Structural Equation Modeling in Educational Psychology

by

Bune Choi

A Thesis Presented to the

FACULTY OF THE GRADUATE SCHOOL

UNIVERSITY OF SOUTHERN CALIFORNIA

In Partial Fulfillment of the

Requirements for the Degree

MASTER OF SCIENCE

(Statistics)

August 1995

UMI Number: 1378407

UNIVERSITY OF SOUTHERN CALIFORNIA
THE GRADUATE SCHOOL
UNIVERSITY PARK
LOS ANGELES, CALIFORNIA 90007

This thesis, written by

Bune Choi

under the direction of h........ Thesis Committee, and approved by all its members, has been presented to and accepted by the Dean of The Graduate School, in partial fulfillment of the requirements for the degree of

MASTERS OF SCIENCE

Alice C. Parker

Dean

Date August 25, 1995

THESIS COMMITTEE

Simon Tavaré

Chairman

Richard Arratia

[illegible]

Dedication

This thesis is dedicated to my parents.

Acknowledgments

I would like to offer my sincerest thanks to my academic advisor, Dr. Tavaré, who provided me with valuable advice and materials. Without his academic and financial support, I could not have completed my study successfully. Also, I would like to acknowledge my committee members, Dr. Arratia and Dr. Goldstein. In addition, Dr. Hocevar deserves recognition for his help with EQS techniques.

Contents

List Of Tables

List Of Figures

Abstract

This thesis focuses on a review of the theoretical foundation, and an empirical example, of structural equation modeling. The theoretical review includes model specification, identification, estimation, assessment of fit, and respecification. An example from the field of educational psychology is provided to illustrate the theory, together with a detailed discussion of the results. The example involves the relations among three latent variables (worry, visualization, and problem-solving ability) on the performance of students in Calculus. The thesis concludes with a brief discussion of some difficulties of this method.

Chapter 1

Introduction

Substantive use of structural equation modeling (SEM) has been growing in psychology and the social sciences. In fact, SEM has been known as a unified model which joins models and methods from econometrics, psychometrics, sociometrics, and multivariate statistics (Bentler, 1992). The generality and wide applicability of the structural equation model approach has been amply demonstrated (Jöreskog & Sörbom, 1989; Bentler, 1992).

The genesis of SEM can be found in the idea of path analysis. Path analysis is defined as a strategy for understanding causal processes through the analysis of correlational data. After development by the geneticist Sewall Wright (1921) as a quantitative aid for biological research, path analysis was introduced to the social sciences by Simon (1954, 1957). Later, Blalock (1961, 1962, 1964) extended and popularized Simon's work. Through additional contributions by Boudon (1965) and Duncan (1966), path analysis became a viable method for rationally inferring causal relationships from correlations, provided certain highly restrictive assumptions are met.

With all these accomplishments of path analysis, Jöreskog (1973), Keesing (1972), and Wiley (1973) developed very general structural equation models that incorporated path diagrams and other features of path analysis into their presentations. These techniques are known by the abbreviation of the JKW model, or more commonly as the LISREL model. Later, Bentler and Weeks (1980), McArdle and McDonald (1984), and others have proposed alternative representations of general structural equations.

As is path analysis, structural equation modeling is a method for estimating the magnitude of the causal relationships that are assumed to operate among the variables in the model. The term "structural" stands for the assumption that the parameters are not just descriptive measures of association but rather that they reveal an invariant "causal" relation. However, converting association into causation is not that simple and needs strong restrictions and assumptions, as many researchers indicate (Cliff, 1983; Freedman 1986, 1993; Anderson and Gerbing, 1988). Therefore, this paper focuses on the theoretical review of SEM for causal inference along with an empirical example of the SEM method. Chapter 2 provides a brief review on the theoretical foundation of SEM. In Chapter 3, a practical example of the use of SEM is presented.

Chapter 2

Structural Equation Model (SEM)

The general structural equation model represents a synthesis of two model types: one is measurement model (or, a confirmatory factor analysis); and the other is a latent variable model (or, structural model). A confirmatory measurement, or factor analysis, model specifies the relation of the observed measures to their posited underlying constructs, with the constructs allowed to intercorrelate freely. On the other hand, a structural model shows the influence of latent variables on each other.

In this chapter, we briefly summarize model specification, identification, estimation, assessment of fit, and respecification from Bollen (1989) and Byrne (1994).

2.1 Model Specification

The first component of the structural equation is the latent variable model,

$$\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}$$

where $\boldsymbol{\eta}$ is the $m \times 1$ vector of latent endogenous random variables; $\boldsymbol{\xi}$ is the $n \times 1$ latent exogenous random variables; $\boldsymbol{B}$ is the $m \times m$ coefficient matrix showing the influence of the latent endogenous variables on each other; and $\boldsymbol{\Gamma}$ is the $m \times n$ coefficient matrix for the effects of $\boldsymbol{\xi}$ on $\boldsymbol{\eta}$. The matrix $(\boldsymbol{I} - \boldsymbol{B})$ is assumed to be nonsingular. $\boldsymbol{\zeta}$ is the disturbance vector that is assumed to have an expected value of zero $E(\boldsymbol{\zeta}) = \mathbf{o}$ and which is uncorrelated with $\boldsymbol{\xi}$. Also, it is assumed that $E(\boldsymbol{\eta}) = \mathbf{o}$ and $E\ (\boldsymbol{\xi}) = \mathbf{o}$.

The second component of the general structural equation model is the measurement model,

$$\boldsymbol{y} = \boldsymbol{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\epsilon}$$

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

where the $\boldsymbol{y}(p \times 1)$ and $\boldsymbol{x}(q \times 1)$ vectors are observed variables; $\boldsymbol{\Lambda}_y(p \times m)$ and $\boldsymbol{\Lambda}_x(q \times n)$ are the coefficient matrices that show the relation of $\boldsymbol{y}$ to $\boldsymbol{\eta}$ and $\boldsymbol{x}$ to $\boldsymbol{\xi}$, respectively; and $\boldsymbol{\epsilon}(p \times 1)$ and the $\boldsymbol{\delta}(q \times 1)$ are the errors of measurement for $\boldsymbol{y}$ and $\boldsymbol{x}$, respectively. The errors of measurement $\boldsymbol{\epsilon}$ and $\boldsymbol{\delta}$ are assumed to be uncorrelated with $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$, and with each other. $\boldsymbol{\eta}$, $\boldsymbol{\xi}$, $\boldsymbol{\epsilon}$, and $\boldsymbol{\delta}$ are also assumed to have an expected value of zero. All of these relations and assumptions are summarized in the example of the path diagram in Figure 2.1.

The procedures of SEM emphasize covariances rather than cases. In SEM, we minimize the difference between the sample covariances and the covariances predicted by the model, instead of minimizing functions of observed and predicted

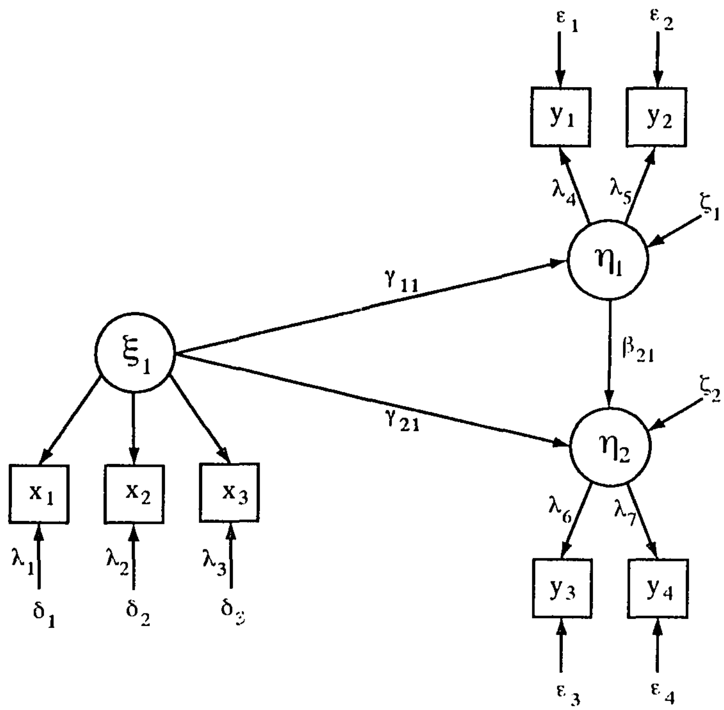

Figure 2.1: Example of path diagram for SEM

individual values. The fundamental hypothesis for the structural equation procedures is that the covariance matrix of the observed variables is a function of a set of parameters:

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$$

where $\boldsymbol{\Sigma}$ is the population covariance matrix of observed variables; $\boldsymbol{\theta}$ is a vector that contains the model parameters; and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is the covariance matrix written as a function of $\boldsymbol{\theta}$. That is, each element of the covariance matrix is a function of one or more model parameters.

