity increased, regardless of factor correlation size when the inter-item correlations were not held constant (Figure 23). When they were held constant, however, decreasing fit was shown when factor correlation was set to .1. When factors were mildly correlated (factor correlations set to .4), holding the average inter-item correlation constant did little to alter index behavior. Such a finding points to the fact that GFI and SRMR are unable to appropriately detect a worsening of fit as a one-factor model is fit to data with increasingly complex latent structures when the factors in the data are even moderately correlated.

The behaviors of GFI, SRMR, and RMSEA in Figures 23 and 24 suggest that in addition to the effects of the inter-item correlation, loading size has an effect on these indices when detecting misspecification in the latent structure as well. When loadings were moderate to large (.6 or above), GFI, SRMR, and RMSEA were able to detect differences in fit when a one-factor model was fit to true models with a varying number of latent factors when loadings were moderate to large (.6 or above). When loadings were small (.4, .5), these indices showed about the same degree of misfit at all levels of $k$. Therefore, it can be said that
the stronger the indicators load onto specific factors, the more able these indices are to detect misspecification when a fitted model includes the wrong number of latent factors.

While factor correlation was shown to not impact the overall patterns of GFI, SRMR, and RMSEA, the behavior of CFI was shown to be dependent on the strength of the relationship between factors. When factor correlation was greater than .1, CFI behaved like GFI, SRMR, and RMSEA and showed an improvement in fit as the number of latent factors present in the data increases. It was only when factor correlation was .1 that CFI showed the behavior we desired. That is, CFI showed a decrease in fit for a one-factor model as the number of factors present in the true model increased.

In addition to loading size and factor correlation size, model balance was found to play a role in index performance. It was expected that better fit would be seen for a one-factor model fit to a more imbalanced true model than for a one-factor model fit to a balanced true model. As was found for misspecifications involving an omitted error covariance, this proved to be the case. The results confirmed what was expected and make sense intuitively. As the true models become more and more imbalanced, the more the covariance structures resemble that of a one-factor model than a two-factor model, and thus a truly one-factor model covariance structure will start to fit better. All indices reflected this, showing worse fit for the balanced true model and best fit for the most imbalanced true model.

The size of the model was also found to affect index performance. In this chapter, the effects of model balance were examined both in terms of the number of indicator variables included in the model \( p \) and the relationship between the number of indicator variables and the number of latent factors \( k \).
Independent of the number of factors in a model, the size of \( p \) was shown to have an influence on index behavior. Regardless of loading size and factor correlation size, fit improved as \( p \) increased for CFI and RMSEA and worsened for SRMR and GFI (Figure 20). AGFI shows a worsening of fit as \( p \) increased, but only when loadings were small to moderate (less than .7). The fact that SRMR and GFI showed a decrease in fit as the number of indicators in the model increases suggests that the addition of more indicators increases the sensitivity of these indices to latent misspecification. The improvement in fit seen for CFI and RMSEA when models increase in size may be of concern to researchers, but an examination of Figure 20 shows that while fit improves, only when loadings are small (.4) do RMSEA values fall within range of the commonly-accepted cutoff value of .06. That is, a one factor-model being fit to a true model with two factors will be rejected using RMSEA when \( p = 18 \) and accepted when \( p = 30 \), but only if loadings are .4 or smaller.

Knowing that CFI and RMSEA show better fit while SRMR and GFI show worse fit as the number of indicators increases could help guide researchers in situations where index values may lead to conflicting results. A researcher dealing with a misspecification at a latent level could be confronted with, for example, CFI and GFI values suggesting different conclusions regarding the fit of a model. The findings here suggest that the researcher could examine the size of the model to determine whether or not the size is cause of the difference. If the model is large enough, it may be a possibility that the size of the model is what is driving the difference between CFI and GFI values, rather than some other model component or additional misspecification. Suppose, for example, that a researcher with a large model was confronted with a CFI value that indicates good fit and a GFI value that suggests poor fit. Knowing that larger models exhibit better fit by CFI and worse fit by GFI, the researcher
could possibly test a smaller version of their model and examine the fit indices again. If CFI still shows acceptable fit for the smaller model but GFI now shows good fit as well, the researcher could conclude that the reason behind GFI indicating poor fit in the original model was due to the size of the model itself, not to a large misspecification.

