SEM Isn't Just the Schoolwide Enrichment Model Anymore: Structural Equation Modeling (SEM) in Gifted Education

D. Betsy McCoach

Structural equation modeling (SEM) refers to a family of statistical techniques that explores the relationships among a set of variables. Structural equation modeling provides an extremely versatile method to model very specific hypotheses involving systems of variables, both measured and unmeasured. Researchers can use SEM to study patterns of interrelationships among variables, compare different groups to each other, study change over time, and do many other types of sophisticated analyses. This paper will present an overview of SEM, present an illustration of research using SEM, and provide suggestions for ways that this powerful technique can be used to answer a variety of research questions within the field of gifted education.

Introduction: Structural Equation Modeling

Structural equation modeling (SEM) refers to a family of techniques, including path analysis, confirmatory factor analysis, structural regression models, autoregressive models, and latent change models (Raykov & Marcoulides, 2000), that utilizes the analysis of covariances and means to explore the relationships among a set of variables and to explain maximum variance within a specified model (Kline, 1998). Structural equation modeling is extremely versatile; it places very few restrictions on the kinds of models that can be tested (Hoyle, 1995). Structural equation modeling has been hailed as "a more comprehensive and flexible approach to research design and data analysis than any other single statistical model in standard use by social and behavioral scientists" (Hoyle, p. 15).

Over the past decade, SEM has become an increasingly popular data-analytic technique, used by researchers in all fields of education and psychology. What was once considered a rather burdensome and complex technique by nonmethodologically oriented educational researchers has become a mainstay of many quantitative studies in the field of education. Several Windows-based SEM

D. Betsy McCoach is Assistant Professor in Residence in the Educational Psychology Department at the University of Connecticut, Storrs.

Journal for the Education of the Gifted. Vol. 27, No. 1, 2003, pp. 36–61.

programs, including LISREL, EQS, AMOS, Mx, and M-Plus, compete for the growing market of SEM users. The purpose of this paper is to introduce researchers and research consumers in the field of gifted education to SEM and illustrate its power for answering research questions within our field.

Advantages of SEM

SEM offers several advantages over traditional data-analytic techniques. It allows researchers to estimate the effects of theoretical or hypothetical constructs, commonly called "latent variables" (Raykov & Marcoulides, 2000). In traditional analyses, researchers must confine themselves to estimating effects for measured variables. In SEM, a number of measured variables can be used to estimate the effects of a latent variable. The analysis of latent variables is both statistically and conceptually appealing. With SEM, researchers can include latent constructs, such as hope, motivation, and creativity, in their analyses. Because SEM allows researchers to distinguish between observed and latent variables and to model both types of variables explicitly, researchers are able to test a wider variety of hypotheses than would be possible with most traditional statistical techniques (Kline, 1998). More important, SEM accounts for potential errors of measurement and allows researchers to account for measurement error explicitly (Raykov & Marcoulides, 2000). The ability to separate measurement error or "error variance" from "true variance" is one of the reasons that SEM provides such powerful analyses. In multiple regression, measurement error within the predictor variables attenuates the regression weight from the predictor variable to the dependent variable (Baron & Kenny, 1986; Campbell & Kenny, 1999). Because analyses using SEM use multiple indicators to estimate the effects of latent variables, they correct for the unreliability within the measured predictor variables and allow for more accurate estimates of the effects of the predictor on the criterion.

In addition, structural equation modeling allows researchers to specify a priori models and to assess the degree to which the data fits the specified model. SEM provides a comprehensive statistical approach for testing existing hypotheses about relations among observed and latent variables (Hoyle, 1995). In this way, SEM forces the researcher to think critically about the relationships among the variables of interest and the hypotheses being tested. Further, SEM allows researchers to test competing theoretical models to determine which model best reproduces the observed variance/covariance matrix.