The implied covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ can be decomposed into three pieces: the covariance matrix of $\boldsymbol{y}$, $\boldsymbol{\Sigma}_{yy}(\boldsymbol{\theta})$; the covariance matrix of $\boldsymbol{y}$ with $\boldsymbol{x}$, $\boldsymbol{\Sigma}_{yx}(\boldsymbol{\theta})$; and the covariance matrix of $\boldsymbol{x}$, $\boldsymbol{\Sigma}_{xx}(\boldsymbol{\theta})$. Consider first $\boldsymbol{\Sigma}_{yy}(\boldsymbol{\theta})$, the implied covariance matrix of $\boldsymbol{y}$:

$$\begin{aligned}
\boldsymbol{\Sigma}_{yy}(\boldsymbol{\theta}) &= E\,(\boldsymbol{y}\boldsymbol{y}') \\
&= E\left[(\boldsymbol{\Lambda}_y\boldsymbol{\eta}+\boldsymbol{\epsilon})(\boldsymbol{\eta}'\boldsymbol{\Lambda}_y'+\boldsymbol{\epsilon}')\right] \\
&= \boldsymbol{\Lambda}_y E(\boldsymbol{\eta}\boldsymbol{\eta}')\boldsymbol{\Lambda}_y' + \boldsymbol{\Theta}_\epsilon
\end{aligned}$$

where $\boldsymbol{\Theta}_\epsilon$ = covariance matrix of $\boldsymbol{\epsilon}$.

Since $\boldsymbol{\eta} = (\boldsymbol{I}-\boldsymbol{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\xi}+\boldsymbol{\zeta})$,

$$\boldsymbol{\Sigma}_{yy}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}_y(\boldsymbol{I}-\boldsymbol{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}'+\boldsymbol{\Psi})\left[(\boldsymbol{I}-\boldsymbol{B})^{-1}\right]'\boldsymbol{\Lambda}_y' + \boldsymbol{\Theta}_\epsilon$$

where $\boldsymbol{\Phi}$ = covariance matrix of $\boldsymbol{\xi}$ and $\boldsymbol{\Psi}$ = covariance matrix of $\boldsymbol{\zeta}$.

When referring to the covariance matrix of $\boldsymbol{y}$ with $\boldsymbol{x}$, $\boldsymbol{\Sigma}_{yx}$, as a function of the structural parameters, it is $\boldsymbol{\Sigma}_{yx}(\boldsymbol{\theta})$.

$$\begin{aligned}
\boldsymbol{\Sigma}_{yx}(\boldsymbol{\theta}) &= E(\boldsymbol{y}\boldsymbol{x}') \\
&= E\left[(\boldsymbol{\Lambda}_y\boldsymbol{\eta}+\boldsymbol{\epsilon})(\boldsymbol{\xi}'\boldsymbol{\Lambda}_x'+\boldsymbol{\delta}')\right] \\
&= \boldsymbol{\Lambda}_y E(\boldsymbol{\eta}\boldsymbol{\xi}')\boldsymbol{\Lambda}_x'
\end{aligned}$$

Again using $\boldsymbol{\eta} = (\boldsymbol{I}-\boldsymbol{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\xi}+\boldsymbol{\zeta})$,

$$\boldsymbol{\Sigma}_{yx}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}_y(\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Lambda}_x'$$

Finally, the covariance matrix of $\boldsymbol{x}$, $\sum_{xx}$, written as a function of the structural parameters is:

$$\begin{aligned}\boldsymbol{\Sigma}_{xx}(\boldsymbol{\theta}) &= E(\boldsymbol{x}\boldsymbol{x}') \\ &= E\Big[(\boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta})(\boldsymbol{\xi}'\boldsymbol{\Lambda}_x' + \boldsymbol{\delta}')\Big] \\ &= \boldsymbol{\Lambda}_x E(\boldsymbol{\xi}\boldsymbol{\xi}')\boldsymbol{\Lambda}_x' + \boldsymbol{\Theta}_\delta \\ &= \boldsymbol{\Lambda}_x \boldsymbol{\Phi} \boldsymbol{\Lambda}_x' + \boldsymbol{\Theta}_\delta\end{aligned}$$

where $\boldsymbol{\Theta}_\delta$ = covariance matrix of $\boldsymbol{\delta}$.

Therefore, by assembling these three components, the covariance matrix for the observed $\boldsymbol{y}$ and $\boldsymbol{x}$ variables can be represented as a function of the model parameters:

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\Sigma}_{yy}(\boldsymbol{\theta}) & \boldsymbol{\Sigma}_{yx}(\boldsymbol{\theta}) \\ \boldsymbol{\Sigma}_{xy}(\boldsymbol{\theta}) & \boldsymbol{\Sigma}_{xx}(\boldsymbol{\theta}) \end{bmatrix}$$

2.2 Identification

In the previous section, we showed that the covariance structure $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ implies $\frac{1}{2}(p+q)(p+q+1)$ nonredundant equations of the form $\sigma_{ij} = \sigma_{ij}(\boldsymbol{\theta})(i \leq j)$, where σ_{ij} is the ij element of $\boldsymbol{\Sigma}$ and $\sigma_{ij}(\boldsymbol{\theta})$ is the ij element of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. If an element of $\boldsymbol{\theta}$ can be expressed as a function of one or more σ_{ij}, then this establishes its identification. If all elements of $\boldsymbol{\theta}$ meet this condition, the model is identified.

There are widely used rules for the identification of a general model. The t-rule, the two-step rule, and MIMIC (Multiple Indicators and Multiple Causes) rule are reviewed in this section. Even though none of these rules is a necessary and sufficient condition for model identification, researchers can apply one or more of these to help assess a model's identification.

2.2.1 t-Rule

The t-rule for identification is that the number of nonredundant elements in the covariance matrix of the observed variables must be greater than or equal to the number of unknown parameters in $\boldsymbol{\theta}$:

$$t \leq \frac{1}{2}(p+q)(p+q+1)$$

where $p+q$ is the number of observed variables and t is the number of free and unconstrained parameters in $\boldsymbol{\theta}$. The nonredundant elements of $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ imply $\frac{1}{2}(p+q)(p+q+1)$ equations. If the number of unknowns in $\boldsymbol{\theta}$ exceeds the number of equations, identification is not possible. This rule is a necessary but not sufficient condition of identification.

2.2.2 Two-Step Rule

The two-step rule consists of two parts: the first part is confirmatory factor analysis; and the second part is path analysis of latent variables. In the first step, a model is reformulated as a measurement model, viewing the original $\boldsymbol{x}$ and $\boldsymbol{y}$ as x variables and the original $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ as ξ variables. The only relationships between the latent variables that are of concern are their variances and covariances ($\boldsymbol{\Phi}$). If identification can be established for the confirmatory factor analysis, the second step of the identification can be applied.

The second step concerns establishing identification of latent variable model as if latent variables were observed with no measurement error. That is, treat latent variable model as a structural equation in observed variables. Then it is ready to determine whether $\boldsymbol{B}$, $\boldsymbol{\Gamma}$, and $\boldsymbol{\Psi}$ are identified. If the first step shows that the measurement parameters are identified and the second step shows that the latent variable model parameters also are identified, then this is sufficient to identify the whole model.

2.2.3 MIMIC Rule

MIMIC models contain observed variables that are multiple indicators and multiple causes of a single latent variable. The equations for this model are:

$$\begin{aligned}\boldsymbol{\eta}_1 &= \boldsymbol{\Gamma x} + \boldsymbol{\zeta}_1 \\ \boldsymbol{y} &= \boldsymbol{\Lambda}_y \boldsymbol{\eta}_1 + \boldsymbol{\epsilon} \\ \boldsymbol{x} &= \boldsymbol{\xi}\end{aligned}$$

The equations show that $\boldsymbol{x}$ is a perfect measure of $\boldsymbol{\xi}$ and that only one latent variable, $\boldsymbol{\eta}_1$, is present. The variable $\boldsymbol{\eta}_1$ is directly affected by one or more x variables, and it is indicated by one or more y variables.

Identification of MIMIC models that conform to the above equations follows if p (the number of y's) is two or greater and q (the number of x's) is one or more, provided that $\boldsymbol{\eta}_1$ is assigned a scale. The MIMIC rule that $p \geq 2$ and $q \leq 1$ is a sufficient condition for identification but not a necessary one.

2.3 Estimation

The estimation procedures derive from the relation of the covariance matrix of the observed variables to the structural parameters. In the section on model specification, it was shown that the covariance matrix is:

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\Sigma}_{xx}(\boldsymbol{\theta}) & \boldsymbol{\Lambda}_y(\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Lambda}_x' \\ \boldsymbol{\Lambda}_x(\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Lambda}_y' & \boldsymbol{\Lambda}_x\boldsymbol{\Phi}\boldsymbol{\Lambda}_x' + \boldsymbol{\Theta}_\delta \end{bmatrix}$$

where

$$\boldsymbol{\Sigma}_{xx}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}_y(\boldsymbol{I}-\boldsymbol{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}' + \boldsymbol{\Psi})\left[(\boldsymbol{I}-\boldsymbol{B})^{-1}\right]'\boldsymbol{\Lambda}_y' + \boldsymbol{\Theta}_\epsilon$$

The unknown parameters in $\boldsymbol{B}$, $\boldsymbol{\Gamma}$, $\boldsymbol{\Phi}$, and $\boldsymbol{\Psi}$ are estimated so that the implied covariance matrix $\widehat{\boldsymbol{\Sigma}}$ $(= \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}))$, is close to the sample covariance matrix $\boldsymbol{S}$.

