GOODNESS OF FIT AS A SINGLE FACTOR STRUCTURAL EQUATION MODEL

by

JAMES A. GULLEN

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2000

MAJOR: EDUCATIONAL EVALUATION AND RESEARCH

Approved by:

Shlomo S. Sawilowsky 10/16/00

Advisor Date

UMI Number: 9992202

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

DEDICATION

I dedicate this dissertation to my mother and father. I have never had to look further than to them for excellent examples of how to be a student, a teacher, a professional educator, or a person.

ACKNOWLEDGMENTS

As with any significant project, I could not have completed this dissertation without the help of many people, many of whom I have never met. I would like to thank the community members of the SEMNET discussion group for their support in supplying me with the fit statistics that made up my data set. I can only hope that I will be fortunate enough to have the opportunity to meet each of them to thank them personally. I would like to thank my doctoral committee. Specifically, I would like to thank Dr. Frank Castronova for his helpful insights into how to write a dissertation. I would like to thank Dr. Richard Kaczynski for his enthusiastic support during this project as well as the opportunity to practice statistics in a context other than education. I am grateful to Dr. Lori Rothenberg for exposing me to this fascinating and complicated branch of statistics. Without her, I would not have had the courage to pursue this area of work. I also thank Dr. Shlomo Sawilowsky. While I am honored to have Dr. Sawilowsky as my major advisor, I am even more appreciative for his common sense approach to statistics. Dr. Sawilowsky has nurtured the Educational Evaluation and Research Department at Wayne State, and I am proud to be able to say that I am a part of it.

I must thank my family, especially my wife, Suzanne. Without their steadfast support throughout all aspects of my education, I would not have been able to see this to completion.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Background of the Study

Structural equation modeling is becoming a popular statistical tool in the social sciences for a number of reasons. One of the primary reasons is that the technique is general and thus encompasses a wide range of useful models. These models include confirmatory factor analyses, traditional linear regressions, path analyses, and combinations of the three. Although the techniques used in structural equation modeling are general, the application typically involves five steps: model specification, model identification, model estimation, model evaluation and finally model respecification (Bollen & Long, 1993).

The first step, model specification, involves the explicit statement of the relationships between the variables (both observed and latent) in the model. In software packages, these relationships are typically stated in one of three forms: mathematical equations, matrices, or path diagrams. It is possible to move between these forms. For example, given a structural equation model's path diagram, it is possible to uniquely determine the appropriate matrices for that model.

Model identification is the second step in fitting a structural equation model. In this step, mathematical rules are applied so that it is possible to determine if all of the parameters that need to be estimated can, in fact, have values computed. The situation that is necessary is similar to that in algebra where as many equations as variables are needed in order to determine unique solutions. Unlike in algebra, however, there is a benefit to having more equations than variables. These over-identified models permit the calculation of fit statistics for the evaluation of model fit.

Model estimation involves using variance/covariance matrix data from the observed variables to calculate the parameters of the model that need to be estimated. A number of different methods are commonly used to fit structural equation models to data. Some of the more popular methods are maximum likelihood, generalized least squares, weighted least squares and asymptotic distribution free (Browne, 1984). The choice on method of estimation is based on a number of considerations. Some things that influence the choice of estimator include properties of the sample data (e.g. normality or lack thereof), whether the researcher is looking to maximize the amount of variance explained by the model, and the availability of significance tests for model parameters.

Once a structural equation model is specified and the free-parameters estimated, the fit of the proposed model to the sample data is evaluated. Several methods can be used to evaluate the model. A Chi-squared test of discrepancy is available, residuals can be analyzed, and many fit indices are calculated in software packages to aid the researcher in assessing the proposed model. These procedures are necessary for the evaluation of structural equation models as the testing of these models is not as straightforward as is the testing of models in other multivariate dependence techniques (Hair, Anderson, Tatham, & Black, 1995). In fact, structural equation models should not, in most instances, be evaluated using just one fit statistic. Shumacker and Lomax (1996) made this point succinctly that, "*no one* index serves as a definite criterion for testing a hypothesized structural equation model." (p. 135).

Model re-specification may or may not be part of a researcher's use of structural equation modeling. Whether or not a researcher modifies a model after evaluation is determined by the researcher's purpose. Jöreskog (1993) drew distinctions between three uses of structural equation modeling: confirmatory, comparative, and model generating.

Only model generating would permit respecification after the initial, hypothesized model has been estimated. Jöreskog asserted, however, that this is the most frequent use of structural equation modeling (Jöreskog, 1993).

Structural equation modeling also has the desirable ability to model relationships between latent (unobserved) variables (referred to as factors) and relationships between latent variables and observed variables (referred to as indicators). In this way, structural equation modeling, in its most general form, can be thought of as the union between factor analysis and path analysis. Introductory text book authors start with path analysis and factor analysis as departure points toward structural equation models because path and factor analyses are special cases of the full structural equation model (see, e.g., Maruyama, 1998).

What most sets structural equation modeling apart from other multivariate techniques is its ability to take measurement error into account. By contrast, the other multivariate techniques that employ the general linear model assume that measurements on the observed variables are free from error. Although many of the general linear modeling techniques have corrections for measurement error, structural equations model these measurement errors explicitly. The researcher (based on the reliability of the instrument) may specify the errors or the computer software may estimate the errors. In this way, the modeling of error makes models more realistic and thus more useful (Bollen, 1989).

Aside from application, the two areas of structural equation modeling that are receiving the most research are model estimation and model evaluation. New estimation methods are being introduced that provide appropriate estimates of parameters under a variety of conditions contained in real-world data, (e.g., non-normality). New indices that attempt to evaluate a model in a number of different ways have been developed. This is a result of the complexities of structural equation models. At the present, a fit index has not been developed that is able to capture all of the elements that would be present in a good model in a single fit index.

The Need for Studies on Fit Indices of Structural Equation Models

Although there is no shortage of fit indices proposed to evaluate the fit of structural equation models, there is little consensus as to what values should be used to indicate a good fit of the model to the data. Most of the fit indices currently used in structural equation modeling have unknown sampling distributions. It is, therefore, not possible to enter a table for one of the fit indices and retrieve the number associated with, for example, the critical value associated with the most extreme 5% of the statistic's distribution. Rules of thumb have been developed for some fit statistics as to what values represent a good fitting model, but most of these rules of thumb have been called into question due to the results of simulation work (see, e.g., Hu and Bentler, 1999).

The formulas for various fit indices also complicate model assessment. The formulae for many fit statistics are not mathematically derived. Different fit indices are designed to assess specific aspects of model fit. For example, one group of fit indices is designed to look at the relative complexity of the model and reward (with better fit statistic values) a more simple, or parsimonious model. To accomplish this, another fit statistic is typically adjusted by a factor based on model complexity. Unfortunately, there is no consensus on the correct way to account for model complexity. This has led to a number of different fit statistics that attempt to measure the model fit to the data and the relative complexity of the model. Structural equation models are complex and it is proving to be difficult to find statistics that capture the many aspects of good model fit.

Although numerous fit statistics assess different aspects of a model it is generally held that all fit statistics will generally agree as to the quality of the model even if they may take on different values. Tabachnick and Fidell (1996) summarized this conventional wisdom, "Good-fitting models produce consistent results on many different indices in many, if not most, cases" (p. 752).In fact, it is a commonly held belief that although the fit indices do provide different information about the model, they will all tend to give the same message about the model. Therefore, it seems appropriate for a researcher new to structural equation techniques to focus on one or two fit statistics until they are more familiar with the finer points of structural equation modeling.

The Problem, Its Purpose and Significance

The purpose of this study is to examine the conventional rule of thumb that although the structural equation model fit statistics consider different hypotheses and behave differently under various conditions, they tend to agree with each other for typical structural equation models. This is an important methodological concern because it is a rule of thumb that is given to students when they are first learning to use the techniques of structural equation modeling. It is important to know if, in fact, this convention is useful and accurate in real world work.

Overview of the Study

This study attempts to determine the viability of the conventional rule of thumb that all structural equation fit indices tend to say the same thing about any given model through the employment of a confirmatory factor analysis. Sets of popular fit indices will be obtained from researchers using structural equation modeling techniques. Sets of fit indices will be obtained from classroom examples as well as from researchers who participate in the discussion group listserver for structural equation modeling (SEMNET).

Additionally, sets of indices will be included that were present in textbooks that cover structural equation modeling techniques and provide samples of output from software packages. Examples covered in the manuals for structural equation modeling software packages also will be included in the analysis. The fit statistics were chosen for inclusion in this study based on their availability in currently popular structural equation modeling software packages (i.e. AMOS, EQS, LISREL, SAS). Only those fit indices that are provided by each of these software packages were included in the confirmatory factor analysis. AMOS 4.0 (Arbuckle, 1999) was used to compute the factor analysis on the sets of fit indices obtained.

Research Questions

The primary question that the current study seeks to answer is whether or not the wide range of fit indices available to researchers employing structural equation modeling techniques tend to say the same thing about any particular model. Is the level of agreement between the proposed model and the obtained data that is expressed by fit indices consistent across the range of fit indices available? The underlying issue is whether or not a given model may have good fit according to some fit statistics and may be shown to have poor fit when different fit indices are consulted. If the results of this study indicate that the notion that all of the fit indices tend to agree with each other is not tenable, another question will be raised: which of the structural equation model fit indices do tend to agree with each other? Furthermore, it will be interesting to see if the groups of fit statistics that do load on common factors follow along the lines of the current classification guidelines proposed for these indices.

Assumptions of the Study

1. Each set of fit indices consists of fit indices measuring the same structural equation model. Each statistic is related to the same model in terms of specification, estimation method, and parameter values.

2. If the fit indices do tend to agree with each other for any particular structural equation model, this fact should show up in a factor analysis as a single dominant factor.

3. If the factor analysis yields more than one dominant factor, the multiple factors will fall along lines that have important implications for the application of structural equation modeling.

Limitations of the Study

The largest limitation of the present study concerns the sets of fit indices that will be obtained. Although a large number of fit indices will be obtained in an attempt to get a representative sample of the multitude of possible structural equation models possible, the sample obtained will not be random. It is possible that certain structural model configurations will not be represented. Furthermore, the exact models are not known. The present study will not be able to detect, for example, if the fit indices behave differently under different estimation methods. The present study is also too general to provide recommendations as to values for the various fit indices that would be indicative of "close fit" between the obtained data and the estimated model.

Chapter 2

Review of Literature

The assessment of model fit in structural equation models has seen a rapid growth over the past twenty years. Traditional statistical tests of discrepancy were a natural, first attempt to distinguish accurate models from poor ones. It was soon learned, however, that the traditional chi-square test of discrepancy developed too much power due to the fact that structural equation models required such large data sets in order to estimate model parameters. Models that showed trivial amounts of misfit were rejected due to the fact that such large numbers of data points were being used in the analysis (Fan, Thompson, & Wang, 1999).

This realization that the chi-square statistic was not performing satisfactorily led to an explosion of proposed fit indices. Many of these newly proposed fit indices had at their heart the chi-square test statistic. Different adjustments were proposed and incorporated into the formulas for these new indices that were supposed to attenuate the undesirable properties of the stand-alone chi-square.

Although tests based on model-data discrepancy are important and useful, another avenue towards assessing the accuracy of a model was pursued. The path analysis/ linear regression aspect of structural equation modeling suggested the use of residuals to try to capture the degree to which the proposed model fit the obtained data. A number of fit indices are based on the relative sizes of various residuals after a structural equation model has been fit to data (Hair, Anderson, Tatham, & Black, 1995).