Finally, it was found that for GFI, the ratio of indicators per factor played a role in index performance over that of just the number of indicators present in the model. When $p$ and $k$ both increased but $p:k$ was held constant (at a ratio of 3:1), GFI was unable to detect the change in the degree of misspecification as a one-factor model was fit to true models with increasingly complex latent structures. When both $p$ and $k$ increased but $p:k$ was no longer held constant across the different levels of factor complexity, GFI was more sensitive to the changes in misspecification and appropriately indicated worse fit as the factor complexity of the data increased. This finding suggests that, apart from GFI, the exact ratio of indicators per factor does not play a role in indices’ abilities to detect an increasingly large latent misspecification over and above the role of increasing the overall model size.
Chapter 5: Conclusion and Overall Discussion

The goals of this research were threefold. First, we wished to discover to what extent the relationship between the amount of model misspecification and the performance of several popular population-defined fit indices was moderated by model features other than the degree of misspecification itself. Second, we wished to determine whether the universal application of commonly-used cutoff values could be considered appropriate given how indices performed with respect to the above listed model features. Finally, based on the findings here, we aim to provide a set of loose, practical guidelines for researchers wishing to apply cutoff values in their own SEM-related research.

In this section, results are first presented by the different nuisance variables studied, including parameter size, model size, and model balance. This is followed with results broken down by index, focusing on some of the more striking findings for each index and offering some suggestions for their use based on what was found. Finally, the thesis concludes with a general discussion of the results, as well as some suggested avenues for future research.

5.1 Effects of Model Components

5.1.1 Loading size

The findings of this study support prior research showing that loading size influences index behavior in the direction of increased sensitivity. For both misspecifications involving omitted error covariances and misspecifications involving the latent structure, all indices were affected by loading size, with larger loadings consistently corresponding to increased sensitivity to misspecification than smaller loadings. However, not all indices are affected by loading size to the same degree.
RMSEA and GFI values were found to change the most as loadings changed, while SRMR values were generally least affected by loading size, especially when the degree of misspecification was small (e.g., a single omitted error covariance of .2). Knowledge of how these indices respond to changes in loading size can be beneficial to researchers, particularly when confronted with conflicting results based on different indices. For example, if the model is deemed as a good fit using SRMR but a poor fit by GFI or RMSEA, and the estimated loadings are high, the researcher may choose to evaluate model fit using SRMR because of the known sensitivity of GFI and RMSEA to loading size. Based on the results of this study, we support examining loading sizes when faced with conflicting results from different indices regarding model fit.

5.1.2 Factor correlation size

Factor correlation size was found to have minimal effects on index value in misspecification scenarios involving omitted error covariances, regardless of whether this covariance involved indicators of same or different factors. We found no instance where the size of the factor correlation would alter a researcher’s decision regarding model fit. When misspecification is thought to arise from an incorrectly specified latent structure, factor correlation size, in conjunction with loading sizes, exerts a greater influence on the fit index values. The weaker the correlation between factors, the greater the effects are of loading size on index’ ability to detect misspecification. That is, when factors approach orthogonality, all indices apart from CFI appear highly sensitive to latent misspecifications when loadings are high (.8, .9), but appear highly insensitive to such misspecification if loadings are low (.4, .5?). For example, if the data have an underlying two-factor structure where the factors are
orthogonal, a one-factor model would fit this data according to all indices apart from CFI if loadings were small enough (.4, .5).

5.1.3 Model size

The effects of model size were perhaps the most varied of all the model components we studied as to their effects on index value. In cases where misspecification was due to omitted error covariances, a nonlinear relationship was found between index value and the number of indicator variables included in the model. Such a relationship may suggest that there is a certain model size at which indices are most sensitive to this type of misspecification. The only other instance of this nonlinear behavior in the literature was documented by Savalei (2010) for RMSEA, also in a situation in which misspecification was due to an omitted error covariance.