Many different fitting functions for the task are possible. The fitting functions $F(\boldsymbol{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ are based on $\boldsymbol{S}$, the sample covariance matrix, and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, the implied covariance matrix of the structural parameters. The fitting functions have the following properties:

(1) $F(\boldsymbol{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ is a scalar

(2) $F(\boldsymbol{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta})) \geq 0$

(3) $F(\boldsymbol{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta})) = 0$ if and only if $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{S}$

(4) $F(\boldsymbol{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ is continuous in $\boldsymbol{S}$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$

Three such fitting functions, maximum likelihood (ML), unweighted least squares (ULS), and generalized least squares (GLS) are reviewed here.

2.3.1 Maximum Likelihood (ML)

In deriving F_{ML}, the set of N independent observations are of the multinormal random variables $\boldsymbol{y}$ and $\boldsymbol{x}$. If we combine $\boldsymbol{y}$ and $\boldsymbol{x}$ into a single $(p+q) \times 1$ vector $\boldsymbol{z}$, where $\boldsymbol{z}$ consists of deviation scores, its probability density function is

$$f(\boldsymbol{z}; \boldsymbol{\Sigma}) = (2\pi)^{-(p+q)/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left[-\frac{1}{2}\boldsymbol{z}'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}\right]$$

For a random sample of N independent observations of $\boldsymbol{z}$, the joint density is

$$f(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N; \boldsymbol{\Sigma}) = f(\boldsymbol{z}_1; \boldsymbol{\Sigma}) f(\boldsymbol{z}_2; \boldsymbol{\Sigma}) \cdots f(\boldsymbol{z}_N; \boldsymbol{\Sigma})$$

With a given sample, the likelihood function is

$$L(\boldsymbol{\theta}) = (2\pi)^{-N(p+q)/2} |\boldsymbol{\Sigma}(\boldsymbol{\theta})|^{-N/2} \exp\left[-\frac{1}{2}\sum_{i=1}^{N} \boldsymbol{z}_i'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{z}_i\right]$$

The log of the likelihood function is

$$\log L(\boldsymbol{\theta}) = \frac{-N(p+q)}{2}\log(2\pi) - \frac{N}{2}\log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| - \frac{1}{2}\sum_{i=1}^{N} \boldsymbol{z}_i'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{z}_i$$

The last term of the log of the likelihood function can be rewritten as

$$
\begin{aligned}
-\frac{1}{2}\sum_{i=1}^{N} \boldsymbol{z}_i' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{z}_i &= -\frac{1}{2}\sum_{i=1}^{N} \mathrm{tr}\left[\boldsymbol{z}_i' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{z}_i\right] \\
&= -\frac{N}{2}\sum_{i=1}^{N} \mathrm{tr}\left[\boldsymbol{N}^{-1}\boldsymbol{z}_i\boldsymbol{z}_i' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right] \\
&= -\frac{N}{2}\mathrm{tr}\left[\boldsymbol{S}^{*}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right]
\end{aligned}
$$

where $\boldsymbol{S}^{*}$ is the sample ML estimator of the covariance matrix which employs N rather than $(N-1)$ in the denominator. Using the rewritten term, the log of the likelihood function can be represented as

$$
\begin{aligned}
\log L(\boldsymbol{\theta}) &= \mathit{constant} - \frac{N}{2}\log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| - \frac{N}{2}\mathrm{tr}\left[\boldsymbol{S}^{*}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right] \\
&= \mathit{constant} - \frac{N}{2}\left\{\log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| + \mathrm{tr}\left[\boldsymbol{S}^{*}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right]\right\}
\end{aligned}
$$

Based on the log of the likelihood function, the fitting function that is minimized can be represented as

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

The unbiased sample covariance matrix $\boldsymbol{S}$ is used in F_{ML}, while the ML estimator $\boldsymbol{S}^{*}$ is used in $\log L(\boldsymbol{\theta})$. Since $\boldsymbol{S}^{*} =$ [(N-1)/N] $\boldsymbol{S}$, these matrices will be essentially equal in large samples.

2.3.2 Unweighted Least Squares (ULS)

The ULS fitting function is

$$F_{ULS} = \frac{1}{2}\text{tr}\left[(\boldsymbol{S} - \boldsymbol{\Sigma}(\boldsymbol{\theta}))^2\right]$$

F_{ULS} minimizes one-half the sum of squares of each element in the residual matrix $(\boldsymbol{S} - \boldsymbol{\Sigma}(\boldsymbol{\theta}))$. The residual matrix in this case consists of the differences between the sample variances and covariances and the corresponding ones predicted by the model. See Bollen (1989) for further details.

2.3.3 Generalized Least Squares (GLS)

Since F_{ULS} is not scale invariant, it would seem reasonable to apply a GLS fitting function that weights the elements of $(\boldsymbol{S} - \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ according to their variances and covariances with other elements. A general form of the GLS fitting function is

$$F_{GLS} = \frac{1}{2}\text{tr}\left(\{[\boldsymbol{S} - \boldsymbol{\Sigma}(\boldsymbol{\theta})]\,\boldsymbol{W}^{-1}\}^2\right)$$

where $\boldsymbol{W}^{-1}$ is a weight matrix for the residual matrix.

For selecting the "correct" weighting matrix $\boldsymbol{W}^{-1}$, we make two assumptions:

(1) $E(s_{ij}) = \sigma_{ij}$

(2) the asymptotic distribution of the elements of $\boldsymbol{S}$ is multinormal with means of σ_{ij} and asymptotic covariances of s_{ij} and s_{gh} equal to $N^{-1}(\sigma_{ig}\sigma_{jh} + \sigma_{ih}\sigma_{jg})$

If the assumptions are satisfied, then $\boldsymbol{W}^{-1}$ should be chosen so that $\text{plim}\boldsymbol{W}^{-1} = c\boldsymbol{\Sigma}^{-1}$, where c is any constant (typically $c = 1$).

Although many $\boldsymbol{W}^{-1}$ are consistent estimators of $\boldsymbol{\Sigma}^{-1}$, the most common choice is $\boldsymbol{W}^{-1} = \boldsymbol{S}^{-1}$:

$$\begin{aligned} F_{GLS} &= \frac{1}{2}\mathrm{tr}\left(\{[\boldsymbol{S} - \boldsymbol{\Sigma}(\boldsymbol{\theta})]\,\boldsymbol{S}^{-1}\}^2\right) \\ &= \frac{1}{2}\mathrm{tr}\left\{[\boldsymbol{I} - \boldsymbol{\Sigma}(\boldsymbol{\theta})\boldsymbol{S}^{-1}]^2\right\} \end{aligned}$$

This F_{GLS} is found in both LISREL and EQS (Jöreskog & Sörbom, 1989; Bentler, 1992).

2.4 Assessment of Fit

After estimating model parameters, given a converged and proper solution, a researcher should assess how well the specified model accounted for the data. To help in the evaluation of a model, a number of statistical measures of fit have been proposed. For example, the LISREL program provides the probability value associated with the chi-square likelihood ratio test, the goodness-of-fit index, and the root-mean-square residual (Jöreskog & Sörbom, 1986). The chi-square probability value and the normed and nonnormed fit indices (Bentler & Bonett, 1980) are obtained from the EQS program (Bentler, 1985). For the proper model evaluation, it is suggested to examine two kinds of fit measures: overall model fit measures and component fit measures (Bollen, 1989).

2.4.1 Overall Model Fit Measures

The covariance structure hypothesis is that $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$. The overall fit measures help to assess whether the hypothesis is valid, and if not, they help to measure the departure of $\boldsymbol{\Sigma}$ from $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. Since $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ are unavailable, their sample counterparts $\boldsymbol{S}$ and $\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})$ are examined. The $\boldsymbol{S}$ is the usual sample covariance matrix, and $\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})$ is the implied covariance matrix evaluated at the estimate of $\boldsymbol{\theta}$ which minimizes either F_{ML}, F_{GLS}, or F_{ULS}.

2.4.1.1 Residuals

The residual matrix is perhaps the simplest function of $\boldsymbol{S}$ and $\widehat{\boldsymbol{\Sigma}}$ for assessing the overall model fit. Since the null hypothesis, H_0, is $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$, $\boldsymbol{S} - \widehat{\boldsymbol{\Sigma}}$ can be used as the counterpart of $\boldsymbol{\Sigma} - \boldsymbol{\Sigma}(\boldsymbol{\theta})$. The individual sample residual covariances are $(s_{ij} - \hat{\sigma}_{ij})$ where s_{ij} is the ijth element in $\boldsymbol{S}$ and $\hat{\sigma}_{ij}$ is the corresponding element in $\widehat{\boldsymbol{\Sigma}}$. The individual residuals can help in assessing model fit as Jöreskog and Sörbom (1986) proposed:

$$RMR = \left[2 \sum_{i=1}^{q} \sum_{j=1}^{i} \frac{(s_{ij} - \hat{\sigma}_{ij})^2}{q(q+1)} \right]^{1/2}$$

For a "good" model, all individual residuals should be near zero. However, the sample residuals are affected by several factors:

(1) the departure of $\boldsymbol{\Sigma}$ from $\boldsymbol{\Sigma}(\boldsymbol{\theta})$

(2) the scales of the observed variables

(3) sampling error

The most interesting factor is (1), that is, whether $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$. When $\boldsymbol{\Sigma} \neq \boldsymbol{\Sigma}(\boldsymbol{\theta})$, one or more of the covariances or variances of the observed variables are not exactly predicted by the model. Also, the magnitudes of the individual and mean of the residuals are altered if the observed variables are measured in different units. A big residual can be due to an observed variable with scale units that have a much larger range than that of the other variables. In addition, the expected magnitude of the sample residuals depends on N, even when the null hypothesis is true. Under fairly general conditions, $\boldsymbol{S} - \widehat{\boldsymbol{\Sigma}}$ converges to $\boldsymbol{\Sigma} - \boldsymbol{\Sigma}(\boldsymbol{\theta})$ as $N \to \infty$. For a given model, $(s_{ij} - \hat{\sigma}_{ij})$ tends to be smaller, the bigger is the sample. So, in judging the residuals for small samples, it is expected that bigger residuals than when examining residuals in large samples, when the model is true in both sample.