With the numerous fit indices available, as well as the increased application of structural equation modeling techniques by researchers, the properties of the fit indices began to be investigated. It was subsequently determined that many of the fit indices behaved differently from each other under a number of different conditions. Furthermore, the properties of a "good" fit index for structural equation models became a topic for consideration. This consideration led to classification schemes for the fit indices (Tanaka, 1993).

As the desirable properties for fit indices were becoming better defined, a large number of studies were undertaken to determine the behaviors of specific fit indices. Many of these studies employed computer simulations (e.g., Ding, Velicer, & Harlow, 1995). Some of these studies consisted of models fit to simulated data or Monte Carlo studies (e.g., Chou, Satorra, & Bentler, 1991). Other studies tried to clarify the behaviors of fit statistics as models were fit to the data that were misspecified to varying degrees (e.g., Bandalos, 1997). Other factors that have been hypothesized to influence fit indices and have been studied are sample size and estimation method (e.g. Fan, Thompson, & Wang, 1999).

The literature on structural equation modeling is expanding quickly with studies determining the effects of very specific situations on particular fit indices. It is not known how well these computer generated data sets reflect the characteristics of data sets that researchers used in practice. Furthermore, the results of these simulations have not, to a large degree, been compared to what researchers are dealing with in their applications of structural equation modeling. Studies are elucidating the differences between fit statistics in specific situations. The fact that so many differences in the properties of the fit indices are being detected naturally raises a question that has, to this point, gone unanswered. Specifically, do the fit indices tend to indicate the same thing about the quality of a particular structural equation model even though it is known that the indices themselves behave differently for a number of reasons?

The Null Hypothesis of a Structural Equation and a Statistical Test of Discrepancy

The null hypothesis for a structural equation model is:

$$\Sigma = \Sigma(\Theta)$$

The null hypothesis is that the population covariance matrix equals the implied covariance matrix formed by the sample and the model parameters. In the equation, Θ is the vector of model parameters.

The null hypothesis states that two matrices are identical. A natural statistical test to employ, then, is a test of discrepancy. Mathematically, it can be shown that (N-1)F is distributed as chi-squared. In this expression, N is the sample size and F is the fitting function minimized by the estimation method employed. Chi-squared tests are evaluated for a given number of degrees of freedom. The degrees of freedom for the chi-squared test is

$$df = \tfrac{1}{2}\,(p+q)(p+q+1) - t$$

where (p + q) is equal to the number of observed variables and t equals the number of free parameters estimated for the model (Jöreskog & Sörbom, 1996).

An undesirable property of the chi-squared test statistic in general is the fact that its power levels are affected by sample size. The test may lack adequate power to detect meaningful departures from the null hypothesis for small samples. Conversely, if samples sizes are too large, the chi-squared test becomes too powerful and detects unimportant departures from the null hypothesis (Cochran, 1952). The result is that all models and their associated hypotheses will be rejected on statistical grounds (Bentler & Bonnet, 1980; Jöreskog, 1969). Bollen (1989) also pointed out that the chi-squared statistic is based on fundamental assumptions. These assumptions include a multinormal distribution of observed exogenous variables, the analysis of the covariance matrix, adequate power in the test via sufficient sample size, and a correctly specified model. Bollen stated that, in practice, at least one these assumptions is violated (Bollen, 1989. pp. 266-8).

A test that is closely related to the chi-squared is the relative chi-squared (Wheaton, Muthen, Alwin, & Summers, 1977). This statistic is formed by dividing the obtained chi-squared statistic by its associated degrees of freedom. Because this index is based on the chi-squared statistic it is influenced by the sample size. However, there are no tables of critical values for this index, although there are rules of thumb based on experience. One common rule of thumb is that $\chi^2/df > 2.0$ is indicative of poor fit (Byrne, 1989). Although rules of thumb are useful, particularly in the model generating aspect of structural equation modeling, more interpretable measures are desirable for model confirming purposes.

Other Statistics Used to Evaluate Structural Equation Models

A useful measure of model-data fit that does not rely on the chi-squared distribution is the root mean square residual (RMR). James Arbuckle (1997) succinctly interpreted the RMR as "...the square root of the average squared amount by which the sample variances and covariances differ from their estimates obtained under the assumption that your model is correct." (p. 571). Although residual analysis is useful for assessing the quality of a model, particularly via model comparison, there are no absolute standards relating the size of the RMR to a well fitting model. As with all residuals, however, values of RMR closer to zero indicate better model-data agreement. An additional consideration with the RMR is that it is based on an arithmetic mean and therefore is not robust to the presence of outliers in the data, which in this case are the residuals. Although it is probably easiesr to understand the RMR conceptually from the standpoint of a residual, the formula to calculate the RMR is given as:

$$\mathrm{RMR} = \{2\ \Sigma\ \Sigma\ (s_{ij} - e_{ij})^2/[p \times (p+1)]\}^{1/2}$$

where s_{ij} are elements of the sample variance/covariance matrix and e_{ij} are elements of the model-implied, or estimated, variance/covariance matrix (Marsh, Balla, & McDonald, 1988).

Jöreskog and Sörbom (1984) provided the goodness of fit index (GFI) in an attempt to have a statistic with upper and lower bounds. The GFI can be thought of as a proportion of variance accounted for by the proposed structural equation model. Practically, it ranges between 0 and 1 with a value of 1 indicating perfect fit. The GFI is defined as

$$\mathrm{GFI} = \mathrm{tr}(\sigma'\mathbf{W}\sigma)/\ \mathrm{tr}(s'\mathbf{W}s)$$

where the numerator is based on the estimated model covariance matrix and the denominator is based on the sample covariance matrix (Tabachnick & Fidell, 1996). Mathematically, it is possible to obtain negative values for the GFI. Fortunately, this happens only when there is a serious discrepancy between the model and the sample data.

The adjusted goodness of fit statistic (AGFI) was proposed by Jöreskog and Sörbom (1984). The AGFI is based on the GFI with an adjustment for model complexity. This adjustment is based on the degrees of freedom in the proposed model relative to the total degrees of freedom possible for any model that could be fit for the particular data set. Like the GFI, a value equal to 1 for the AGFI indicates perfect model fit. The AGFI is given by the formula

$$AGFI = 1 - \{(p+q)(p+q+1)/2d\}(1-GFI).$$

In this formula p+q is equal to the number of observed variables included in the analysis, d is equal to the degrees of freedom for the model and GFI is the value of the goodness-of-fit index (Jöreskog & Sörbom, 1996).

The parsimony goodness of fit index (PGFI) was proposed by Mulaik, et. al. (1989) as an alternative to the AGFI. Like the AGFI, the PGFI adjusts the goodness of fit index by a factor involving the degrees of freedom in the proposed model and the degrees of freedom in the null model. The PGFI is calculated using the formula

$$PGFI = (d/d_b)GFI$$

where the ratio d/d_b is calculated by dividing the degrees of freedom of the model being evaluated by the degrees of freedom for the null model (Arbuckle, 1999). The adjustment contained in the PGFI is different from the AGFI, but interpretation is similar. All other things being equal, a simpler model will have a value closer to 1.0.

The Tucker-Lewis index (Tucker & Lewis, 1973) is another incremental fit statistic that takes into account model complexity. The Tucker-Lewis Index (TLI) was also presented by Bentler and Bonnet (1980) as the non-normed fit index (NNFI). (The TLI was originally developed for factor analysis, a specific application of structural equation modeling. Bentler and Bonnet generalized the fit statistic to structural equation modeling, in general, in the form of the NNFI.) As with the other incremental fit and comparative fit indices, values closer to 1.0 indicate better model data agreement. The calculation of the Tucker-Lewis Index is

$$TLI = (\chi_n^2/df_n - \chi_t^2/df_t)/(\chi_n^2/df_n - 1.0).$$

In the formula for the Tucker-Lewis Index, the subscript n represents a value from the null, or independence model. A subscript of t represents a value from the target, or tested model (Marsh, Balla, & McDonald, 1988).

The incremental fit index (also known as Δ2) was introduced by Bollen (1989) to create a fit index that was independent of sample size and took model complexity into account. Like many of the other fit indices, values close to 1.0 indicate very good fit and values greater than 1.0 may indicate overfitting the model to the data (i.e. not having enough degrees of freedom in the model). The formula for the IFI is

$$IFI = x^2_b - x^2_m / x^2_b - df_m ,$$

where the subscripts denote which model the value comes from. A subscript b indicates a value from the null or baseline model, and a subscript m indicates a value from the model being tested.

The root-mean-square error of approximation or RMSEA (Brown & Cudek, 1993) is a fit index that has generated widespread interest among researchers using structural equation models. The RMSEA is defined as the square-root of the discrepancy function minimized by the estimation method divided by the degrees of freedom in the model (Arbuckle, 1999, Steiger & Lind , 1980, and Steiger , 1990). Like the TLI, Steiger (1980) developed this index for factor analysis. Browne and Cudeck (1993) adapted and presented the RMSEA for use with the general approach of structural equation modeling. The RMSEA is unlike many of the previously mentioned fit indices in that values close to 1.0 do not indicate good model fit. As its name implies, the RMSEA is an average error, or residual. As in other multivariate techniques, good models have relatively small residuals. Therefore, values of the RMSEA close to 0.0 are desired. Browne and Cudeck (1993) presented a rule of thumb for the values of the RMSEA that indicate reasonable fit. They asserted that values less than 0.08 are evidence of reasonable model fit.

Classification of Structural Equation Fit Indices

As the number of fit indices easily accessible to researchers increased, it soon became apparent that the different fit indices had some important differences. These differences provide the basis for useful classification systems that can aid researchers in choosing appropriate fit indices for their studies.

The first widely used classification scheme for fit indices consisted of two classifications; so called "stand-alone" indexes, and "incremental" indexes (of which there were two types) (Marsh, et al. 1988). The chi-square/df ratio and the RMR are examples of stand alone fit indices. Stand alone indices attempt to assess the fit of the hypothesized model in absolute terms. Incremental indexes, on the other hand, attempt to assess model fit by comparing the hypothesized model to some other "baseline" model. Typically, this baseline model is the "null" model, a model in which all of the observed variables are assumed to be uncorrelated. Incremental fit indices measure the improvement in model fit for the hypothesized model compared to this null model. Bentler and Bonett (1980) pointed out that the incremental fit indices can be used for any two nested models, not just the hypothesized and null models.

Hu and Bentler (1995) further refined and delineated the two types of incremental fit indices into two groups; relative indexes and absolute indices. Relative indices are a slightly broader class of indices than the incremental indexes. With relative indices the hypothesized model may be compared to the null model and to the "just identified" or "saturated" model. The just identified is a model in which there are no degrees of freedom. This model represents a hypothetical best fitting model. The null model and the saturated model provide the endpoints of a continuum along which the hypothesized model's fit can be placed (Maruyama, 1998). Judgements as to the quality of a model's fit are then determined by the proximity of the model's fit statistic to either of these endpoints. Models whose fit index is relatively closer to the fit statistic value of the null model are determined to be poor fitting while models whose fit statistic is close to the value of the saturated model's show good fit. Hu and Bentler (1995) also proposed adjusted indices that would take into account the relative complexity of the hypothesized model compared to the null and saturated models. These fit indices are useful for gauging the parsimony of the hypothesized model. All other things being equal, researchers typically prefer simpler models.