The effect of model size is of little practical concern in cases involving an omitted error covariance, however. When loadings are small to moderate, changes in index value as $p$ increases does not affect the accept/reject choice for any index. Only when both the omitted error covariance and the loadings are large should researchers be aware of the fact that their model could be judged as a poor fit to the data at one value of $p$ but be judged as a well-fitting model at just a slightly higher or lower number of indicators.

The nonlinear pattern is present at least to some extent for all indices, both for those that explicitly include either the degrees of freedom or $p$ in their calculations (NNFI, AGFI, SRMR, RMSEA, and gamma, see Equations 2, 4, 5, 6, and 7) and those that do not (CFI and GFI, see Equations 1 and 3). The nonlinear behavior appears more prominently for indices that involve the degrees of freedom or $p$ in their calculations. The effects of model size were minimal when loadings were small and, in general, were small enough such that an increase
or decrease in the number of indicators in a model does not alter whether or not a model is accepted using the commonly-used cutoff values of any of the indices studied here.

5.1.4 Model balance

The current research examined the role of model balance on fit index behavior, a model feature that has previously been unexamined in the literature. In the case of an omitted error covariance, it was found that all indices showed best fit when the omitted error covariance occurred between variables loading onto the larger factor. In the case of latent misspecification, all indices showed best fit when a model was fit to the most severely imbalanced true model. Both these findings can be justified theoretically. Specifically, when it comes to an omitted error covariance, the size of the misspecification will be more “diluted” when it occurs within a larger factor as compared to a smaller factor, leading to better fit as reflected by the index. When it comes to a misspecified latent structure, the effects of model imbalance on index performance is somewhat analogous to the effects of increasing the overall number of indicators in a one-factor model. That is, in each case, the severity of the misspecification is “diluted” within the larger matrix (for a one-factor model with a greater $p$) or within the portion of the matrix corresponding to the larger factor (for a two-factor model where one factor is larger than the other).

The effects of model imbalance on index values were generally small, particularly when the misspecification was due to an omitted error covariance, unless a model was severely imbalanced (e.g., one factor has seven indicators while the other has 17). Thus, an unequal number of indicators per factor will most likely not affect the decision regarding model fit given commonly used cutoff values.
5.2 Index-Specific Effects

5.2.1 CFI and NNFI

Perhaps the most striking set of findings emerging from the research presented here surrounds the performance of CFI and NNFI in misspecification scenarios involving omitted error covariances. Regardless of any other model components (factor loading size, factor correlation size, model size, or model balance), CFI and NNFI exhibit an alarming nonlinear trend with respect to the size of a single or multiple omitted error covariances.

Relying on either of these indices could lead a researcher to accept a model as well-fitting when that model contains either a small or large omitted error covariance, but would lead a researcher to reject a model if the model contained an omitted error covariance of a moderate size (e.g., $\psi = .4$). Based on the fact that a model with a larger misspecification could fit better than a model with a smaller misspecification according to these indices, we advocate abandoning the use of CFI and NNFI in cases where it is suspected that misspecification is suspected to be minor (e.g., due to omitted error covariances rather than misspecified latent structure) and advocate using either index with caution when there is little knowledge as to what is causing the misspecification.

To gain a better understanding of the nonlinear behaviors of these indices, we turn to the equations for CFI and NNFI, reprinted here:

$$\text{CFI} = 1 - \frac{F_M}{F_I}$$  \quad (4)

$$\text{NNFI} = 1 - \frac{F_M}{F_I} \times \frac{df_I}{df_M}$$  \quad (5)

In the population, these indices are functions of both the minimized fit function of the proposed model ($\widehat{F}_M$) and the minimized fit function of the null or baseline model ($\widehat{F}_I$). As in
most applications of CFI and NNFI, the baseline model defined in this study was one in which all variables were modeled as uncorrelated in the population.