Considering the negative factors previously stated, some researchers propose corrected residuals. For the scale problem, Bentler (1985) suggests correlation residuals. Each correlation residual is $r_{ij} - \hat{r}_{ij}$ where r_{ij} is the sample correlation of the ith and jth variables, and $\hat{r}_{ij}$ is the model predicted correlation. Individually $(r_{ij} - \hat{r}_{ij})$ gauges how well a correlation (or a standardized variance for $i = j$) is reproduced. A correlation residual should be fairly close to zero for most well-fitting models. On the other hand, Jöreskog and Sörbom (1986) propose a normalized residual:

$$\text{Normalized residual} = \frac{(s_{ij} - \hat{\sigma}_{ij})}{\left[(\hat{\sigma}_{ii}\hat{\sigma}_{jj} + \hat{\sigma}_{ij}^2)/N\right]^{1/2}}$$

The numerator is the residual and the denominator is the square root of its estimated asymptotic variance. This normalized residual provides an approximate correction for such sample size effects and for scaling differences. The largest absolute values of the normalized residuals indicate the s_{ij} elements that are most poorly fit by the model.

2.4.1.2 A Chi-Square (χ^2) Test

The quantities of $(N-1)F_{ML}$ or $(N-1)F_{GLS}$ provide chi-square estimators to test H_0: $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$. Since H_0 is equivalent to the hypothesis that $\boldsymbol{\Sigma} - \boldsymbol{\Sigma}(\boldsymbol{\theta}) = 0$, the chi-square test is a simultaneous test that all residuals in $\boldsymbol{\Sigma} - \boldsymbol{\Sigma}(\boldsymbol{\theta})$ are zero.

Here, we review how the likelihood ratio rationale for the asymptotic chi-square distribution of $(N-1)F_{ML}$ can be established. The null hypothesis indicates that the specification of the fixed, free, and constrained parameters in $\boldsymbol{\Lambda}_x$, $\boldsymbol{\Phi}$, and $\boldsymbol{\Theta}_\delta$ is valid. Under H_0, we have ML estimators of the free and constrained parameters in these matrices that together with the fixed parameters comprise the estimated matrices $\widehat{\boldsymbol{\Lambda}_x}$, $\hat{\boldsymbol{\Phi}}$, and $\widehat{\boldsymbol{\Theta}_\delta}$. Let $\log L_0$ represent the log of the likelihood function corresponding to H_0, and $\log L_1$ represent that corresponding to an alternative hypothesis, H_1. When evaluated at $\boldsymbol{S}$ and $\widehat{\boldsymbol{\Sigma}}$, the $\log L_0$ is

$$\log L_0 = -\frac{N-1}{2}\left\{\log|\widehat{\boldsymbol{\Sigma}}| + \mathrm{tr}(\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{S})\right\}$$

This is the log of the numerator for the likelihood ratio test. If $\widehat{\boldsymbol{\Sigma}}$ is set to $\boldsymbol{S}$ the sample covariance matrix, the $\log L_1$ is at its maximum value. So the $\log L_1$ is

$$\begin{aligned}\log L_1 &= -\frac{N-1}{2}\{\log|\boldsymbol{S}| + \operatorname{tr}(\boldsymbol{S}^{-1}\boldsymbol{S})\} \\ &= -\frac{N-1}{2}\{\log|\boldsymbol{S}| + q\}\end{aligned}$$

This is the log of the denominator for the likelihood ratio. Since H_1 sets $\widehat{\boldsymbol{\Sigma}}$ to $\boldsymbol{S}$, comparing $\log L_1$ to $\log L_0$ evaluates H_0 vis-à-vis a perfect fit, H_1. The natural logarithm of the likelihood ratio, $\log(L_0/L_1)$, when multiplied by -2 is distributed as chi-square variate when H_0 is true and $(N-1)$ is large. In this case

$$\begin{aligned}-2\log\left(\frac{L_0}{L_1}\right) &= -2\log L_0 + 2\log L_1 \\ &= (N-1)\left[\log|\widehat{\boldsymbol{\Sigma}}| + \operatorname{tr}(\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{S})\right] - (N-1)(\log|\boldsymbol{S}| + q) \\ &= (N-1)\left(\log|\widehat{\boldsymbol{\Sigma}}| + \operatorname{tr}(\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{S}) - \log|\boldsymbol{S}| - q\right)\end{aligned}$$

In the last line of the right-hand side of the previous equation, the quantity within parentheses is the fitting function F_{ML} evaluated at $\boldsymbol{S}$ and $\widehat{\boldsymbol{\Sigma}}$. So, the expression of the last line shows that $(N-1)$ times the fitting function F_{ML} evaluated at $\hat{\boldsymbol{\theta}}$ is approximately distributed as a chi-square variate. Its degrees of freedom are $\frac{1}{2}q(q+1) - t$, where the first term is the number of nonredundant elements in $\boldsymbol{S}$ given q observed variables, and t is the number of free parameters in $\boldsymbol{\theta}$.

For the chi-square test of the SEM, the null hypothesis H_0 is that the constraints on $\boldsymbol{\Sigma}$ implied by the model are valid (i.e., $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$). The standard of comparison is the perfect fit of $\widehat{\boldsymbol{\Sigma}}$ equals to $\boldsymbol{S}$. The probability level of the calculated chi-square is the probability of obtaining a χ^2 value larger than the value obtained if H_0 is correct. The higher the probability of the χ^2, the closer is the fit of H_0 to the perfect fit.

The chi-square approximation makes use of several assumptions:

(1) $\boldsymbol{x}$ has no kurtosis

(2) the covariance matrix is analyzed

(3) the sample is sufficiently large

(4) the H_0: $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ hold exactly

When all of these assumptions are satisfied, $(N-1)F_{ML}$ (or $(N-1)F_{GLS}$) is a good approximation to a chi-square variable suitable for tests of statistical significance. However, if one or more of the preceding conditions is violated, then the χ^2 test loses some of its value.

2.4.1.3 Additional Measures of Overall Model Fit

There are several other ways to measure overall model fit. Jöreskog and Sörbom (1986) propose a Goodness of Fit Index (GFI) and an Adjusted GFI for models fitted with F_{ML} and with F_{ULS}:

$$GFI_{ML} = 1 - \frac{\mathrm{tr}\left[(\widehat{\boldsymbol{\Sigma}}^{-1} - \boldsymbol{I})^2\right]}{\mathrm{tr}\left[(\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{S})^2\right]}$$

$$AGFI_{ML} = 1 - \left[\frac{q(q+1)}{2df}\right][1 - GFI_{ML}]$$

$$GFI_{ULS} = 1 - \frac{\mathrm{tr}\left[(\boldsymbol{I} - \widehat{\boldsymbol{\Sigma}})^2\right]}{\mathrm{tr}(\boldsymbol{S}^2)}$$

$$AGFI_{ULS} = 1 - \left[\frac{q(q+1)}{2df}\right][1 - GFI_{ULS}]$$

where

$$df = \left(\tfrac{1}{2}\right)(p+q)(p+q+1) - (\text{number of parameters to be estimated})$$

The GFI_{ML} measures the relative amount of the variances and covariances in $\boldsymbol{S}$ that are predicted by $\widehat{\boldsymbol{\Sigma}}$ The $AGFI_{ML}$ adjusts for the degrees of freedom of a model relative to the number of variables.

Also, Tanaka and Huba (1985) propose GLS versions:

$$GFI_{GLS} = 1 - \frac{\mathrm{tr}\left[(\boldsymbol{I} - \widehat{\boldsymbol{\Sigma}}\boldsymbol{S}^{-1})^2\right]}{q}$$

$$AGFI_{GLS} = 1 - \left[\frac{q(q+1)}{2df}\right][1 - GFI_{GLS}]$$

2.4.2 Component Fit Measures

In addition to the measure of overall fit, an examination of the components of the model is essential since nonsense results for individual parameters can occur. For the component fit of a model, several measures are suggested: parameter estimates, asymptotic standard errors, asymptotic correlation matrix of parameter estimates, and $R^2_{x_i}$ for observed variables. Among them, parameter estimates and $R^2_{x_i}$ for observed variables are discussed here.

For the parameter estimates of $\boldsymbol{\Lambda}_x$, $\boldsymbol{\Phi}$, and $\boldsymbol{\Theta}_\delta$, misspecification of the model could give improper solutions. Improper solutions refer to sample estimates that take values that are impossible in the population, such as negative variances and correlations greater than one. Improper solutions can be caused by several factors. First, the covariance (correlation) matrix analyzed may have outliers or influential observations that lead to distorted measures of association for the observed variables, which in turn affect the parameter estimates. Second, there could be a fundamental fault of specification in the model. The model requires reconstruction based on the researcher's substantive knowledge.