Tanaka (1993) provided a comprehensive and complete classification scheme for structural equation fit indices. Tanaka expanded on the work of Marsh, et al. (1988), and Hu, Bentler & Kano (1992) and proposed six dimensions along which fit indexes could be classified. This scheme accommodated the ground work that was laid by Marsh, et al.(1988), and Hu and Bentler. For example, the absolute and relative fit indexes can be placed along a dimension that ranged from fit indexes that were absolute (i.e. stand alone) to those that are relative (i.e. incremental). Hu and Bentler's adjusted fit indexes could be placed along a simplicity vs. complexity (p. 16) continuum. This represents the degree to which a fit index penalized models in which many parameters were estimated.

The new dimensions for fit indexes that were provided by Tanaka were "population vs. sample" based, "normed vs. non-normed", "estimation method free vs. estimation method specific", and "sample size independent vs. sample size dependent" (Tanaka, 1993, p. 16). Population vs. sample is determined by whether or not the fit index estimates a known population parameter. Normed fit indices are designed to have values that range from 0 to 1. Estimation method free indexes will have values that do not change across estimation methods (i.e. maximum likelihood or generalized least-squares).

Sample size independent fit indexes will have values that do not fluctuate due to differing sample sizes for estimation of the same model (Tanaka, 1993).

Differences Between the Fit Indexes

The classification schemes for the fit indices employed by researchers using structural equation models allow the choice of fit indices that target specific aspects of their models. For example, a model producing the smallest possible residual may be more desirable than the most parsimonious model. These schemes also delineated the ways in which the fit indices differ and have led to a number of research studies investigating how they differ from one another under a variety of circumstances.

Although researchers were having success classifying the fit indexes according to their similarities, research was beginning to show that the fit indexes behaved differently from each other for a number of reasons. Kaplan (1988) looked at the effect of specification error and estimation method on structural equation models and found that different estimation methods would yield biased parameter estimates for different sets of parameters within the same model. This research showed that ML estimation yielded biased parameter estimates across the model when misspecification was present while another estimation method, two-stage least squares, only produced biased parameter estimates around the model's misspecification. Interestingly, in neither case was bias detected in the parameters associated with the observed indicators of latent endogenous variables. The prospect of bias in parameter estimates is insidious. Kaplan (1988) pointed out that parameter bias in the structural model propagated into the measurement model. Although it might be expected this would tend to produce lower values of the goodness-of-fit functions, it was found that the relative chi-square was not able to detect models with severe levels of parameter bias in certain models. Furthermore, it is probably not possible to determine, ahead of time, what models lead to this situation (Kaplan, 1988).

Structural equation techniques are primarily designed for use on data that is continuous and, for some estimation methods, multivariate normally distributed. Techniques have been developed, however, that allow the techniques to be used on categorical data. These techniques are necessary because research has shown that when data sets containing ordinal data are analyzed using product moment correlation coefficients, the estimates of structural parameters are biased (Ethington, 1987).

Green, et al. (1987) showed that the number of scale points employed in Likert scale data had an effect on the chi-square statistic employed in the evaluation of confirmatory factor analyses. They found that as the number of scale points increased, the likelihood that the chi-square statistic would indicate spurious additional factors in data generated from a dataset reflecting a single factor decreased. This finding was contrary to previous research which they cited (e.g. Bernstein & Teng, 1989).

Coenders, Satorra, and Saris (1997) showed the use of different types of correlation coefficients would also affect the accuracy of the point estimates of the model's parameters when ordinal data was used. They compared the Pearson product-moment coefficient, the polyserial coefficient and the conditional polychoric correlation. As the underlying continuous distribution(s) reflected by the ordinal scale(s) became less normal, the polyserial and conditional polychoric coefficients tend to produce less biased point estimates for model parameters.

It has also been shown that as ordinal data exhibits increasing non-normality, the fit indices are affected, although to different degrees. Further complicating matters is the interaction between estimation method and non-normality. Research has shown that the effects of non-normal data on fit indices are tempered in some instances by the estimation method used to fit the structural equation model to the data. For example, the use of weighted least-squares to fit a model will offset somewhat the effects of non-normality in the data for some fit indices such as the chi-square statistic and RMSEA, but not in fit indices such as the NNFI or the CFI (Hutchinson & Olmos, 1998).

Nonnormal data pose problems in structural equation models even if the data are continuous. Curran, West, & Finch (1996) showed that ML estimation showed increasing bias in its chi-square based fit indexes as nonnormality in the data increased. This same bias was not present when ADF was used with appropriate sample sizes. Their study indicated that non-normal data tended to impede the chi-square statistic's ability to detect model misspecification.

Nonnormal data have also been shown to affect the estimation of structural parameters (variances and regression weights) in structural equation models. Nonnormality affects the estimation of the disturbances for latent endogenous variables to a greater extent than it affects the estimation of other model parameters. Furthermore, non-normality tends to produce standard errors for the model parameters that underestimate the variability of the parameter estimates. This makes testing for significant model parameters difficult, even for the parameter estimates that have been shown to not be overly influenced by nonnormality (e.g., factor loadings) (Wang, Fan, & Willson, 1996).

In more complicated structural equation models, researchers may not be interested in individual model parameters. Rather, it may be that the effect of one latent variable on another latent variable that is of substantive interest. The total effect of one latent variable on another is equal to the sum of the direct effect (regression coefficient) between the two variables and any indirect effects. An indirect effect is a path from the first latent variable to the second that passes through at least one other latent variable. An indirect effect is also referred to as a mediated effect. Nonnormality makes the estimation and significance testing of the total effect of one latent variable on another difficult to because of differential influence on the standard errors. The estimates of standard errors for mediated effects show less bias under nonnormality than do the estimates of the standard errors for direct effects (Finch, West, & MacKinnon, 1997).

Research has also shown that the estimation method used to fit the structural equation model to the empirical data set can have an effect on the fit indexes for that model (Ding, Velcier, & Harlow, 1995). It was shown that many but not all of the fit indices studied were significantly affected by estimation method. The results of this study indicated that less bias was present in the estimation of parameters when GLS was employed compared to when ML was used with the model. This finding was supported by Fan, Thompson, & Wang (1999). Their research also showed that estimation method significantly affected the fit indices of structural equation models. This study also expanded upon the results of Ding, Velcier, & Harlow (1995) by showing that misspecified models demonstrated more bias in their fit indexes than did correctly specified models. This is important since research using structural equation models often employs the comparison of competing models. Bollen and Long (1993) state the point well; "it is better to consider several alternative models than to examine only a single model. Often knowledge in an area is not detailed enough to provide a single specification of a model" (p. 7). If two or more models are being compared, at least one will be misspecified.

Marsh (1998) showed that the method in which a researcher deals with missing data can also result in bias in fit indexes based on the chi-square test statistic. Listwise deletion of missing data is often undesirable due to the large sample sizes required to estimate structural equation models. Pairwise deletion allows for the use of all available information in the calculation of the covariances used in model estimation. Unfortunately, a major problem with pairwise deletion is that as the number of deletions increases, the likelihood that the variance covariance matrix will become non positive-definite increases. If the variance-covariance matrix is non positive-definite, it cannot be inverted and a solution cannot be estimated. Even if pairwise deletion does not result in a not positive-definite matrix, the values of the chi-square test statistic (and all fit indices that use the chi-square value) will be biased. This bias does not appear in the parameter estimates themselves.

As the affects of such things as measurement level, estimation method, and missing data are becoming clearer in broad ways, research is shifting to look at how specific fit indexes behave and compare to each other under very controlled conditions. Computer simulation (e.g. Monte Carlo, bootstrap) continues to be the method of choice for this research.

Marsh (1995) has compared the TLI to Bollen's Δ2 fit index and found the TLI to perform better. The Δ2 fit index was shown to be systematically biased as sample size changed. Marsh also contends that the parsimony correction built into Δ2 actually accomplishes the opposite, that it penalizes the simpler model. To further complicate matters, this penalty also varies with sample size. The penalty is larger for small N than for larger N (Marsh, 1995) In Bollen's original derivation of Δ2, the parsimony correction is explained and appeared to be correct (Bollen, 1989, 271-2). Marsh interpreted the same correction in a different way and arrives at the opposite conclusion. It has not been determined which interpretation is correct.

Other fit indexes are also being questioned on philosophical grounds. Rigdon (1996) compared the CFI with the RMSEA citing research showing that incremental fit indexes like the CFI are less stable across estimation methods than absolute fit indexes. Rigdon questions whether the null model is the most appropriate model to compare the hypothesized and saturated models to. He conceded that abandoning this convention would result in a number of very similar fit indexes that are not directly comparable and do not have sampling distributions on which to base an assessment of overall model fit. In the end, Rigdon (1996) agreed with (and cited) Sobel & Bohrnstedt's (1985) recommendation that the CFI should be used in exploratory (model generating) endeavors and that the RMSEA should be used in a confirmatory role (Sobel & Bohrnstedt, 1985).

The difficulties encountered in the above research dealing with adjustments for parsimony do not come as a surprise. Marsh, Balla, & McDonald (1988) suggested that an ideal fit index is one that is independent of sample size, adjusts for parsimony, and adjusts for overall model fit. McDonald and Marsh (1990) expanded this list to include that a fit index that reflects the true population model when it is known. Compared to this standard, all of the currently available adjusted for parsimony fit indexes fail to meet all of these criteria. The fit indexes fail to meet these criteria in different aspects and to different degrees (Williams & Holahan, 1994).

Based on the previously cited research showing that estimation method, scale of measurement, sample size, normality of the data, and portion of the model under analysis can affect fit indices it is reasonable to question, as Tanaka (1993) does, whether in fact all of the different fit indexes calculated for any given structural equation model will tend to agree in their reflection of the quality of model-data fit. Tanaka stated the point well:

"Given the ambiguity that can arise from multiple fit measures, it is essential to develop some unifying set of principles for comparing fit indices" (p. 15).

Chapter 3

Methodology

A request was made to the SEMNET Internet discussion group (SEMNET@BAMA.UA.EDU) via the SEMNET listserver. Researchers were asked to e-mail the portion of computer output that contains the fit indices for structural equation models that they had estimated. Sets of fit indices that were produced by AMOS, EQS, LISREL, or PROC CALIS (SAS) were included in this confirmatory factor analysis. These packages were chosen since they are the most widely used software packages used for structural equation modeling. For consistency, only fit indices that are calculated by all four software packages were included in the analysis.

The common core of fit indices that were included in this study consisted of 9 statistics; chi-square, the relative chi-square, the root mean square residual (RMR), the goodness of fit index (GFI), the adjusted goodness of fit index (AGFI), the parsimony goodness of fit index (PGFI), the Tucker-Lewis index (TLI, also referred to as the non-normed fit index (NNFI)), the incremental fit index (IFI, also referred to as Bollen's $\Delta 2$), and the root mean square error of approximation (RMSEA). These fit indices were used as the data set for a confirmatory factor analysis consisting of one latent factor (goodness-of-fit) with nine observed indicators (the 9 fit indexes). The confirmatory factor analysis model was estimated by AMOS 4.0 (Arbuckle, 1999) on an IBM personal computer running the Windows 95 operating system.

The target sample size for this study was 300 sets of fit indices. The sample was then randomly divided into two groups, each consisting or approximately 150 complete sets of fit indices. The first group was used to estimate the initial model and the second was used to cross-validate the model once appropriate modifications had been made. This sample size is based on the need to estimate eighteen model parameters in a one-factor model. Kline (1998) opined that that a ratio of between 10:1 and 20:1 is appropriate to have between number of subjects and number of estimated free parameters (Kline, 1998, p. 211). The target sample size of 150 in each group meets this ratio. This sample size is also appropriately sized so as to not over-power the chi-square test statistic (Hayduk, 1987).