Perhaps the most significant advantage of SEM is that it allows researchers to model the direct, indirect, and total effects of a system of variables. Therefore, SEM allows researchers to test for and model mediation within their models. A *mediator variable* is a "middle man," an intervening variable that explains the relationship between a predictor variable and a dependent variable (Baron & Kenny, 1986). An *indirect effect* refers to the relationship between two variables that is mediated by one or more intervening variables (Raykov & Marcoulides, 2000). "If an indirect effect does not receive proper attention, the relationship between two variables of interest may not be fully considered." (p. 7). Because mediational models allow researchers to treat a single variable as both an independent variable and a dependent variable, they provide the researcher with the opportunity to test a variety of complex models. However, the use of mediational models is not without peril. Measurement error in the mediator tends to produce an underestimate of the effect of the mediator and an overestimate of the effect of the independent variable on the dependent variable when all of the path coefficients are positive (Baron & Kenny, 1986). Luckily, using latent variable models to develop mediational models eliminates this problem, as SEM accounts for the measurement error in the mediator.

Besides using SEM models to examine the relationships of a system of variables, they can also be used to model effects across groups or growth across time. SEM models can be used to compare patterns of interrelationships across groups of people by using a procedure called *multiple-groups SEM* or *multisample SEM*. SEM models can be used to analyze means, as well as covariances. For instance, SEM models are often used to model growth over time. This procedure is commonly referred to as "growth curve analysis" or "latent curve analysis" (Duncan, Duncan, Strycker, Li, & Alpert, 1999). In addition, researchers can also use SEM techniques to model latent means, a procedure referred to as "mean structure analysis." In short,

> Although partial correlation, ANOVA [analysis of variance], and multiple regression analysis can be used to isolate putative causal variables from other variables, SEM is more flexible and comprehensive than any of these approaches, providing means of controlling not only for extraneous or confounding variables but for measurement error as well. (Hoyle, 1995, p. 10)

Basics of SEM

How easy is it to for a novice user to learn to analyze his or her data using SEM techniques? A researcher who wants to learn to use SEM will have a much easier time mastering the basics than would have been the case a decade ago. Recently, several good introductory SEM texts have introduced researchers to the field of SEM (e.g., Diamantopoulos & Siguaw, 2000; Kelloway, 1998; Kline, 1998; Maruyama, 1998; Raykov & Marcoulides, 2000; Schumacker & Lomax, 1996). Further, SEM has become exponentially more popular as readily available, user-friendly computer software programs enable researchers to conduct their own analyses within a Windows platform. In the early days of SEM, "model fitting programs usually required users to generate a lot of arcane code for each analysis, a time-consuming, tedious, and highly error prone process" (Kline, 1998, p. 6). Today, most major SEM programs are becoming increasingly slick and easy to use. The major SEM computer programs, such as LISREL, EQS, AMOS, and M-Plus, can handle most types of SEM models. Therefore, once a researcher has mastered the basics of one computer program, he or she can use that program to model a seemingly infinite array of different types of models (Maruyama, 1998).

Understanding SEM

The basic building block of any structural equation model is the variance/covariance matrix. In fact, all of the information needed to perform a SEM analysis is contained in the covariance matrix.[1] Therefore, an analyst can create and analyze a SEM without the raw data file. When a researcher publishes the covariance or correlation matrix, other interested researchers should be able to replicate his or her results using the published covariance or correlation matrix.[2] As we delve into what might seem to be a complex barrage of symbols, jargon, and numbers, keep in mind that at the heart of SEM is the covariance matrix. SEM simply represents a way to use the covariance matrix to explain complex patterns of interrelationships among variables.

Path Diagrams

Path diagrams are visual displays of structural equations and are, perhaps, the most intuitive way to conceptualize the process of developing and testing a specified model. Most causal or predictive models can be reconceptualized as a path model. For example, Figure 1 illustrates a path diagram of a multiple regression model with three predictors and a dependent variable. The curved lines