Another measure of component fit is $R^2_{x_i}$ for each x_i variable. It is estimated as

$$R^2_{x_i} = 1 - \frac{\mathrm{Var}(\delta_i)}{\hat{\sigma_{ii}}}$$

where $\hat{\sigma_{ii}}$ = variance of x_i predicted by the model. The $R^2_{x_i}$ is analogous to the squared multiple correlation coefficient, with x_i as the dependent variable and the latent variables (ξ) as the explanatory variables. Generally, the goal is finding measures with high $R^2_{x_i}$'s.

2.5 Respecification

There are many potential causes for low measures of overall fit. A common cause is a misspecified model. The incorrect inclusion or exclusion of a parameter can be the error. So a common response to a poorly fitting model is to respecify it. However, respecification decisions should not be based on statistical considerations alone but rather in conjunction with theory and content considerations. The potentially richest source of ideas for respecification is the theoretical or substantive knowledge of the researcher.

The first respecification necessary is in response to nonconvergence or an improper solution. Nonconvergence can occur because of a fundamentally incongruent pattern of sample covariances that is caused either by sampling error in conjunction with a properly specified model or by a misspecification. Relying on content, one can obtain convergence for the model by respecifying one or more problematic indicators to different constructs or by excluding them from further analysis.

Considering improper solutions, Van Driel (1978) presented three potential causes: sampling variations in conjunction with true parameter values close to zero, a fundamentally misspecified model, and underidentification of the model. Recently Gerbing and Anderson (1987) found that for improper estimates due to sampling error, respecifying the model with the problematic parameter fixed at zero has no appreciable effect on the parameter estimates of other factors or on the overall goodness-of-fit indices.

Given a converged and proper solution but unacceptable overall fit, Anderson and Gerbing (1988) suggest four basic ways for respecification:

(1) relate the indicator to a different factor

(2) delete the indicator from the model

(3) relate the indicator to multiple factors

(4) use correlated measurement errors

Chapter 3

Empirical Application

In this Chapter, an empirical example is presented to show how a structural model can be used for latent variables. The example is a three-factor model with two or three indicators (See Figure 3.1). The structural model is designed to examine the relationships among three latent variables: worry (ξ_1), visualization (ξ_2), and problem-solving ability (η_1). The latent variable worry has two indicators, worry1 (x_1) and worry2 (x_2). Visualization construct also has two indicators, visualization1 (x_3) and visualization2 (x_4). The construct of problem-solving ability has three indicators, self-efficacy (y_1), metacognition (y_2), and cognitive strategy (y_3). For this example, the EQS program (Benter, 1992) is used to analyze the data set.

3.1 Data

The subjects consist of 113 students in a calculus course at University of Southern California. At the end of the semester (Spring, 1995), they were asked to complete

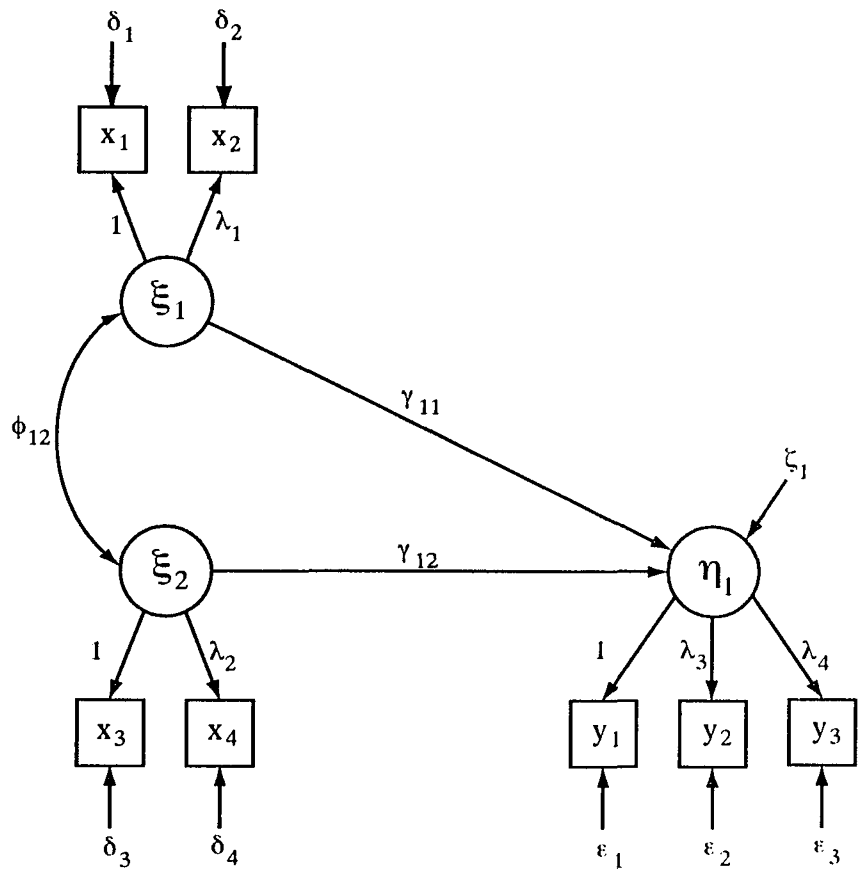

Figure 3.1: Confirmatory Factor Analysis Model for Worry (ξ_1), Visualization (ξ_2), and Problem-Solving Ability (ξ_3)

a questionnaire booklet used to measure three constructs in this study. The Paper Folding Test (French, Ekstrom, & Price, 1963) was used to measure visualization skills and the Self-Assessment Questionnaire (O'Neil & Abedi, 1992) was used to measure worry, self-efficacy, metacognition, and cognitive strategy.

3.2 Model Specification

Based on the theoretical discussion on model specification in the previous chapter, the model is specified in two components: the structural model of the latent variables; and the measurement model of the relationships between latent variables and indicators. The general system for the structural model containing both components is expressed as:

$$\boldsymbol{\eta} = \boldsymbol{B\eta} + \boldsymbol{\Gamma\xi} + \boldsymbol{\zeta}$$

$$\boldsymbol{y} = \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\epsilon}$$

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

3.2.1 Structural Model

The data represents the relations of worry, visualization, and problem-solving ability in Calculus. In many previous research (Hembree, 1992; Malpass, 1994; Mayer, 1990, 1993), worry is considered as a negative effect on problem-solving ability in Mathematics, while visualization is often seen as enhancing it. In addition, worry

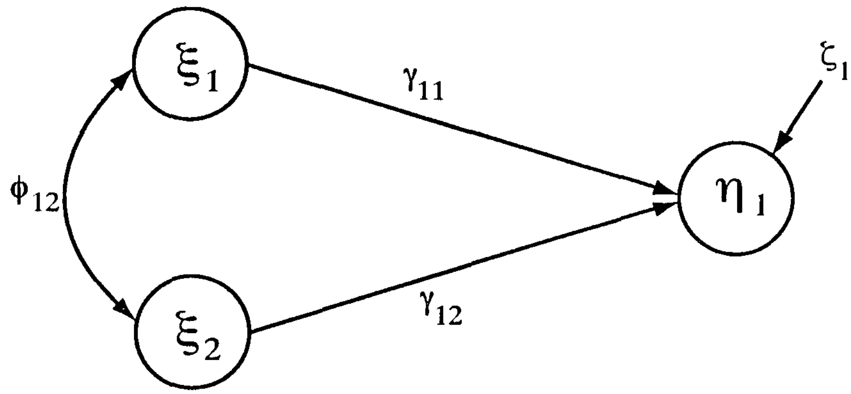

Figure 3.2: A Latent Variables Model of Worry (ξ_1), Visualization (ξ_2), and Problem-Solving Ability (η_1)

and visualization may have negative correlational relation. These idea suggest that the latent variable model should have worry (ξ_1) and visualization (ξ_2) influencing problem-solving ability (η_1). Based on this description, we can form the elements of $\boldsymbol{\Gamma}$ and the latent variable model:

$$\eta_1 = \gamma_{11}\xi_1 + \gamma_{12}\xi_2 + \zeta_1$$

$$[\eta_1] = \begin{bmatrix} \gamma_{11} & \gamma_{12} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + [\zeta_1]$$

The relationships in the latent variable model are represented in the path diagram as shown in Figure 3.2.