It was suspected that the nonlinear behavior observed for CFI and NNFI when misspecification was due to omitted error covariances had to do with the relative rates of change of $\bar{F}_M$ and $\bar{F}_I$ as the size of the omitted error covariances increased. To gain a visual understanding of the rates of change, we plotted the rate of change of $\bar{F}_M$ and $\bar{F}_I$ separately against an increasing omitted error covariance in a one-factor model with eight indicators that failed to include an error covariance present in the population (this is the same model used in Figure 1).

The plots (not included here) revealed that the rates of change were indeed different and depended both on the size of the omitted error covariance and the size of the loadings. The rates of change of $\bar{F}_M$ and $\bar{F}_I$ differed not only from each other but at different sizes of omitted error covariance as well, suggesting that there is some relationship between the size of the misspecification and its influence on the minimized fit functions for the baseline model and the minimized fit functions for the proposed model. The fact that similar nonlinear patterns were found for both CFI and NNFI suggests that the rates of change of $\bar{F}_M$ and $\bar{F}_I$ are independent of the degrees of freedom of the proposed and baseline models.

Contrary to the troubling nonlinear behavior seen when misspecification is due to omitted error covariances, the behavior of CFI and NNFI in cases where misspecification is due to an incorrectly modeled latent structure gives us reason to support the continuing use of these indices. Of the population indices we examined, CFI and NNFI were found to be the most sensitive to misspecifications involving an incorrect number of latent factors. They accurately reflect the increase in the degree of misspecification when a one-factor model is fit
to true models with increasingly complex latent structures. This is an encouraging finding, as it suggests that while CFI and NNFI may show strange behavior when the source of misspecification is due to omitted error covariances, the two indices are more sensitive to and can more accurately reflect latent misspecification, which can be considered a more severe form of misspecification than misspecification due to omitted error covariances.

We note, however, that researchers should be aware of the degree to which CFI and NNFI are sensitive to latent misspecification. In some of the situations we examined involving latent structure misspecification, CFI and NNFI values never increased above .8 regardless of the influence of other modeling components such as loading size or model size. This suggests that CFI and NNFI are extremely sensitive to latent misspecification. This sensitivity can be considered one of the strong characteristics of the index.

When index values are consistently low in a given modeling situation, it may make it easier to interpret such values in relation to a given cutoff value, such as .95 (the most commonly used cutoff value for CFI and NNFI). Obtaining a CFI value of .8 provides researchers with the assurance that if their model does not contain the appropriate number of factors (that is, the number of factors that are present in the true model), the misspecification will always lead to model rejection using CFI.

The use of CFI values in model comparison situations can benefit researchers as well. If a researcher is interested in determining which of two models fits better in a situation that may involve latent misspecification, they can make use of CFI values to compare the fit of the two models to each other even though, on the absolute scale, both values indicate general poor fit by the CFI.
5.2.2 GFI and AGFI

The current research indicates that unlike CFI and NNFI, both GFI and AGFI appropriately reflect changes in model fit when misspecification is due to an omitted error covariance. That is, as the size of a single omitted error covariance increases (or the number of omitted error covariances in the model increases), GFI and AGFI show worse fit. Thus, we recommend the use of GFI or AGFI over the use of CFI or NNFI if it is suspected that misspecification in a model is due to one or more omitted error covariances.

In the context of latent misspecification, both GFI and AGFI show an improvement in fit as the number of latent factors present in the true model increases, while the 1-factor model is fit to data. This is contrary to the desired behavior. However, at moderate to high loading sizes, this increase in fit is of no practical concern, as index values remain below the commonly used cutoff value of .95, indicating that the proposed one-factor model does not fit the true model for any value of \( k \) greater than one. Thus, GFI and AGFI can be safely used in the context of latent misspecification if loadings are of at least a moderate size (greater than .4 or .5).