The initial model estimated in the confirmatory factor analysis was a one factor model in which all nine fit indices are specified to load upon the single latent factor. It is appropriate to specify and estimate a single factor confirmatory factor analysis as a starting point when previous research does not give guidance as to the number of factors that are in the model (Kline, 1998).

The quality of the one factor model was not judged by only the chi-squared discrepancy test. The squared-multiple correlations for each observed variable were computed along with the chi-square discrepancy test. The squared-multiple correlation for each observed variable is the proportion of the variance in that variable that is explained by the factor. A discrepancy in the values of the squared-multiple correlations may indicate that the number of latent factors specified is incorrect (Kline, 1998).

The modification indices were also computed to see other potential model modifications. Modification indices give a conservative estimate of the decrease in the chi-squared test if certain model parameters (correlations) were freed to estimation rather than being constrained to equal zero (Arbuckle & Woethke, 1999). Modification indices are calculated based solely on statistical criteria and not on theoretical considerations. For this reason, it is possible that some, or all, of the covariances that are indicated as capable of improving model fit will not be freed to estimation. As an example, theory would support allowing the errors in the chi-squared test and the relative chi-squared test to correlate. Both of these fit statistics are based directly on the chi-squared statistic. As an aside, this particular correlation was left out of the initial model in the interests of parsimony. On the other hand, theory probably would not support error in the chi-squared statistic directly correlating with the error in the root mean-squared residual since they are based on different functions.

Once the model had been modified based on theoretical and appropriate statistical concerns, the specification for this revised model was applied to the second half of the collected data. As with the initial model, the fit of the revised model was assessed by the chi-square fit statistic and by the squared-multiple correlations. Various fit indices are also reported to help facilitate independent future research.

Chapter 4

Results

Participants in the SEMNET discussion group returned 331 complete sets of fit indices for expressed use in this study. Four cases were multivariate outliers as determined by significant Mahalanobis distances. These four cases were excluded from the sample. This sample was then split randomly using SPSS 9.0 into 2 groups, group 1 for use in initial model building and group 2 for cross-validation.

The first group consisted of 159 sets of fit indices. All nine fit indices displayed some degree of non-normality. However, the chi-squared and root-mean-residual were more atypical than the other 7 fit indices as shown in Table 1.

Fit Index	Skewness	Kurtosis
Chi2	2.521	6.201
RelChi2	1.528	1.441
RMSEA	.794	1.218
RMR	3.735	13.173
GFI	-1.135	.502
AGFI	-.644	-.149
PGFI	-.313	-.645
TLI	-.590	.023
IFI	-1.170	1.508

Table 1: Skewness and Kurtosis Statistics for Group 1.

Based on the relatively greater values of skewness and kurtosis compared to the other seven fit indices, these two variables were excluded from the analysis. This decision appears reasonable in light of Kline's characterization of skew values greater than 3 as "extreme" (Kline, 1998, p.82).

The first step in the model generation process consisted of fitting a one-factor model to the seven fit indices in the data set. The path diagram for this model is presented in Figure 1:

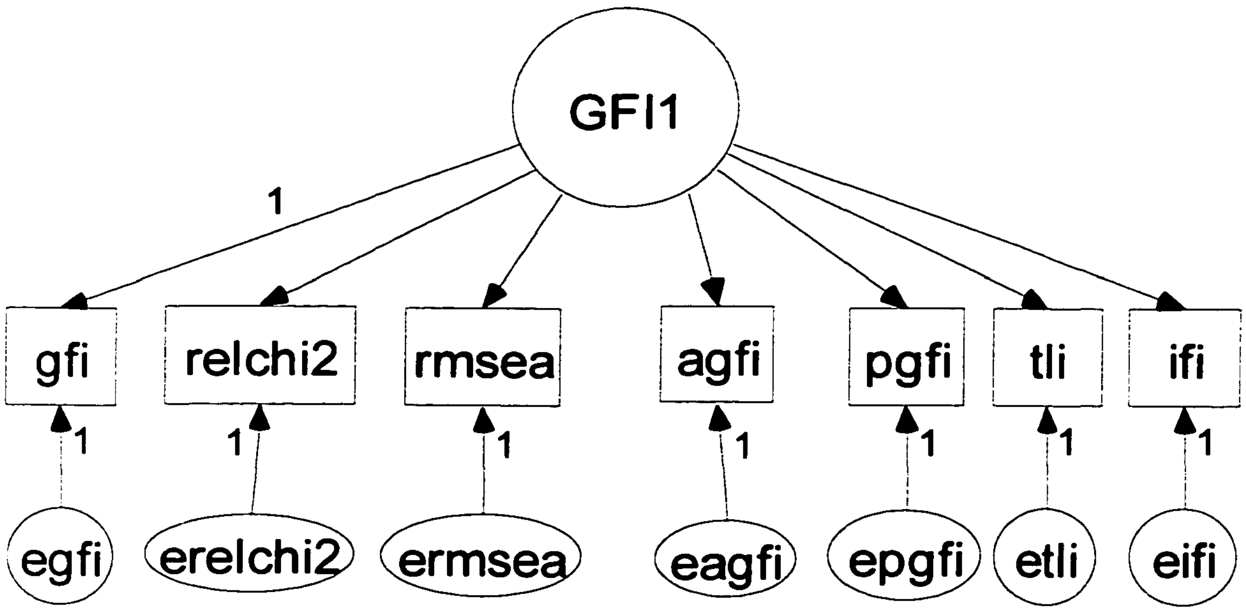


Figure 1: The One Factor Model.

The one factor model does not fit the data in group 1 well. The obtained chi-squared value of 801 (14 df) is significant at the traditional .05 level. The other fit indices also indicate poor fit (GFI = .577, RMSEA = .412, incremental fit indices < 0.65). The squared multiple correlations for observed variables are presented in Table 2.

Observed Variable	Squared Multiple Correlation
IFI	0.593
TLI	0.733
PGFI	0.071
AGFI	0.892
RMSEA	0.395
RelChi2	0.154
GFI	0.855

Table 2: Squared Multiple Correlations Resulting from the One Factor Model.

The squared multiple correlations show that there may be promise in the factorial representation of the model. Four observed indicators have over half of their variance accounted for by the single factor. A further indicator as to the potential for the model is the fact that the factor loading of each observed variable, with the exception of GFI, is statistically significant at the 0.05 level. The factor loading of GFI on the latent factor cannot be tested for significance because it was fixed at 1.0 to satisfy the identifiability of the model. The fit of the one factor model, however, is not good. The next step involved modifying the model, using the same data, in order to try to get a better fitting model to use in cross validation.

Statistical guidance in modifying structural equation models can typically be found in two places. If there are non-significant paths in the original model, those paths may be eliminated from the model. Additionally, software packages calculate modification indices that give an indication of how much the model fit will improve if specified paths are added to the model. Each of these methods was used in modifying the one factor model and the results are summarized below. Obviously, one must be careful in deleting or adding paths from the *a priori* model. The conceptual ramifications cannot be overshadowed by the statistical criteria of a path significance, or the value of a modification index.

The one-factor model did not have any non-significant paths. No paths will be deleted in the first round of model modification. Model modification will consist of adding a second factor to the model. This addition is justified on two levels. First, the variance of three of the observed indicators is not well accounted for by the single factor. Second, the modification indices indicate that model fit can improve significantly if certain measurement errors are allowed to correlate with each other. Correlated error terms are often indicators of a missing factor in the equations of the observed variables whose measurement errors are indicated to covary (Schumacker & Lomax 1996, p.86). This missing factor may either be a factor in the model that has no direct effect on the two observed variables or it may be an unmodeled factor. Since there is only a single factor in the first model, an unmodeled second factor is clearly indicated.

The indication for the addition of a second factor complicates the model when considered in conjunction with the fact that there are no non-significant factor loadings in the one factor model. This means that the second factor will "overlay" the first factor. Stated another way, some observed variables will be directly affected by both factors. This will result in a model that doesn't possess the characteristic of unidimensional measurement. This fact will manifest itself in the need to impose some additional constraints on the model in order to achieve identifiability. The path diagram for the two factor model is presented in Figure 2:

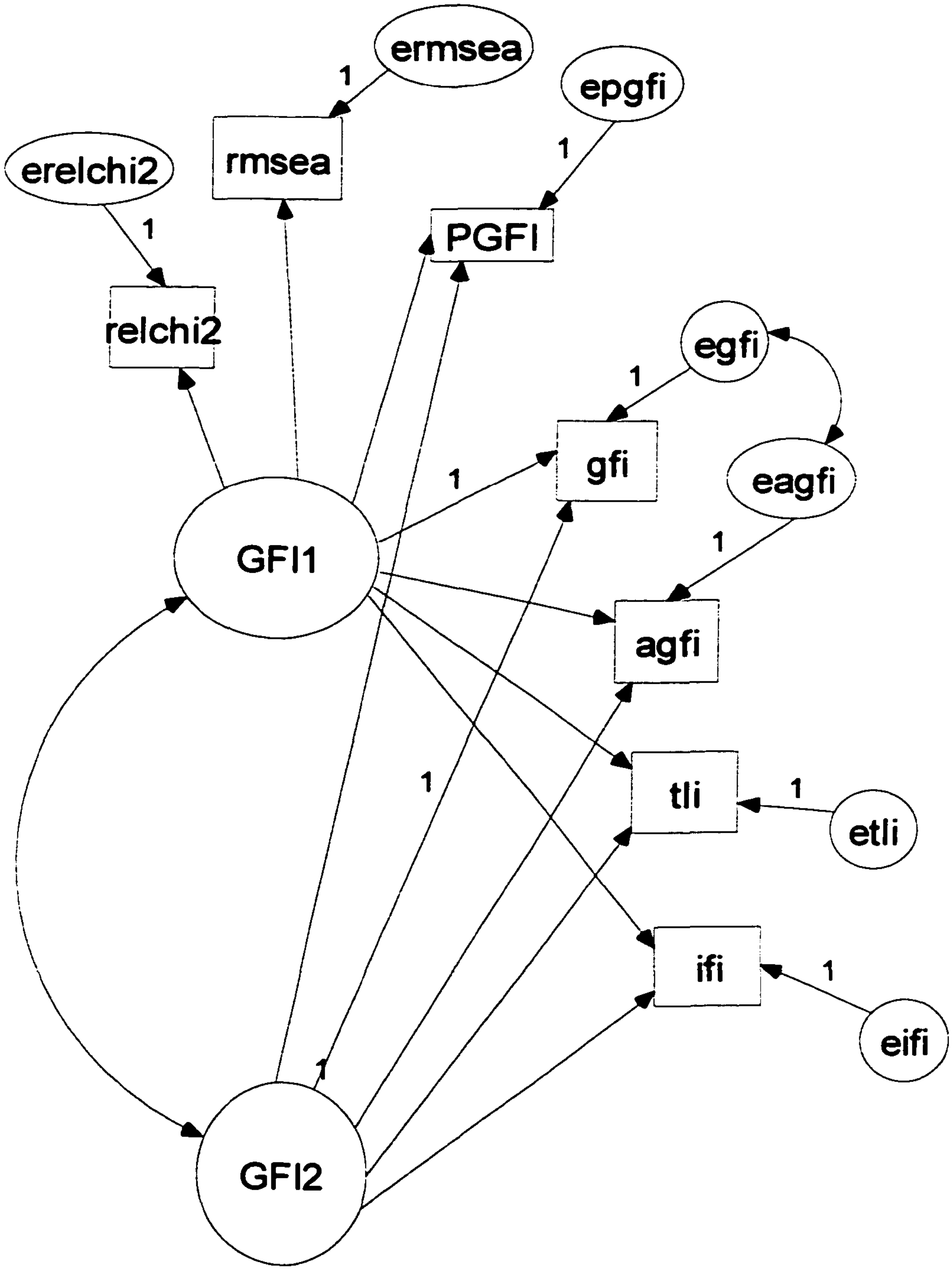


Figure 2: The Two Factor Model.