3.2.2 Measurement Model

For the measurement model, seven indicators are used to measure three constructs of x part and y part. For x part equations, two indicators are used for each construct of worry (ξ_1) and visualization (ξ_2). The indicators of worry1 (x_1) and worry2 (x_2) and the indicators of visualization1 (x_3) and visualization2 (x_4) are for the worry and the visualization constructs, respectively. The scale of ξ_1 is set to x_1, and that for ξ_2 is set to x_3. Furthermore the coefficient linking ξ_1 and x_2, and that linking ξ_2 and x_4 are unconstrained. The preceding information reveals the pattern of $\boldsymbol{\Lambda}_x$. The $\boldsymbol{x}$ measurement equation is

$$\begin{aligned} x_1 &= \xi_1 + \delta_1 \\ x_2 &= \lambda_2 \xi_1 + \delta_2 \\ x_3 &= \xi_2 + \delta_3 \\ x_4 &= \lambda_4 \xi_2 + \delta_4 \end{aligned}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \lambda_2 & 0 \\ 0 & 1 \\ 0 & \lambda_4 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{bmatrix}$$

For y part, three measures of problem-solving ability (η_1) are self-efficacy (y_1), metacognition (y_2), and cognitive strategy (y_3). Problem-solving ability (η_1) is set to the scale of y_1, whereas coefficients showing η_1's influence on y_2 and y_3 are unconstrained. The $\boldsymbol{y}$ measurement equation is

$$\begin{aligned} y_1 &= \eta_1 + \epsilon_1 \\ y_2 &= \lambda_6 \eta_1 + \epsilon_2 \\ y_3 &= \lambda_7 \eta_1 + \epsilon_3 \end{aligned}$$

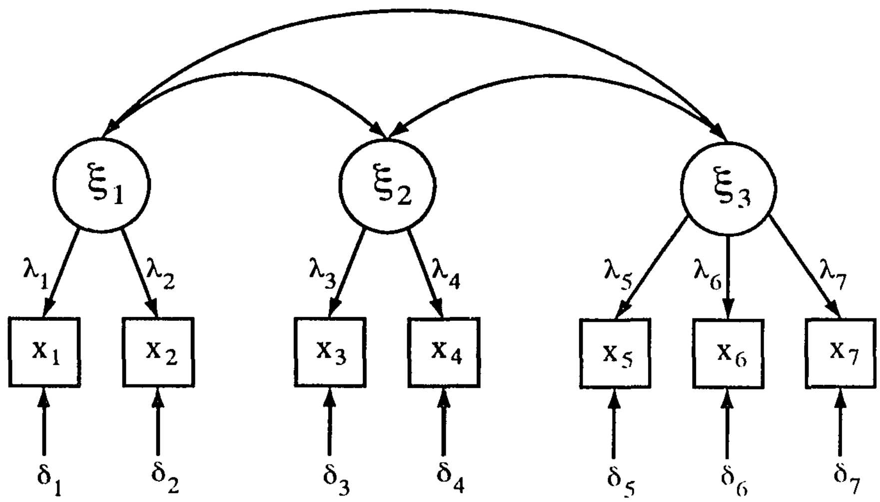

Figure 3.3: Confirmatory Factor Analysis Model for Worry (ξ_1), Visualization (ξ_2), and Problem-Solving Ability (ξ_3)

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ \lambda_6 \\ \lambda_7 \end{bmatrix} [\eta_1] + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}$$

All of these relations are summarized in the path diagram in Figure 3.3. Using the model we have specified, we want to examine the relationships among the latent variables. Similarily with other studies, we expect visualization has a positive effect on problem solving ability, while worry influences it negatively. We expect to investigate the negative relation between worry and visualization as well.

3.3 Identification

The equations for the path diagram of this model is

$$[\eta_1] = \begin{bmatrix} \gamma_{11} & \gamma_{12} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + [\zeta_1]$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \lambda_2 & 0 \\ 0 & 1 \\ 0 & \lambda_4 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ \lambda_6 \\ \lambda_7 \end{bmatrix} [\eta_1] + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}$$

The covariance matrix of the observed variables $\boldsymbol{\Sigma}$ is

$$\begin{bmatrix} \mathrm{Var}(x_1) & & & & & & \\ \mathrm{Cov}(x_2,x_1) & \mathrm{Var}(x_2) & & & & & \\ \mathrm{Cov}(x_3,x_1) & \mathrm{Cov}(x_3,x_2) & \mathrm{Var}(x_3) & & & & \\ \mathrm{Cov}(x_4,x_1) & \mathrm{Cov}(x_4,x_2) & \mathrm{Cov}(x_4,x_3) & \mathrm{Var}(x_4) & & & \\ \mathrm{Cov}(y_1,x_1) & \mathrm{Cov}(y_1,x_2) & \mathrm{Cov}(y_1,x_3) & \mathrm{Cov}(y_1,x_4) & \mathrm{Var}(y_1) & & \\ \mathrm{Cov}(y_2,x_1) & \mathrm{Cov}(y_2,x_2) & \mathrm{Cov}(y_2,x_3) & \mathrm{Cov}(y_2,x_4) & \mathrm{Cov}(y_2,y_1) & \mathrm{Var}(y_2) & \\ \mathrm{Cov}(y_3,x_1) & \mathrm{Cov}(y_3,x_2) & \mathrm{Cov}(y_3,x_3) & \mathrm{Cov}(y_3,x_4) & \mathrm{Cov}(y_3,y_1) & \mathrm{Cov}(y_3,y_2) & \mathrm{Var}(y_3) \end{bmatrix}$$

Substituting the parameter matrices for the above model into the implied covariance matrix derived in the previous chapter shows that $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is

$$\begin{bmatrix} \phi_{11} + \mathrm{V}_1 & & & & & & \\ \lambda_2\phi_{11} & \lambda_2^2\phi_{11} + \mathrm{V}_2 & & & & & \\ \phi_{12} & \lambda_2\phi_{12} & \phi_{22} + \mathrm{V}_3 & & & & \\ \lambda_2\lambda_4\phi_{12} & \lambda_2\lambda_4\phi_{12} & \lambda_4\phi_{22} & \lambda_4^2\phi_{22} + \mathrm{V}_4 & & & \\ \theta_2 & \lambda_2\theta_2 & \theta_3 & \lambda_4\theta_3 & \theta_1 + \mathrm{V}_5 & & \\ \lambda_6\theta_2 & \lambda_2\lambda_6\theta_2 & \lambda_6\theta_3 & \lambda_4\lambda_6\theta_3 & \lambda_6\theta_1 & \lambda_6^2\theta_1 + \mathrm{V}_6 & \\ \lambda_7\theta_2 & \lambda_2\lambda_7\theta_2 & \lambda_7\theta_3 & \lambda_4\lambda_7\theta_3 & \lambda_7\theta_1 & \lambda_6\lambda_7\theta_1 & \lambda_7^2\theta_1 + \mathrm{V}_7 \end{bmatrix}$$

where

$$\begin{aligned} \mathrm{V}_i &= \begin{cases} \mathrm{Var}(\delta_i), & 1 \le i \le 4 \\ \mathrm{Var}(\epsilon_j), & 5 \le i \le 7, 1 \le j \le 3 \end{cases} \\ \theta_1 &= \gamma_{11}^2\phi_{11} + 2\gamma_{11}\gamma_{12}\phi_{12} + \gamma_{12}^2\phi_{12} + \psi_{11} \\ \theta_2 &= \gamma_{11}\phi_{11} + \gamma_{12}\phi_{12} \\ \theta_3 &= \gamma_{11}\phi_{12} + \gamma_{12}\phi_{22} \end{aligned}$$

$$\boldsymbol{\theta}' = \begin{matrix} [\lambda_2 & \lambda_4 & \lambda_6 & \lambda_7 & \gamma_{11} & \gamma_{12} & \phi_{11} & \phi_{12} & \phi_{22}, & \\ & & \mathrm{Var}(\epsilon_1) & \mathrm{Var}(\epsilon_2) & \mathrm{Var}(\epsilon_3) & \mathrm{Var}(\delta_1) & \mathrm{Var}(\delta_2) & \mathrm{Var}(\delta_3) & \mathrm{Var}(\delta_4) & \psi_{11}] \end{matrix}$$

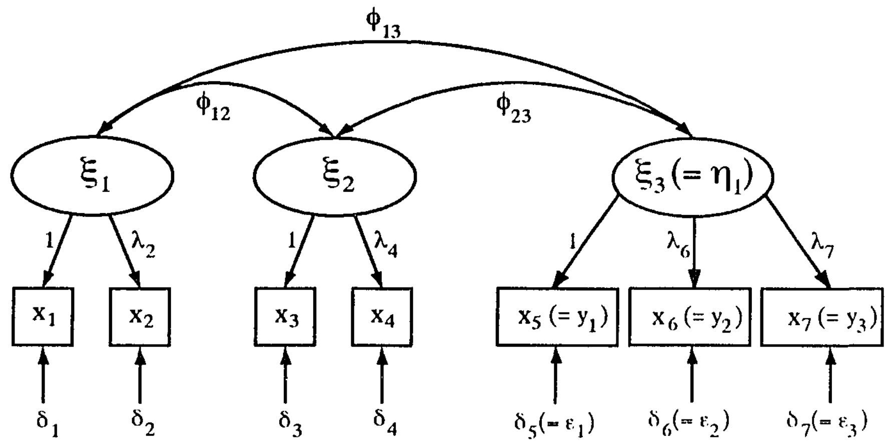

Figure 3.4: The First Step of Reformulation for Two-Step Rule

A quick way to detect some underidentified models is by the t-rule. The covariance structure of $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ leads to twenty eight $[= \frac{1}{2}(7)(8)]$ equations in seventeen unknowns (number of free parameters). Thus the model may be identified.

To further examine its identification, we apply the two-step rule. The first step establishes that all parameters in the measurement model are identified, including the covariance matrix of the latent variables. So, for this example, η_1 is redefined as ξ_3, y_1, y_2, and y_3 are now x_5, x_6, and x_7, ϵ_1, ϵ_2, and ϵ_3 are δ_5, δ_6, and δ_7, and η_1, γ_{11}, and γ_{21} are not considered. Instead, we now examine the variances and covariance of ξ_1, ξ_2, and the new $\xi_3(= \eta_1)$. This reformulation is represented in the Figure 3.4. In the second step, the latent variable model parameters are identified if the latent variables are treated as perfectly measured variables. Figure 3.5 shows the latent variable model.