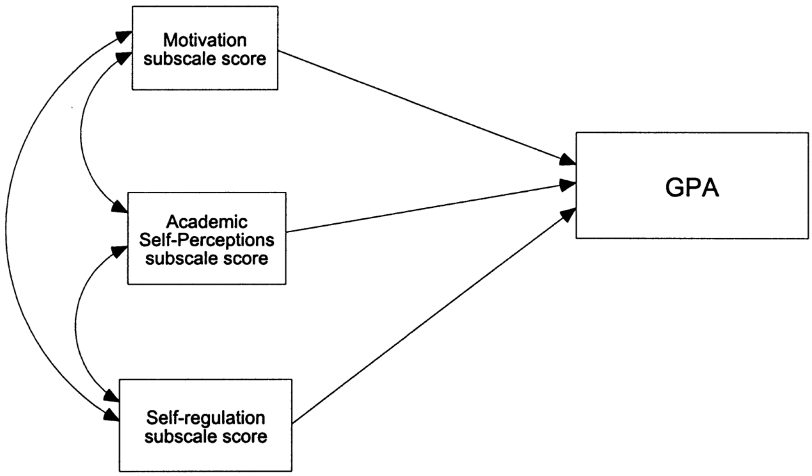

Figure 1. A multiple regression model with three predictor variables shown as a path model.

among the three predictor variables symbolize the correlations among the variables. Straight arrows connect each of the independent variables to the dependent variable. These arrows are commonly called "paths." Just as in multiple regression, these paths represent a measure of the relationship between the predictor variable and the dependent variable *after controlling for* the other variables in the model. Typically, the reported value of a path is the standardized regression coefficient. Notice that all of the variables in the model are indicated by rectangles. In a path diagram, rectangles indicate observed or measured variables. An observed variable is one that is actually measured. For example, a student's score on a test or a subscale is an observed variable.

In contrast, latent variables are the hypothetically existing constructs of interest in a study. Examples of latent variables include peace, intelligence, and apathy. Latent variables cannot be directly measured. Rather, they must be inferred or derived from the relationships among observed variables that are thought to be measures of the latent variable. In a path diagram, circles indicate latent variables. Generally, the value of a latent variable is estimated by using multiple observed variables as indicators of the latent variable. For example, if a researcher wants to use creativity as a latent variable in his or her model, he or she might use several observed variables, including scores on a divergent-thinking task, self-report measures, and peer-report measures. Creativity, the latent construct, is esti-

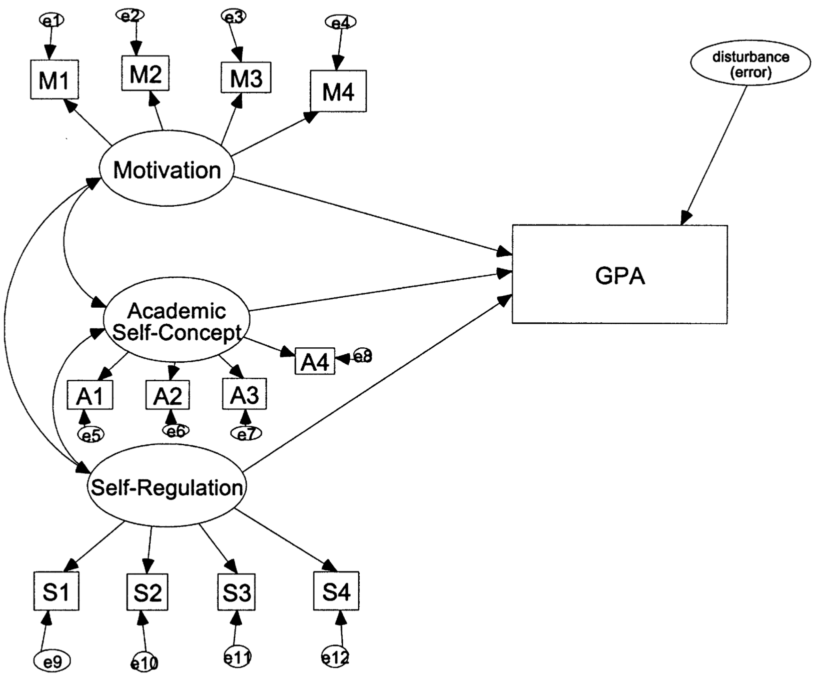

Figure 2. A latent variable structural equation model of academic achievement.