Two additional model modifications were made to the one-factor model. The variance of the second factor was set to 1.0. This was required in addition to fixing the factor loading of the observed variable GFI to achieve identifiability of the model. Modification indices indicated that the errors between the observed variables GFI and

AGFI should covary even after the addition of the second factor. The covariance between the measurement errors of GFI and AGFI is defensible since the value of GFI is used directly in the computation of AGFI.

As anticipated, the fit of the two factor model was significantly better than that of the one factor model. The chi-square statistic was significantly reduced to 44.4 with 8 degrees of freedom. The other fit indices also demonstrated improved fit (GFI = .933, RMSEA = .17, Incremental fit indices > .86). The values of these fit indices do not quite demonstrate good fit. They are much better than those of the original model, and are quite close to the "traditionally accepted" empirical cut-off values.

As with the one-factor model, all of the factor loadings of the observed variables were statistically significant. Squared-multiple correlations for observed variables are presented in Table 3.

Observed Variable	Squared Multiple Correlaton
IFI	0.474
TLI	0.922
PGFI	0.520
AGFI	0.486
RMSEA	0.867
RelChi2	0.524
GFI	0.554

Table 3: Squared Multiple Correlations from the Two Factor Model.

As can be seen in Table 3, four of the squared multiple correlations increased as a result of including the additional factor. Although three squared multiple correlations did decrease in value, no manifest variable has a less than 47% of its variance accounted for by the model as opposed to 7% in the one factor model.

No further model modification is appropriate. All paths are statistically significant and should be retained. Modification indices do not indicate that there are more paths to add, at least none that would be worth the loss of an additional degree of freedom. The two factor model with two correlated errors (as shown in Figure 2) will be the model used for cross-validation with group 2 of the data set.

Group 2 consisted of 168 complete sets of fit indices. As with group 1, the variables were not normally distributed. The degrees of non-normality in the manifest variables were comparable to those seen in group 1 with the exception of the relative chi-squared variable. Relative Chi-squared exhibited much more skew in group 2 (skewness >4) than it did in group 1. This variable was not excluded from the analysis since this sample is being used to confirm the model generated with group 1. Fit indices for both groups are presented in Table 4.

Group	Relative Chi-Squared	GFI	TLI/IFI	RMSEA
Group 1	5.55	.93	.89/.96	.17
Group 2	8.08	.92	.89/.96	.21

Table 4: Fit Indices for Groups 1 and 2 for the Two Factor Model.

Table 5 presents the squared multiple correlations for observed variables in both groups. As in group 1, all paths were statistically significant (<.05) when the second group data were fit to the two factor model.

Observed Variable	Squared Multiple Correlations Group 1 / Group 2
IFI	0.474 / 0.922
TLI	0.922 / 0.994
PGFI	0.520 / 0.455
AGFI	0.486 / 0.742
RMSEA	0.867 / 0.589
RelChi2	0.524 / 0.216
GFI	0.554 / 0.972

Table 5: Squared Multiple Correlations from Groups 1 and 2 for the Two Factor Model.

Chapter 5

Discussion

Assessing model-data fit in structural equation models is at once both a rapidly developing, as well as a tricky area of research. Perhaps the vigor with which this area is being explored is most evident by the terms researchers use to describe it. Tabachnick & Fidell describe this are of research as "lively" (1996, p. 748), while Bollen stated that this area is surrounded by "the most heated controversies" (Bollen & Long, 1993, p.2). This study does not provide the information necessary to resolve these controversies. It does, however, share some insight regarding a fundamental concept that, to this point, had been taken for granted. This clarification does move the area forward in two ways. First, pedagogical practices are addressed and clarified and secondly, a focused direction for future research is specified.

The results of this study show that the conventional wisdom that all the fit indices tend to say the same thing is an oversimplification. It is well known that the fit indices will be highly correlated with each other (Marsh, et al., 1988). This study showed that this inter correlation did not manifest itself in a one factor model. More specifically, the fit indices did not load on a single factor that could be thought of as the construct "good fit". Rather than good fit being thought of as a single unified construct with the fit indices being the observed indicators, multiple latent factors, each representing an aspect of good fit, affected the fit indices. This coincided nicely with Tanaka's multi-dimensional classification scheme for the fit indices (Tanaka, 1993).

The null hypothesis in structural equation modeling is represented by

$$\Sigma = \Sigma(\theta)$$

where Σ is the population covariance matrix and $\Sigma(\theta)$ is the covariance matrix formed by the sample moments and the vector of free model parameters, θ. In an attempt to test this model, the sample covariance matrix was compared to an implied covariance matrix based on the model parameters and the discrepancy was noted. Each of the fit indices in this study attempted to capture this discrepancy. Factor 1, which affected each fit index, could be thought of as the factor representing the discrepancy between Σ and $\Sigma(\theta)$. This factor of absolute model discrepancy is common throughout statistical hypothesis testing. At their hearts, even the most basic statistical tests are trying to determine whether there is 'too big' a discrepancy between the data at hand and the specified model. Structural equations do tend to have more complicated models than, say, t-tests do. When we specify, *a priori*, that a particular statistic is equal to some value (or another statistic) we have a model that can be tested. This process of testing makes use of the discrepancy factor, no matter how simplistic the model.

The second factor was less clear in its meaning. Factor 2 did not directly impact either the relative chi-square, or the RMSEA fit statistic. These two fit indices were the most straightforward measures of discrepancy used in this study. Neither of these two fit indices compared the results of the estimated model to anything related to another model. The five fit indices that factor 2 did affect do compare results from the estimated model to some other model. Factor 2 could be thought of as comparative discrepancy.

At first consideration, it may not be clear that the GFI is a fit index that compares the results from two models. In fact the GFI is typically considered to be analogous to the multiple R-square used so predominately in multiple regression (Tanaka & Huba, 1989). Thought of in this light, GFI appears to be an absolute index of fit and contrary to the notion that factor 2 in this study represents comparative discrepancy with respect to some additional model.

Joreskog & Sorbom (1984) described their GFI in slightly different terms. They explained the GFI as a proportion. The numerator of this proportion is the value of the minimum fit function for the model under consideration. The denominator is the value of the fit function before any model has been fit to the data (1996, p 29). Looking at the GFI in this manner, we see that we are comparing models. The GFI compares the results of the specified model to the results of no model, a null model in the absolute truest sense.

The AGFI and the PGFI both employ a model comparison strategy via the GFI. This model comparison is implicit in the fact that both the AGFI and the PGFI have at their hearts the GFI. Therefore, the AGFI and the PGFI were also indicators of the comparative discrepancy factor.

The formulas for the TLI and the IFI are clearer in their use of model comparison. Due to this, it was easier to see how they are manifest variables of the comparative discrepancy factor. In each of these fit indices, chi-squared discrepancy statistics are calculated based on a null model. The null model represents a very poorly fitting model and provides a basis for evaluating the fit of the hypothesized model. The TLI and the IFI make use of the two discrepancy statistics and the attendant degrees of freedom for the two models in the calculation of their values.

A natural question was raised at this point. For the five fit indices that were affected by both the factor of absolute discrepancy and the factor of comparative discrepancy, do the factors affect the fit indices differentially? Although this was not one of the research questions that this study was initially designed to answer this question was able to be addressed in a reasonable manner.

In an attempt to gain insight into the nature of any differential effects the 2 factors have on the fit indices (if there were any), the two factor model was re-estimated using data from both the initial and confirmatory data sets. Both data sets were combined so that the factor loadings would be based on as much data as was available and thus be as robust as possible. The factor loadings for the five fit indices that were affected by both latent factors are presented in Table 6.

Fit Index	Factor 1 Loading (Absolute Discrepancy)	Factor 2 Loading (Comparative Discrepancy)
GFI	1.00*	1.00*
AGFI	.732	.772
PGFI	-2.001	-1.937
IFI	.731	.765
TLI	.506	.578

*The factor loadings of GFI were fixed to 1.0 for identifiability.

Table 6: Factor Loadings for Fit Indices Affected by Both Latent Factors

The factor loadings for the GFI index were fixed at 1.0 to identify the model. The most intriguing result were the factor loadings of PGFI. The factor loadings of PGFI were of twice the magnitude of the loadings of the other, non-fixed fit indices. The sign of the factor loadings of PGFI were also of the opposite sign compared to the others.

It is well known that in structural equation modeling overall model-data fit can be improved by simply freeing more model parameters for estimation. Obviously, the more model parameters that are estimated from the data set, the better the model will fit the data set. Fit indices such as the AGFI and PGFI incorporate a correction to the standalone GFI fit statistic based on the number of model parameters estimated. More specifically, a downward adjustment is made to the GFI based on the number of model parameters that are freed for estimation. The magnitude of this downward adjustment increases as the number of free parameters increases. The results of this study showed that the correction incorporated in the PGFI is much more extreme than the correction incorporated in the AGFI. This is important in light of the findings of Marsh, Balla, & McDonald that the AGFI already overcorrects for adding additional free parameters to a particular target model (1988, p. 398). Another possibility is that the data set consists of many sets of fit indices that are based on models that were close to being saturated.

This result related to the PGFI does not come as a complete surprise. Tabachnik and Fidell (1996) pointed out that, unless there are many more covariances than there are free parameters, the PGFI will have a value much lower than the other fit indices (1996, p.751). Williams and Holahan (1994) found that the PGFI was one of the two best fit indices for differentiating between nested models that progress from a target model toward a saturated model. This result clearly showed that the PGFI is very strongly influenced by the addition of free parameters to the model.

The other interesting result from looking at the loadings on the absolute discrepancy and comparative discrepancy factors was that the TLI had smaller loadings than the IFI or the AGFI. This may mean that the TLI is more stable in the face of discrepancy between the model and the data. Additionally, the TLI had the highest squared multiple correlations in both the exploratory and confirmatory data sets. These two results, taken together, would seem to indicate that the Tucker-Lewis Index is the best fit index for use in assessing a structural equation model in terms of both absolute and relative discrepancy.

Limitations of the Study

As with any study, there are certain qualifiers that need to be explicitly stated. The major limitation of this study relates to the data set. While the sample size was appropriate for the testing of the structural equation models, it is made up of fit indices that represent a very small portion of all the possible structural equation models that are possible. Another limitation of this study is the number of fit indices used in the data set. More specifically, having to reduce the number of observed variables from nine to seven is disappointing. The exclusion of the two variables for violation of distributional assumptions does increase confidence in the results that this study produced. Finally, the results of this study should be carefully scrutinized due to the fact the final structural equation model exhibits somewhat questionable model-data fit.

Suggestions for Further Research

While all of the model's parameters were statistically significant, and much of the observed variables' variance was explained, the fit indices for the model fell slightly outside traditionally accepted values. This fact raised an interesting question that should be investigated: What does it mean to have a model account for a lot of variance, yet not fit the data well?