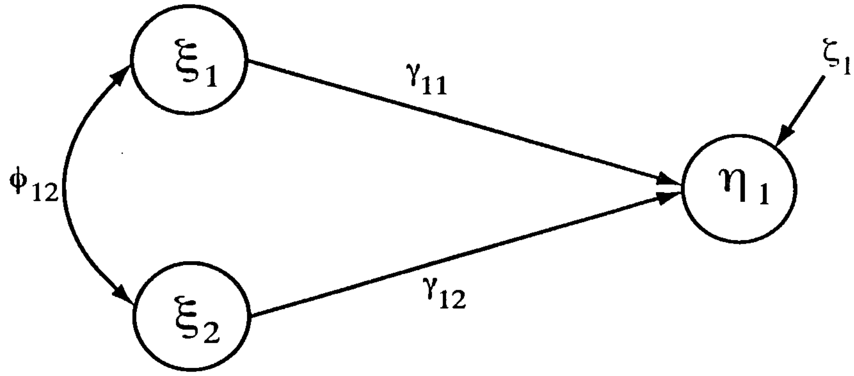

Figure 3.5: The Second Step of Latent Variable Model for Two-Step Rule

3.4 Estimation and Model Evaluation

The hypothesis of our model is $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$. Thus, given the sample covariance matrix of the observed variables, $\boldsymbol{S}$, how can we choose $\boldsymbol{\theta}$ so that $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is close to $\boldsymbol{S}$? In this study, the ML fitting function is used. As shown earlier, the ML fitting function is

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

We want to minimize this function with respect to $\boldsymbol{\theta}$. Similarly, the GLS and ULS fitting functions can be used as many researchers have shown (Bentler, 1992; Bollen, 1989; Jöreskog & Sörbom, 1989).

Based on the fit assessment of the base model and the result of the LMtest, the model is respecified as shown in Figure 3.6. Two correlational relations are added:

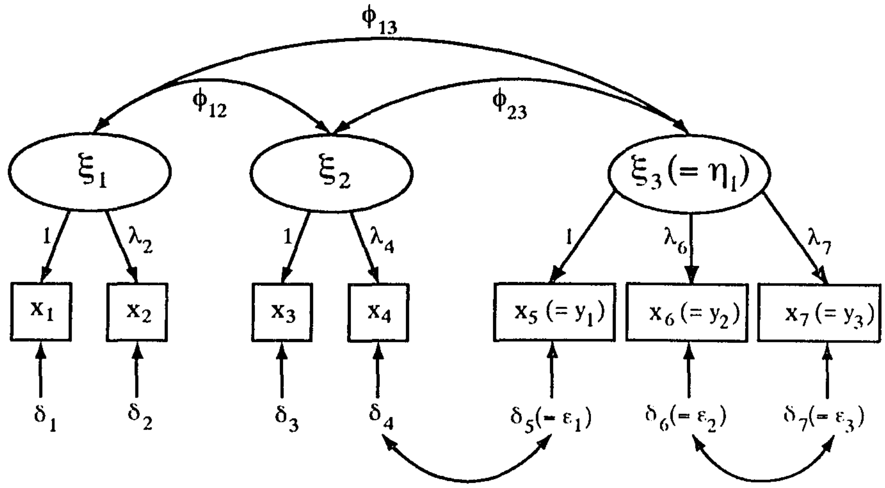

Figure 3.6: The respecified CFA model

visualization2 (x_4) and self-efficacy (x_5): and metacognition (x_6) and cognitive strategy (x_7). The correlational relations of these errors are uniqueness of the indicator shared in common rather than errors.

In our example, the lower half of the sample covariance matrix, $\boldsymbol{S}$, is in Table 3.1.

The residual matrix ($\boldsymbol{S} - \widehat{\boldsymbol{\Sigma}}$) after applying F_{ML} is

$$\boldsymbol{S} - \widehat{\boldsymbol{\Sigma}} = \begin{bmatrix} 0.000 & & & & & & \\ 0.000 & 0.000 & & & & & \\ 0.425 & -0.084 & 0.000 & & & & \\ 0.501 & 0.108 & -0.042 & -0.088 & & & \\ -0.112 & 0.066 & -0.152 & -0.108 & 0.045 & & \\ 0.333 & 0.249 & 0.431 & 1.037 & 0.301 & 0.000 & \\ -0.901 & -0.403 & 0.392 & 0.009 & -0.354 & 0.000 & 0.000 \end{bmatrix}$$

None of the elements of this matrix seems large. The average absolute value of the residuals is 0.219, while the average off-diagonal absolute covariance residuals is reported as 0.286. Compare to the magnitude of the elements in $\boldsymbol{S}$, this is small. Since some elements of a covariance matrix are exactly predicted for a given model

Table 3.1: Covariance Matrix and Standard Deviation for Observed Variables

	x_1	x_2	x_3	x_4	y_1	y_2	y_3
x_1	3.504						
x_2	3.998	8.352					
x_3	0.223	-0.485	2.657				
x_4	0.266	-0.357	2.730	5.443			
y_1	-2.199	-4.083	0.440	1.767	12.292		
y_2	-3.245	-6.863	1.446	2.213	14.283	46.567	
y_3	-5.310	-9.169	1.643	1.459	16.881	52.850	87.068
sd	3.506	6.824	9.331	1.872	2.890	1.630	2.333

and fitting function regardless of the sample covariance matrix, we have some zero elements in the residual matrix.

For the measurement model, the ML estimates $\widehat{\boldsymbol{\Lambda}_x}$, $\widehat{\boldsymbol{\Lambda}_y}$, $\widehat{\boldsymbol{\Theta}_\delta}$, and $\widehat{\boldsymbol{\Theta}_\epsilon}$ are

$$\widehat{\boldsymbol{\Lambda}_x} = \begin{bmatrix} 1.00 & 0.00 \\ 1.98 & 0.00 \\ 0.00 & 1.00 \\ 0.00 & 1.59 \end{bmatrix}$$

$$\widehat{\boldsymbol{\Lambda}_y} = \begin{bmatrix} 1.00 \\ 1.71 \\ 2.11 \end{bmatrix}$$

$$\text{diag}\left(\widehat{\boldsymbol{\Theta}_\delta}\right) = \begin{bmatrix} 1.49 & .40 & .25 & 2.32 \end{bmatrix}$$

$$\text{diag}\left(\widehat{\boldsymbol{\Theta}_\epsilon}\right) = \begin{bmatrix} 4.09 & 22.60 & 50.65 \end{bmatrix}$$

The ML estimates for the path model parameters $\widehat{\boldsymbol{\Gamma}}$, $\widehat{\boldsymbol{\Phi}}$, and $\widehat{\boldsymbol{\Psi}}$ are

$$\widehat{\boldsymbol{\Gamma}} = \begin{bmatrix} -1.022 & .161 \end{bmatrix}$$

$$\widehat{\boldsymbol{\Phi}} = \begin{bmatrix} 2.01 & & \\ -.203 & 2.39 & \\ -2.086 & .592 & 8.15 \end{bmatrix}$$

Table 3.2: The ML Estimates for the parameters

Parameter	ML Estimates	Standard Errors	Standardized values
λ_1	1.000^c	-	.758
λ_2	1.988	.369	.976
λ_3	1.000^c	-	.949
λ_4	1.159	.881	.762
λ_5	1.000^c	-	.816
λ_6	1.714	.376	.717
λ_7	2.113	.489	.647
γ_{11}	-1.022	.229	-.507
γ_{12}	.161	.221	.087
ϕ_{11}	2.011	.540	
ϕ_{12}	-.202	.227	-.092
ϕ_{22}	2.392	1.842	
Var(ϵ_1)	4. 090	1.713	
Var(ϵ_2)	22.599	5.585	
Var(ϵ_3)	50.654	10.239	
Var(δ_1)	1.493	.389	
Var(δ_2)	.404	1.323	
Var(δ_3)	.265	1.808	
Var(δ_4)	2.317	2.449	
ψ_{11}	5.930	1.821	

Note: c = constrained to equal 1.00

$$\hat{\Psi} = \left[\ 5.93 \ \right]$$

These estimates of parameters are reported with their standard errors and standardized values in Table 3.2.

As far as the overall fit of the model is concerned, the measures are summarized in Table 3.3. The overall fit of this model is good. Since the probability value for the χ^2 statistics (.197) with 9 *df* is higher than an α level of .05, we don't reject the

Table 3.3: Summary of Overall Fit

Fit Index	$\chi^2(df)$	NFI	NNFI	CFI	Standardized χ^2	p-value
Estimates	12.29 (9)	.96	.97	.99	12.11	.197

Table 3.4: The Component Fit Measures ($R^2_{x_i}$'s and $R^2_{y_i}$'s)

	x_1	x_2	x_3	x_4	y_1	y_2	y_3
$R^2_{x_i}$	.58	.96	.04	.60	-	-	-
$R^2_{y_i}$	-	-	-	-	.68	.53	.43

hypothesis ($\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$). Besides, all fit indices are extremely high. Therefore, our model is supported by the data.