We also found GFI and AGFI values to be strongly influenced by loading size, becoming more sensitive to both types of misspecification as the size of the loadings increases. We suggest that researchers be particularly aware of loadings when employing GFI or AGFI, as changes in loading size may affect whether a model is accepted as fitting. Particularly, GFI and AGFI are most sensitive to misspecification when loadings are high, and models with a particular-sized omitted error covariance may be accepted when loadings are small but rejected when loadings are large.
Finally, in all modeling scenarios examined here, GFI and AGFI performed extremely similarly, even in cases where the number of indicator variables was the model component being varied. This is an interesting finding, considering that Joreskog and Sorbom (1981) developed AGFI as an adjustment to GFI that was supposed to compensate for the increase in fit seen for GFI as more parameters are included in a model.

5.2.3 RMSEA and gamma

Both RMSEA and gamma performed as expected in misspecification cases involving omitted error covariances. That is, these indices were sensitive to increases in either the size of an omitted error covariance or the inclusion of multiple omitted error covariances and indicated worse fit accordingly.

As was the case with GFI and AGFI, the size of the loadings had a significant effect on the ability of RMSEA and gamma to detect misspecification, both in the case of omitted error covariances and in the case of a misspecified latent structure. Of the indices studied, RMSEA and gamma showed the greatest differences in index value across different loading sizes. Thus, we suggest that researchers who utilize RMSEA or gamma be particularly aware of the effect of loading size, as it may lead to different decisions regarding model acceptance or rejection.

The most interesting results for RMSEA and gamma involved models with misspecified latent structure. When a one-factor model was fit to data with an increasing number of latent factors, other indices showed either a decrease in fit (CFI and NNFI) or an increase in fit (GFI, AGFI, and SRMR). RMSEA and gamma, on the other hand, appeared to be almost completely insensitive to an increase in misspecification due to a one-factor model being fit to true models with increasingly complex latent structures. Holding the average
inter-item correlation constant across different levels of factor complexity improved the sensitivity of these two indices such that worse fit was shown as the number of latent factors in the true model increased. Thus, part of the reason there is little change in index value as the number of latent variables in the true model increases is the sensitivity of RMSEA and gamma to the average inter-item correlation.

Finally, even though gamma further adjusts RMSEA for parsimony (see Equation 10), reliance on gamma versus RMSEA did not lead to differences in model acceptance or rejection in the situations studied in this research. Because RMSEA is already an extremely popular index, there is no advantage to encouraging users to use the gamma index in addition to or instead of RMSEA.

5.2.4 SRMR

As already noted, SRMR was found to be the least affected by loading size of all the indices studied here. Particularly in cases where misspecification is due to omitted error covariances, changes in loading size resulted in little change in SRMR values. This lack of sensitivity to loading size suggests that SRMR has as an advantage over indices such as RMSEA. For example, in a model with high estimated loading values, RMSEA would be extremely sensitive to even the most trivial misspecification, while SRMR would not be. However, SRMR was found to have a serious disadvantage as well. Perhaps the most notable finding with respect to SRMR is its general insensitivity to both of the misspecification types examined in our study. Regardless of whether misspecification is due to omitted error covariances or to an incorrectly modeled latent structure, SRMR was less able to detect misspecification using the traditional cutoff value of .08 than the other indices studied. Even in cases involving some of the most severe misspecifications (e.g., a model with ten
moderately-sized omitted error covariances), SRMR values failed to increase above .08. However, if SRMR values are high, a researcher could be fairly certain that the fitted model is seriously misspecified, and furthermore that the index value is not an artifact of high loadings or other model features.

SRMR’s lack of sensitivity using the cutoff value of .08 may also suggest that a lower cutoff should be used. While some instances in the literature suggest raising the cutoff value to .09 or .10, a lower cutoff value has never been suggested for SRMR. The results here demonstrate that cutoff values of .08, .09 and .10 may be too liberal for the actual behavior of SRMR.