The results of this study need to be cross-validated with many more sets of fit indices. The model should also be tested against data sets that are much larger than the one used in this study. Using larger data sets would allow three things. First, it could be investigated whether there are more factors that affect the fit indices. Secondly, tests for a higher order factor structure could be conducted. It is possible that there is a higher order factor that affects factors representing good fit including the two factors brought out in this study. Finally, using larger data sets would allow for the model to be fit using asymptotic-distribution free estimation. This would allow the use of all nine fit indices originally intended for use in this study, since the distributional assumptions are relaxed from what they are in maximum likelihood estimation. This would also allow for additional fit indices to be included in the model, as mentioned above, without having to worry about their distributional properties.

The relationship between the GFI based fit indices should be explored further. The results of this study showed two interesting results related to the PGFI that are not totally clear. First the sign and magnitude of the PGFI factor loadings are not consistent with those of the other fit indices that are affected by both absolute discrepancy and comparative discrepancy. Perhaps more interestingly, the modification indices did not indicate a covariance between the measurement error of the PGFI as was present between the GFI and the AGFI. If this correlated error is indicative of an unmodeled factor, it appears to be a factor that does not impact the PGFI to the same extent that it impacts the GFI and the AGFI. This relationship between the three indices needs to be clarified.

Additional fit indices should be included in future research to determine which factors they load on and what the magnitudes of those loadings are. The seven fit indices ultimately used in this study are just a small portion of the over thirty fit indices that have been developed for use in evaluating model fit in structural equation modeling. It is important to determine how additional fit indices relate to the factors of absolute discrepancy and comparative discrepancy.

APPENDIX A

The covariance matrix obtained from the fit indices in group 1 is:

	rmsea	relchi2	PGFI	ifi	tli	agfi	gfi
rmsea	0.004						
relchi2	0.246	27.918					
PGFI	-0.002	-0.222	0.024				
ifi	-0.002	-0.098	-0.004	0.009			
tli	-0.004	-0.235	-0.002	0.005	0.008		
agfi	-0.004	-0.120	-0.003	0.005	0.008	0.016	
gfi	-0.001	-0.007	-0.006	0.004	0.005	0.010	0.008

The covariance matrix obtained from the fit indices in group 2 is:

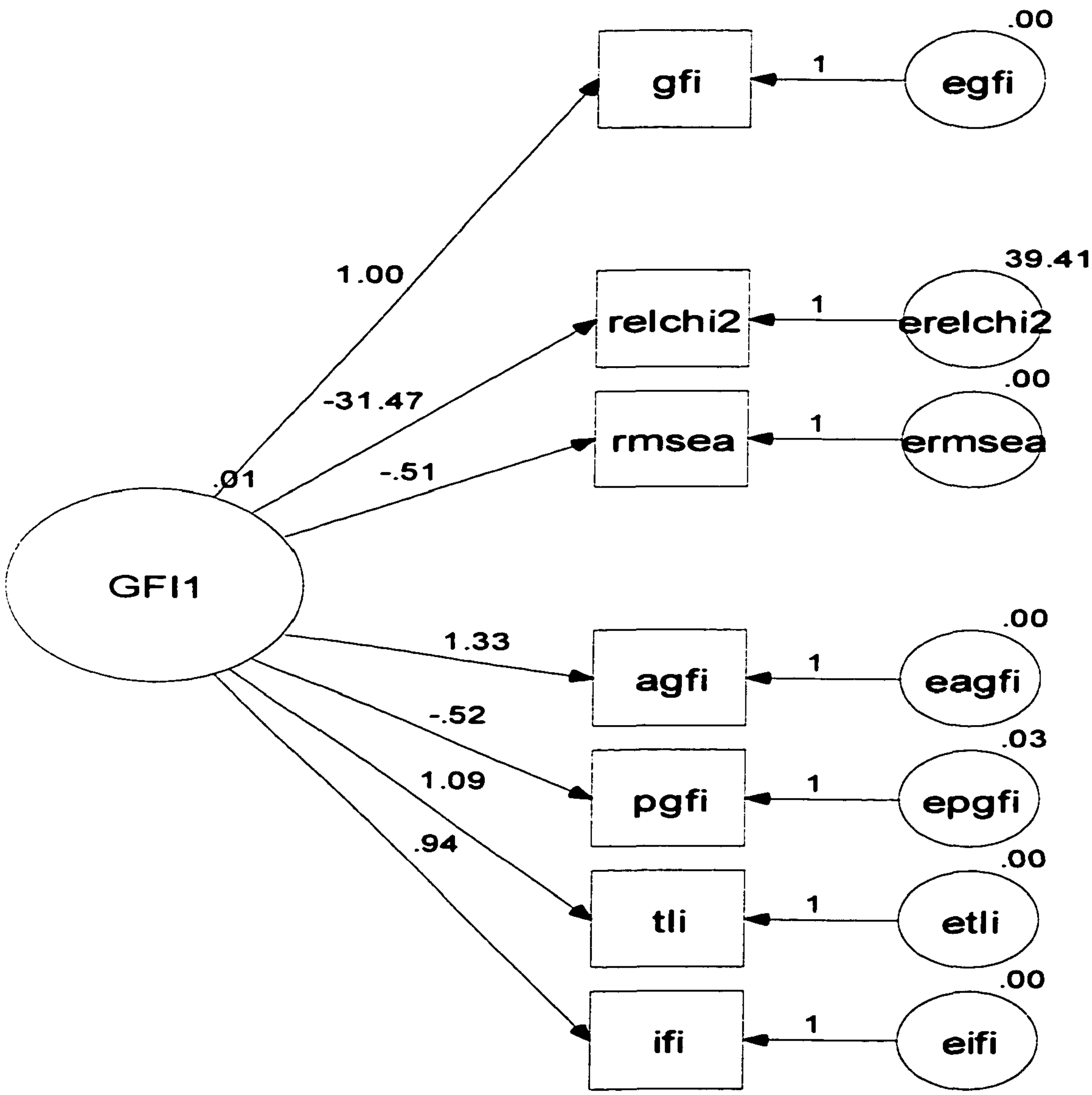
	rmsea	relchi2	PGFI	ifi	tli	agfi	gfi
rmsea	0.003						
relchi2	0.222	42.373					
PGFI	0.000	-0.222	0.030				
ifi	-0.003	-0.091	-0.005	0.005			
tli	-0.004	-0.201	-0.003	0.006	0.008		
agfi	-0.003	-0.108	-0.006	0.005	0.006	0.007	
gfi	-0.001	0.010	-0.007	0.004	0.004	0.005	0.004

The covariance matrix from the complete data set is:

	rmsea	relchi2	PGFI	ifi	tli	agfi	gfi
rmsea	0.005						
relchi2	0.338	46.785					
PGFI	-0.002	-0.276	0.027				
ifi	-0.004	-0.269	-0.004	0.010			
tli	-0.006	-0.385	-0.002	0.008	0.011		
agfi	-0.005	-0.266	-0.004	0.007	0.010	0.014	
gfi	-0.003	-0.132	-0.006	0.006	0.007	0.010	0.008

APPENDIX B

The output path diagram from the one factor model is:



The values of the path coefficients for the one factor model and their test statistics are:

Regression Weights

		Estimate	S.E.	C.R.	P
gfi	<--GFI1	1.000			
relchi2	<--GFI1	-31.469	4.218	-7.460	0.000
rmsea	<--GFI1	-0.510	0.038	-13.558	0.000
agfi	<--GFI1	1.330	0.043	30.942	0.000
pgfi	<--GFI1	-0.516	0.106	-4.872	0.000
tli	<--GFI1	1.095	0.046	23.886	0.000
ifi	<--GFI1	0.940	0.049	18.994	0.000

The values of the modification indices from the one factor model are:

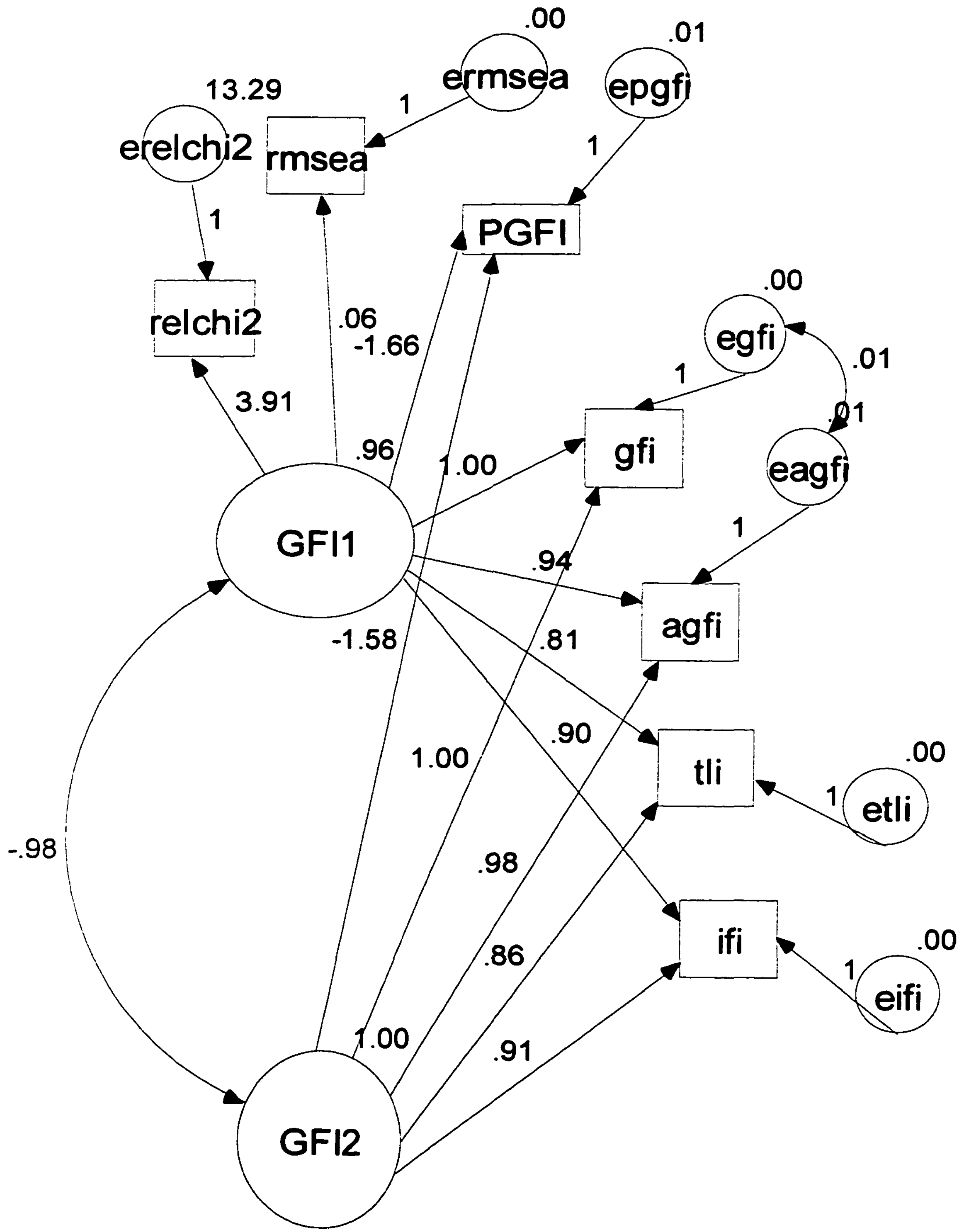
Covariances:			M.I.	Par Change
etli	<-->	eifi	106.034	0.002
eagfi	<-->	eifi	74.525	-0.002
ermsea	<-->	etli	115.551	-0.002
ermsea	<-->	epgfi	55.014	-0.004
erelchi2	<-->	etli	67.386	-0.172
erelchi2	<-->	epgfi	50.287	-0.391
erelchi2	<-->	ermsea	148.984	0.231
egfi	<-->	epgfi	100.838	-0.004
egfi	<-->	eagfi	64.387	0.001
egfi	<-->	ermsea	129.807	0.001
egfi	<-->	erelchi2	74.413	0.124