For better fit assessment of the model, we need to examine the component measures. The $R^2_{x_i}$'s for the worry and visualization indicators and the $R^2_{y_i}$'s for the problem solving ability indicators are shown in Table 3.4. Except for the first indicator of the visualization factor, the squared multiple correlation coefficients for each indicator are moderate to high. Particularly, the second indicator of worry factor (x_2) is extremely high, showing 96% of the variance in x_2 accounted for by the latent worry variable (ξ_1).

In summary, the assessment of fit, both the overall and component fit measures suggest that the model properly matches the data. As we expected, worry (ξ_1) has negative influence on problem solving ability (η_1), whereas visualization skills (ξ_2) enhance the problem solving ability in Calculus. Worry and visualization are negatively correlated. Worry alone explained a modest amount of the variance (about 26%) of problem solving ability. Since the effect of visualization on problem solving ability is not significant, the combined influence of worry and visualization led to about 26% explained variance in problem solving ability. In addition, the indicators of worry and problem solving ability were fairly good, with 43% to 96% of the variance explained by the latent factors, while the indicators of visualization were relatively low with the percentage of explained variances 4% and 60%.

Chapter 4

Conclusions

This paper reviews the theoretical foundation of structural equation modeling along with an empirical example. For the theoretical background, a brief discussion on model specification, identification, estimation, assessment of fit, and respecification is presented. For an empirical example, we show a structural model with three latent variables and seven indicators. The model examines the relationships among three latent variables: worry, visualization, and problem solving ability in Calculus. Worry and visualization have two indicators, whereas problem solving ability has three indicators. Briefly, worry has significant negative influence on problem solving ability while visualization doesn't have significant effect on it.

As other researchers have indicated (Cliff, 1983; Freedman 1987, 1993; Anderson and Gerbing, 1988), there are many difficulties in establishing causal relations. Since the basic modeling techniques of SEM is converting association into causation, some caution is needed in application of the causal modeling method. First, identifying the exogenous variables is a problem since results can depend quite strongly on assumptions of exogeneity. Second, the possible effects of variables that are not included in a model must be considered. Third, a model is never confirmed by data; rather, it gains support by failing to be rejected. In spite of good fit of a model, other models with equal fit may exist. Fourth, the nominalistic fallacy—naming something does not necessarily mean that one understands it—must also be considered. So, validity and reliability of observed variables are required.

Reference List

[1] Anderson, J. C., and Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin,* **103**, 411-423.

[2] Bentler, P. M. (1985). *Theory and implementation of EQS: A structural equations program.* Los Angeles, BMDP Statistical Software.

[3] Bentler, P. M. (1992). *EQS: Structural equation program manual.* Los Angeles, BMDP Statistical Software.

[4] Bentler, P. M., and Bonett, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin,* **88**, 588-600.

[5] Bentler, P. M., and Weeks, D. G. (1980). Multivariate analysis with latent variables. In P. R. Krishnaiah and L. Kanal, eds., *Handbook of Statistics.* Vol. 2. Amsterdam: North-Holland, pp. 747-771.

[6] Blalock, H. M. (1961). Correlation and causality: The multivariate case. *Social Forces,* **39**, 246-251.

[7] Blalock, H. M. (1962). Four-variable causal models and partial correlations. *American Journal of Sociology,* **68**, 182-1941.

[8] Blalock, H. M. (1964). *Causal inferences in nonexperimental research.* Chapel Hill: University of North Carolina.

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

[10] Boudon, R. (1965). A method of linear causal analysis: Dependence analysis. *American Sociological Review,* **30**, 365-374.

[11] Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows.* Thousand Oaks, CA: Sage.

[12] Cliff, N. (1983). Some cautions concerning the application of causal modeling methods. *Multivariate Behavioral Research,* **18**, 115-126.

[13] Duncan, O. D. (1966). Path analysis: Sociological examples. *American Journal of Sociology,* **72**, 1-16.

[14] Freedman, D. (1986). As others see us: A case study in path analysis. *Journal of Educational Statistics,* **12**, 101-128.

[15] Freedman, D. (1993). From association to causation via regression. *Talk presented at the Notre Dame Conference on Causality in Crisis.*

[16] French, J. W., Gabel, D. L., and Price, L. A. (1963). *Kit of reference tests for cognitive factors.* Princeton, NJ: Educational Testing Service.

[17] Gerbing, D. W., and Anderson, J. C. (1987). Improper solutions in the analysis of covariance structures: Their interpretability and a comparison of alternate respecifications. *Psychometrika,* **52**, 99-111.

[18] Hembree, R. (1992). Experiments and relational studies in problem solving: A meta-analysis. *Journal for Research in Mathematics Education,* **21**, 33-46.

[19] Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger and O. D. Duncan, eds., *Structural Equation Models in the Social Sciences.* New York: Academic Press, pp. 85-112.

[20] Jöreskog, K. G., and Sörbom, D. (1986). *LISREL VI: Analysis of linear structural relationships by maximum likelihood and least square methods.* Mooresville, IN: Scientific Software, Inc.

[21] Jöreskog, K. G., and Sörbom, D. (1989). *LISREL 7: A guide to the program and applications.* Chicago: SPSS, Inc.

[22] Keesling, J. W. (1972). *Maximum likelihood approaches to causal analysis.* Ph.D. dissertation. Department of Education: University of Chicago.

[23] Malpass, J. R. (1994). *A structural model of self-efficacy, goal orientation, worry, self=regulated learning, and mathematics achievement.* Ph.D. dissertation. School of Education: University of Southern California.

[24] Mayer, R. E. (1990). Problem solving. In M. W. Eysenck, ed., *The Blackwell dictionary of cognitive psychology.* Oxford, UK: Basil Blackwell, pp. 284-288.

[25] Mayer, R. E. (1993). Illustrations that instruct. In R. Glaser, ed., *Advances in instructional psychology.* Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 253-284.

[26] McArdle, J. J., and McDonald, R. P. (1984). Some algebraic properties of the reticular action model for moment structures. *British Journal of Mathematical and Statistical Psychology,* **37**, 234-251.

[27] O'Neil, H. F., Jr., and Abedi, J. (1994). *Development and validation of state metacognitive scales.* Unpublished manuscript.

[28] Simon, H. A. (1954). Spurious correlation: A causal interpretation. *Journal of the America Statistical Association,* **49**, 467-479.

[29] Simon, H. A. (1957). *Models of man.* New York: Wiley.

[30] Tanaka, J. S., and Huba, G. J. (1985). A fit index for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology,* **38**, 197-201.

[31] Van Driel, O. P. (1978). On various causes of improper solutions of maximum likelihood factor analysis. *Psychometrika,* **43**, 225-243.

[32] Wiley, D. E. (1973). The identification problem for structural equation models with unmeasured variables. In A. S. Goldberger and O. D. Duncan, eds., *Structural Equation Models in the Social Sciences.* New York: Academic Press, pp. 69-83.

[33] Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research,* **20**, 557-585.

Appendix A

EQS Program for the CFA Model

```
/title
 thesis: confirmatory factor analysis
/specifications
 case=113;var=7; me=ml; analysis=covariance; matrix=correlation;
/labels
 v1=worry1; v2=worry2;
 v3=visual1; v4=visual2;
 v5=selfeff; v6=metacog;v7=heuristic;
 f1=worry; f2=visual; f3=psability;
/equations
 v1= f1+e1;
 v2=*f1+e2;
 v3= f2+e3;
 v4=*f2+e4;
 v5=f3+e5;
 v6=*f3+e6;
 v7=*f3+e7;
/variances
 f1=1.0*;
 f2=1.0*;
 f3=1.0*;
 e1 to e7=*;
/cov
 f1, f2=*;
 f1, f3=*;
 f2, f3=*;
 e7, e6=*;
 e5, e4=*;
/matrix=correlations
1.000
 .739  1.000
 .073  -.103  1.000
 .061  -.053   .718  1.000
-.335  -.403   .077   .216    1.000
-.254  -.348   .130   .139     .597  1.000
-.304  -.340   .108   .067     .516   .830  1.000
/standard deviations
```

```
1.872  2.890  1.630  2.333  3.506  6.824  9.331
/lmtest
 set=pee;
/wtest
/end
```

Appendix B

EQS program for the Path Models

```
/title
 thesis: Path Analysis
/specifications
 case=113;var=7; me=ml; analysis=covariance; matrix=correlation;
/labels
 v1=worry1; v2=worry2;
 v3=visual1; v4=visual2;
 v5=selfeff; v6=metacog;v7=heuristic;
 f1=worry; f2=visual; f3=psability;
/equations
 v1= f1+e1; v2=*f1+e2;
 v3= f2+e3; v4=*f2+e4;
 v5=f3+e5; v6=*f3+e6; v7=*f3+e7;
 f3=*f1+*f2+d1;
/variances
 e1 to e7=*;
 d1=*;
/cov
 f2, f1=*;
 e7, e6=*;
 e4, e5=*;
/matrix=correlations
1.000
 .739  1.000
 .073  -.103  1.000
 .061  -.053   .718  1.000
-.335  -.403   .077   .216    1.000
-.254  -.348   .130   .139     .597  1.000
-.304  -.340   .108   .067     .516   .830  1.000
/standard deviations
1.872  2.890  1.630  2.333  3.506  6.824  9.331
/lmtest
 set=pee;
/wtest
/end
```