Finally, even if a lower cutoff value were used, SRMR’s behavior in scenarios involving latent misspecification suggests that it may not be an appropriate index to use in such cases. Like GFI and AGFI, SRMR indicates an improvement in fit as the one-factor model is fit to data with an increasing number of latent factors. Unlike GFI and AGFI, however, SRMR values are small enough (less than the commonly used cutoff of .08) to accept the one-factor model as well-fitting even at moderate loading sizes.

If researchers are unsure of the possible source of the misspecification in their model (e.g., whether it arises from omitted error covariances or from a misspecified latent structure), we support the use of multiple indices to best determine the origin of the problem. Even though SRMR (like GFI and AGFI) may not provide reliable information about the degree of misspecification when it is due to an incorrectly modeled latent structure, an examination of SRMR with another index like CFI, which does show a worsening of fit as the latent structure misspecification increases, could provide additional information.
5.3 Concluding Remarks

The goals of this study were three-fold: to assess index behavior with respect to the influences of nuisance variables, to evaluate the appropriateness of applying universal cutoff values across all modeling scenarios, and to offer some guidance as to what indices perform best under what modeling situations. With respect to the first goal, the results of our study clearly show that all indices are affected to some degree by various model features, including loading size, factor correlation size, model size, and model balance.

With respect to the second goal, it is clear that the results of this research do not support a universally applied cutoff value for any of the indices studied. While we did not study the appropriateness of the specific cutoff values plotted in our figures, the wide range of values taken on by the fit indices in the different scenarios presented here makes it clear to us that any universally applied cutoff value would not be appropriate for all modeling situations. If a single variable were to be identified that should affect the chosen cutoff value for any index, it would be (with the exception of SRMR) loading size. Broadly speaking, cutoffs should be higher for lower loadings and lower for higher loadings to detect the same size of misspecification.

With respect to the third goal, brief guidelines to help researchers use and interpret the fit indices were given in Section 5.2. To restate briefly, in cases where it is reasonable to assume that misspecification may be due to something minor like an omitted error covariance, we advise against the use of CFI or NNFI and promote the use of GFI, AGFI, or RMSEA, perhaps in conjunction with SRMR. On the other hand, if there is reasonable evidence to suggest that misspecification is due to an incorrectly modeled latent structure, we support the use of CFI or NNFI. In all cases, researchers should be aware of loading sizes, as
we have found loading size to be influential on index behavior in all scenarios, and
researchers should also be aware of factor correlation sizes when dealing with situations
where misspecification may be due to an incorrectly modeled latent structure.

The present study has several limitations. The investigation was limited to the study
of two misspecification types: misspecification due to omitted error covariance and
misspecification due to an incorrectly specified latent structure. Future research should study
other sources of misspecification to gain a broader understanding of index behavior. Limited
previous research by Savalei (2010) has examined RMSEA behavior in the context of
omitted cross-loadings. Future research could possibly expand on the scenarios explored by
Savalei and include an investigation of the additional indices presented here.

Future research should also expanding on the findings here regarding the nonlinear
behavior of CFI and NNFI found, for example, in Figure 1. Research here was limited to
examining the plots of the relative rates of change of $F_{M}$ and $F_{I}$, and doing so solely in the
case of an omitted error covariance. More information regarding why CFI and NNFI display
this nonlinear behavior could be gathered by plotting the behaviors of $F_{M}$ and $F_{I}$ with respect
to changes in other model components and in cases where misspecification arises from
something other than an omitted error covariance. An analytic approach to understanding the
strange nonlinear behavior of these indices could also be pursued.

Finally, this study was the first to examine the effects of model balance on index
behavior. However, due to the fact that so many other model components were being studied
as well, the effects of model balance were only examined in a few basic scenarios. Future
research could involve a more in-depth examination of how the number of indicators loading
onto the factors of a model effects index value, perhaps by extending imbalance to models
with more than two factors or by examining imbalanced models with different sources of misspecification, such as omitted factor-loadings.
Bibliography


Herting, J. R., & Costner, H. L. (2000). Another perspective on “the proper number of factors” and the appropriate number of steps. *Structural Equation Modeling, 7*(1), 92–110.