Regression Weights:			M.I.	Par Change
tli	<--	rmsea	67.983	-0.396
tli	<--	relchi2	56.562	-0.004
rmsea	<--	pgfi	50.930	-0.130
rmsea	<--	relchi2	124.912	0.005
relchi2	<--	rmsea	87.225	46.739
gfi	<--	pgfi	93.486	-0.134
gfi	<--	rmsea	77.017	0.290
gfi	<--	relchi2	62.604	0.003

The fit indices resulting from the one factor model are:

Fit Measure	Default model	Saturated	Independence
Discrepancy	801.414	0.000	2166.144
Degrees of freedom	14	0	21
P	0.000		0.000
Number of parameters	14	28	7
Discrepancy / df	57.244		103.150
RMR	0.091	0.000	0.136
GFI	0.577	1.000	0.350
Adjusted GFI	0.154		0.134
Parsimony-adjusted GFI	0.288		0.263
Normed fit index	0.630	1.000	0.000
Relative fit index	0.445		0.000
Incremental fit index	0.634	1.000	0.000
Tucker-Lewis index	0.449		0.000
Comparative fit index	0.633	1.000	0.000
Parsimony ratio	0.667	0.000	1.000
Parsimony-adjusted NFI	0.420	0.000	0.000
Parsimony-adjusted CFI	0.422	0.000	0.000
Noncentrality parameter estimate	787.414	0.000	2145.144
NCP lower bound	698.350	0.000	1996.021
NCP upper bound	883.873	0.000	2301.602
FMIN	2.414	0.000	6.525
F0	2.372	0.000	6.461
F0 lower bound	2.103	0.000	6.012
F0 upper bound	2.662	0.000	6.933
RMSEA	0.412		0.555
RMSEA lower bound	0.388		0.535
RMSEA upper bound	0.436		0.575
P for test of close fit	0.000		0.000
Akaike information criterion (AIC)	829.414	56.000	2180.144
Browne-Cudeck criterion	830.105	57.383	2180.489
Bayes information criterion	909.971	217.113	2220.422
Consistent AIC	896.728	190.628	2213.801
Expected cross validation index	2.498	0.169	6.567
ECVI lower bound	2.230	0.169	6.118
ECVI upper bound	2.789	0.169	7.038
MECVI	2.500	0.173	6.568

The output path diagram from the two factor model is:



The path coefficients resulting from the two factor model estimated with group 1 data are:

Regression Weights

			Estimate	S.E.	C.R.	P
gfi	<--	GFI2	1.000			
agfi	<--	GFI2	0.981	0.066	14.939	0.000
tli	<--	GFI2	0.860	0.094	9.149	0.000
ifi	<--	GFI2	0.909	0.122	7.429	0.000
gfi	<--	GFI1	1.000			
agfi	<--	GFI1	0.937	0.068	13.862	0.000
tli	<--	GFI1	0.808	0.096	8.414	0.000
PGFI	<--	GFI1	-1.660	0.212	-7.846	0.000
ifi	<--	GFI1	0.896	0.126	7.134	0.000
rmsea	<--	GFI1	0.062	0.004	14.633	0.000
PGFI	<--	GFI2	-1.582	0.206	-7.687	0.000
relchi2	<--	GFI1	3.909	0.384	10.187	0.000

The covariances of the two factor model fit two data from group 1 are:

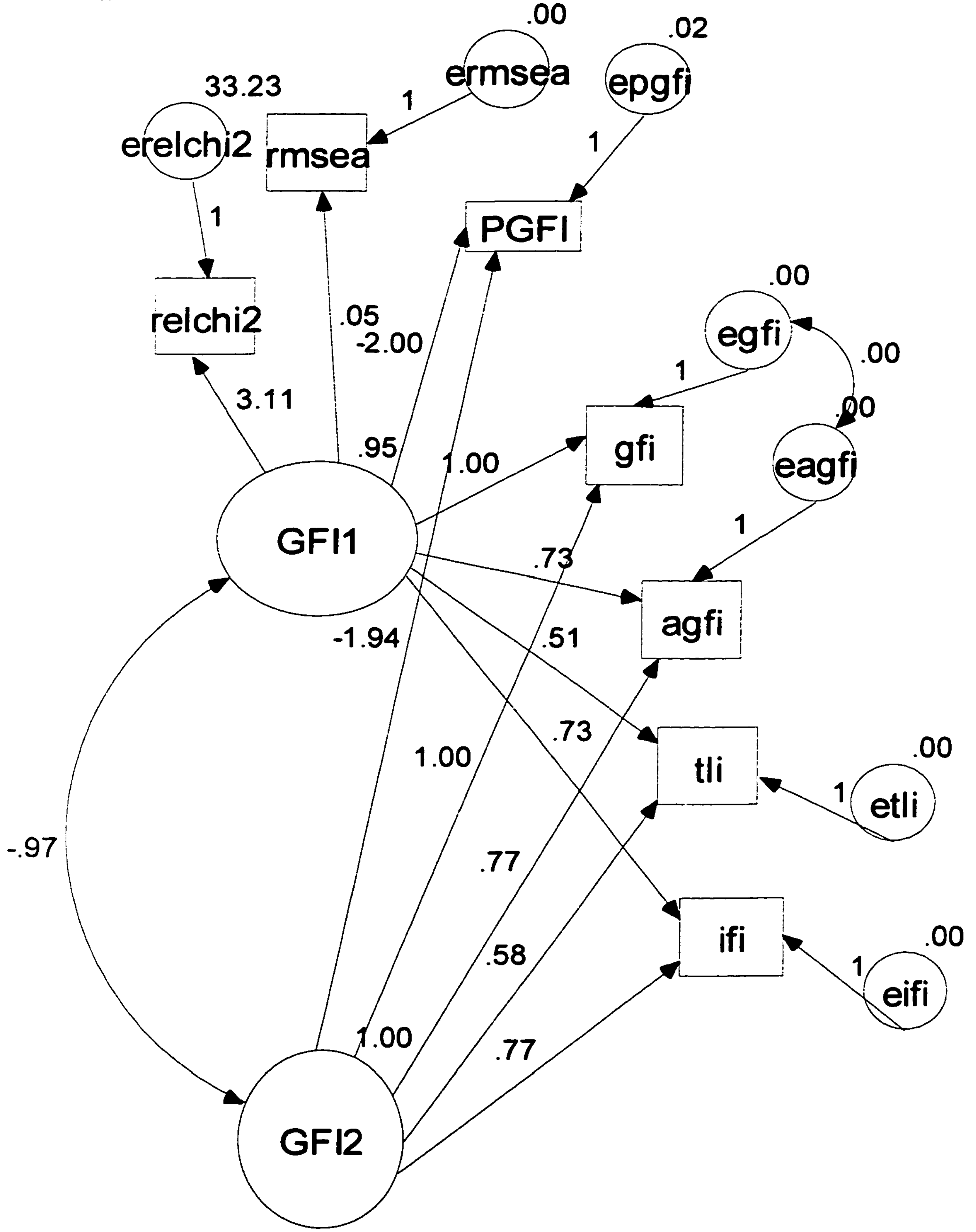
Covariances

			Estimate	S.E.	C.R.	P
GFI1	<-->	GFI2	-0.976	0.008	-130.110	0.000
egfi	<-->	eagfi	0.005	0.001	7.225	0.000

The fit indices for the two factor model fit to the data from group 1 are:

Fit Measure	Default model	Saturated	Independence
Discrepancy	44.353	0.000	866.128
Degrees of freedom	8	0	21
P	0.000		0.000
Number of parameters	20	28	7
Discrepancy / df	5.544		41.244
RMR	0.028	0.000	0.082
GFI	0.933	1.000	0.438
Adjusted GFI	0.766		0.250
Parsimony-adjusted GFI	0.267		0.328
Normed fit index	0.949	1.000	0.000
Relative fit index	0.866		0.000
Incremental fit index	0.958	1.000	0.000
Tucker-Lewis index	0.887		0.000
Comparative fit index	0.957	1.000	0.000
Parsimony ratio	0.381	0.000	1.000
Parsimony-adjusted NFI	0.361	0.000	0.000
Parsimony-adjusted CFI	0.365	0.000	0.000
Noncentrality parameter estimate	36.353	0.000	845.128
NCP lower bound	19.087	0.000	752.558
NCP upper bound	61.128	0.000	945.094
FMIN	0.281	0.000	5.482
F0	0.230	0.000	5.349
F0 lower bound	0.121	0.000	4.763
F0 upper bound	0.387	0.000	5.982
RMSEA	0.170		0.505
RMSEA lower bound	0.123		0.476
RMSEA upper bound	0.220		0.534
P for test of close fit	0.000		0.000
Akaike information criterion (AIC)	84.353	56.000	880.128
Browne-Cudeck criterion	86.487	58.987	880.875
Bayes information criterion	184.650	196.415	915.232
Consistent AIC	165.731	169.929	908.610
Expected cross validation index	0.534	0.354	5.570
ECVI lower bound	0.425	0.354	4.985
ECVI upper bound	0.691	0.354	6.203
MECVI	0.547	0.373	5.575

The output path diagram resulting from fitting the two factor model to the data in group 2 is:



The path coefficients resulting from fitting the two factor model to the data in group 2 are:

Regression Weights

			Estimate	S.E.	C.R.	P
gfi	<--	GFI2	1.000			
agfi	<--	GFI2	0.772	0.056	13.870	0.000
tli	<--	GFI2	0.578	0.070	8.229	0.000
ifi	<--	GFI2	0.765	0.057	13.474	0.000
gfi	<--	GFI1	1.000			
agfi	<--	GFI1	0.732	0.058	12.725	0.000
tli	<--	GFI1	0.506	0.073	6.979	0.000
PGFI	<--	GFI1	-2.001	0.203	-9.840	0.000
ifi	<--	GFI1	0.731	0.059	12.423	0.000
rmsea	<--	GFI1	0.046	0.004	11.674	0.000
PGFI	<--	GFI2	-1.937	0.197	-9.850	0.000
relchi2	<--	GFI1	3.106	0.481	6.456	0.000

The covariances resulting from fitting the two factor model to the data in group 2 are:

Covariances

			Estimate	S.E.	C.R.	P
GFI1	<-->	GFI2	-0.972	0.006	-165.743	0.000
egfi	<-->	eagfi	0.001	0.000	5.007	0.000

The fit indices resulting from fitting the two factor model to the data in group 2 are:

Fit Measure	Default model	Saturated	Independence
Discrepancy	64.660	0.000	1335.299
Degrees of freedom	8	0	21
P	0.000		0.000
Number of parameters	20	28	7
Discrepancy / df	8.082		63.586
RMR	0.045	0.000	0.075
GFI	0.916	1.000	0.341
Adjusted GFI	0.706		0.121
Parsimony-adjusted GFI	0.262		0.255
Normed fit index	0.952	1.000	0.000
Relative fit index	0.873		0.000
Incremental fit index	0.957	1.000	0.000
Tucker-Lewis index	0.887		0.000
Comparative fit index	0.957	1.000	0.000
Parsimony ratio	0.381	0.000	1.000
Parsimony-adjusted NFI	0.363	0.000	0.000
Parsimony-adjusted CFI	0.365	0.000	0.000
Noncentrality parameter estimate	56.660	0.000	1314.299
NCP lower bound	34.674	0.000	1198.207
NCP upper bound	86.122	0.000	1437.766
FMIN	0.387	0.000	7.996
F0	0.339	0.000	7.870
F0 lower bound	0.208	0.000	7.175
F0 upper bound	0.516	0.000	8.609
RMSEA	0.206		0.612
RMSEA lower bound	0.161		0.585
RMSEA upper bound	0.254		0.640
P for test of close fit	0.000		0.000
Akaike information criterion (AIC)	104.660	56.000	1349.299
Browne-Cudeck criterion	106.672	58.818	1350.004
Bayes information criterion	206.057	197.956	1384.788
Consistent AIC	187.139	171.471	1378.167
Expected cross validation index	0.627	0.335	8.080
ECVI lower bound	0.495	0.335	7.384
ECVI upper bound	0.803	0.335	8.819
MECVI	0.639	0.352	8.084

References

Arbuckle, J. (1997). Amos Users' Guide Version 3.6. Chicago: SmallWaters.

Arbuckle, J. & Wothke, W. (1999). Amos Users' Guide Version 4.0. Chicago: SmallWaters.

Bandalos, D. (1997). Assessing sources of error in structural equation models: The effect of sample size, reliability, and model misspecification. Structural Equation Modeling, 4, 177-192.

Bentler, P. & Bonnet, D. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88, 588-606.

Bernstein, I. & Teng, G. (1989). Factoring items and factoring scales are different: Spurious evidence for multidimensionality due to item categorization. Psychological Bulletin, 103. 467-477.

Bollen, K. (1989). Structural Equation Modeling with Latent Variables. New York: John Wiley & Sons.

Bollen, K. & Long, J. (1993). Testing Structural Equation Models. California: Sage Publications.

Browne, M. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. British Journal of Mathematical and Statistical Psychology, 37, 62-83.

Brown, M. & Cudeck, R. (1993). Alternative ways of assessing model fit. In Bollen, K. & Long, J. (Eds.) Testing structural equation models. Newbury Park: Sage. 136-162.

Byrne, B.(1989). A Primer of LISREL: Basic Applications and Programming for Confirmatory Factor Analytic Models. New York: Springer-Verlag.

Chou, C., Bentler, P., & Satorra, A. (1991). Scaled test statistics and robust standard errors for nonnormal data in covariance structure analysis: A monte carlo study. British Journal of Mathematical and Statistical Psychology, 44, 347-357.

Cochran, W. (1952). The χ^2 test of goodness of fit. Annals of Mathematical Statistics, 23, 315-345.

Coenders, G., Satorra, A. & Saris, W. (1997). Alternative approaches to structural modeling of ordinal data: A monte carlo study. Structural Equation Modeling, 4, 261-282.

Curran,P., West, S.,& Finch, J. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. Psychological Methods, 1, 16-29.

Ding, L., Velcier, W., & Harlow, L. (1995). Effects of estimation methods, number of indicators per factor, and improper solutions on structural equation modeling fit indices. Structural Equation Modeling, 2, 119-144.

Ethington, C. (1987). The robustness of LISREL estimates in structural equation models. Journal of Experimental Education, 55. 80-88.

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. Structural Equation Modeling, 6, 56-83.

Finch, J., West, S., & MacKinnon, D. (1997). Effects of sample size and nonnormality on the estimation of mediated effects in latent variable models. Structural Equation Modeling, 4, 87-107.

Green, S., Akey, T., Fleming, K., Hershberger, S. & Marquis, J. (1997). Effect of the number of scale points on chi-square fit indices in confirmatory factor analysis.

Structural Equation Modeling, 4, 108-120.

Hair, J., Anderson, R., Tatham, R. & Black, W. (1995). Multivariate Data Analysis: With Readings (4th ed.). New Jersey: Prentice-Hall.

Hayduk, L. (1987). Structural equation modeling with LISREL: Essentials and Advances. Baltimore: The Johns Hopkins University Press.

Hu, L. & Bentler, P. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), Structural Equation Modeling: Concepts, Issues, and Applications. (pp. 76-99). Thousand Oaks: Sage

Hu, L. & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Structural Equation Modeling, 6 (1), 1-55.

Hu, L., Bentler, P. & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted?. Psychological Bulletin, 112, 351-362.

Hutchinson, S. & Olmos, A. (1998). Behavior of descriptive fit indexes in confirmatory factor analysis using ordered categorical data. Structural Equation Modeling, 5, 344-364.

Jöreskog, K. (1969). A general approach to confirmatory maximum likelihood factor analysis. Psychometrica, 32, 443-482.

Jöreskog, K. (1993). Testing structural equation models. In Bollen, K. & Long, J. (Eds.). Testing Structural Equation Models. (pp. 294-316), Newberry Park, CA: Sage Publications.

Jöreskog, K. & Sörbom, D. (1984). LISREL VI- User's Guide (3rd ed.). Mooresville: Scientific Software.

Jöreskog, K. & Sörbom, D. (1996). LISREL 8: User's Reference Manual. Chicago: Scientific Software International.

Kaplan, D. (1988). The impact of specification error on the estimation, testing, and improvement of structural equation models. Multivariate Behavioral Research, 23, 69-86.

Kline, R. (1998). Principles and Practice of Structural Equation Modeling. New York: The Guilford Press.

Marsh, H. (1995). Δ2 and χ^2I2 fit indices for structural equation models: A brief note of clarification. Structural Equation Modeling, 2, 246-254.

Marsh, H. (1998). Pairwise deletion of missing data in structural equation models: Nonpositive definite matrices, parameter estimates, goodness of fit, and adjusted sample sizes. Structural Equation Modeling, 5, 22-36.

Marsh, H., Balla, J. & McDonald, R. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. Psychological Bulletin, 103, 391-410.

Maruyama, G. (1998). Basics of Structural Equation Modeling. Thousand Oaks: Sage Publications.

McDonald, R., & Marsh, H. (1990). Choosing a multivariate model: Noncentrality and goodness-of-fit. Psychological Bulletin, 107, 247-255.

Mulaik, S., James, L., Van Alstine, J., Bennet, N., Lind, S., & Stillwell, C. (1989). Evaluation of goodness-of-fit indices for structural equation models. Psychological Bulletin, 10,. 430-445.

Rigdon, E. (1996). CFI vs. RMSEA: A comparison of two fit indexes for structural equation modeling. Structural Equation Modeling, 3, 369-379.

Schumacker, R. & Lomax, R. (1996). A Beginner's Guide to Structural Equation Modeling. Mahwah: Lawrence Erlbaum Associates.

Sobel, M., & Bohrnstedt, G. (1985). Use of null models in evaluating the fit of covariance structure models. In N.B. Tuma (Ed.). Sociological Methodology, (pp. 152-78). San Francisco: Josey-Bass.

Steiger, J. (1990). Structural model evaluation and modification: An interval estimation approach. Multivariate Behavioral Research, 25, 173-180.

Steiger, J. & Lind, J. (1980). Statistically based tests for the number of common factors. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Tabachnick, B. & Fidell,L. (1996). Using Multivariate Statistics (3rd ed.). New York: Harper Collins.

Tanaka, J. (1993). Multifaceted conceptions of fit in structural equation models. In Bollen, K. & Long, J. (Eds.). Testing Structural Equation Models (pp. 10-39). Newberry Park, CA: Sage Publications.

Tanaka, J. & Huba, G. (1989). A general coefficient of determination for covariance structure models under arbitrary GLS estimation. British Journal of Mathematical and Statistical Psychology, 42, 233-39.

Tucker, L., & Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. Psychometrica, 38, 1-10.

Wang, L., Fan, X., & Willson, V. (1996). Effects of nonnormal data on parameter estimates and fit indices for a model with latent and manifest variables: An empirical study. Structural Equation Modeling, 3. 228-247.

Wheaton, B.,Muthen, B., Alwin, B., & Summers, G. (1977). Assessing reliability and stability in panel models. In D. Heise (Ed.), Sociological Methodology 1977 (pp.84-136). San Francisco: Jossey-Bass.

Williams, L.,& Holahan, P. (1994). Parsimony-based fit indices for multiple indicator models: Do they work?. Structural Equation Modeling, 1, 161-189.

ABSTRACT

GOODNESS OF FIT AS A SINGLE FACTOR STRUCTURAL EQUATION MODEL

by

JAMES A. GULLEN

December 2000

Advisor: Dr. Shlomo Sawilowsky

Major: Educational Evaluation and Research

Degree: Doctor of Philosophy

The assessment of model data fit in structural equation modeling is a complicated and rapidly expanding area of research in the social sciences. Numerous fit indices have been proposed by researchers which attempt to capture the degree to which a structural equation model agrees with the observed data. Numerous studies have shown that different fit indices are affected to different extents by such things as degree of non-normality of the data, estimation method employed, and mis-specification within the model. Despite these differing properties, new students to structural equation modeling are often told that the fit indices will all tend to agree with each other in their assessment of the model data fit. This study investigated whether or not this agreement manifests

itself in a single factor model in which multiple fit indices load on a latent factor that could be thought of as the construct good fit. The results of this study showed that the one factor model was an over-simplification. Rather than loading on a single factor, the fit indices included in this study were affected by two factors. The two factors can be thought of as absolute discrepancy and comparative discrepancy. Additionally, results of this study showed that the observed fit indices were affected differentially by the two latent factors. Suggestions for further research were made, including one based on a somewhat surprising result found with respect to the goodness of fit index and other fit indices derived from it.

AUTOBIOGRAPHICAL STATEMENT

JAMES ANDREW GULLEN

Birth: Fort Wayne, Indiana; November 8, 1968

Education:
Secondary. Southgate Anderson High School, Southgate Michigan, 1987.
Undergraduate. Wayne State University, Detroit, Michigan, Bachelor's of Science in Mathematics Education, 1992.
Michigan Provisional Secondary Teaching Certificate, 1992.
Michigan Professional Secondary Teaching Certificate, 1996.
Graduate. Wayne State University, Detroit, Michigan, Master's of Education in Educational Evaluation and Research, 1996.
Doctoral. Wayne State University, Detroit, Michigan, Doctor of Philosophy, 2000. Major: Educational Evaluation and Research. Cognate: Community Medicine.

Positions:
Mathematics Teacher, Southgate Community School District, Southgate, Michigan, 1993-1998.
Assessment Consultant, Wayne RESA, Wayne, Michigan, 1998-present.
Adjunct Faculty, College of Education, Wayne State University, Detroit, Michigan, 1999-present.
Associate Editor for Graduate Work, Journal of Modern Applied Statistical Methods, 2000-present.
Section Bassist, Southern Great Lakes Symphony, Flat Rock, Michigan, 2000-present.

Memberships:
American Statistical Association, 1998-present.
National Council on Measurement in Education, 1998-present.
Michigan Educational Research Association, 1999-present.

Publication:
Rothenberg, L. & Gullen, J. (1999). Assessing the appropriateness of different estimators for competing models with likert scaled data using bootstrapping. Poster presented at the Joint Statisitical Meetings of the American Statistical Association, Baltimore, Maryland, August 8-12, 1999.