mated by a variety of observed variables, or indicators. The question of how many observed variables a researcher needs to adequately estimate a latent variable is a very complex issue. From a technical point of view, two observed variables might be adequate if there is more than one latent variable in the model. Using three observed variables is technically adequate to estimate a latent variable. However, having four observed variables is better, and having more "is gravy" (Kenny, 1979, p. 143). From a theoretical point of view, "constructs can differ widely in the extent to which the domain of related observable variables is (1) large or small and (2) specifically or loosely defined" (Nunnally & Bernstein, 1994, p. 85). In general, the more abstract and loosely defined a construct is, the more indicators will be required to measure adequately the latent variable (Nunnally & Bernstein).[3]

Figure 2 illustrates a latent variable model of achievement. Each of the three predictors—motivation, academic self-perceptions, and self-regulation—is measured using four observed variables. These observed variables could be four different subscales that each measure the latent variables. More commonly though, these observed variables are individual items or small clusters of items (sometimes referred to as "item parcels") on the subscale of an instrument. The three latent variables are being used to predict GPA. In this model, there are 13 observed variables: 4 variables on each of 3 latent variables and GPA, an observed variable.

In SEM, a variable is considered to be exogenous or endogenous. Exogenous variables remain at the outside edge of the model; they are only predictors of other variables; they are not affected by any other variables in the model. Exogenous variables may be correlated with other exogenous variables, but they never have paths (straight arrowheads) leading into them. Endogenous variables are caused or predicted by one or more variables in the model. However, endogenous variables can also affect other endogenous variables. In path analysis, the independent variables are exogenous; the dependent variable is endogenous. In addition, every endogenous variable in the model contains an error term, often referred to as a "disturbance." The disturbance or error term represents the sum of all other causes of the endogenous variable that are *not* explicitly specified in the model. The disturbance of a standardized endogenous variable is equal to the square root of 1 minus the R^2 of the endogenous variable. In Figure 2, notice that GPA, which is an endogenous variable, has a disturbance term. In addition, because each of the observed variables that is predicted by a latent variable is an endogenous variable, each has an error term or disturbance.

Building a SEM Model

Specification and Identification. A researcher who plans to use SEM first specifies a hypothesized model (i.e., draws a path diagram). At this stage, it is also important to determine whether the model is identified. For a SEM model to be identified, there must be at least as many elements in the variance/covariance as there are parameters to be estimated within the model. A parameter is an unknown characteristic of the population that we are trying to estimate in the specified SEM model. In addition, both the measurement model and the structural model must be identified in order for the entire latent variable model to be identified. However, having met this necessary condition does not ensure that the specified model is indeed identified. The issue of identification is complex, and presenting all of the rules for identification is beyond the scope of this paper. (For a thorough treatment of identification, see Kenny, Kashy, & Bolger, 1998.)

Data Collection. Next the researcher selects measures to test the hypothesized model and then collects the data. Ideally, a researcher should specify the model to be tested prior to gathering the data (Kline, 1998). A word of caution is warranted here. Some researchers seem to feel that complicated analyses, such as SEM, can help to overcome shortcomings in the data or design of a study. **Though structural equation models may look impressive, SEM cannot salvage studies that contain badly measured constructs, inappropriate samples, or faulty designs.**

Model Specification. There are two components to a structural equation model: the measurement model and the structural model. The measurement model depicts the relationships between the observed variables (also called "indicators") and their underlying latent variables (Kline, 1998). An analysis of a measurement model is commonly called a "confirmatory factor analysis." The structural model depicts the predictive paths and consists of the structural paths between and among the latent variables and any observed variables that are not indicators of an underlying variable. The SEM analyst engages in a two-step modeling process (Kline). Before analyzing the full structural model, the researcher examines the adequacy of the measurement model by conducting a confirmatory factor analysis of all of the latent variables in the structural equation model and evaluates the plausibility of the measurement model. If the fit of the measurement model is unsatisfactory, the fit of the full model will also be unsatisfactory. Therefore, any problems in the measurement model should be addressed before proceeding to the structural analysis. Next, the researcher analyzes the structural model, evaluates the model fit, makes any necessary changes to the model, and evaluates the fit of the revised model (Kline, pp. 49–50). Sometimes, a researcher may compare the fit of two or more competing models.

Model Fit. How do we know if the data "fits" the model? Hypothesis testing in SEM departs from traditional tests of significance. In most statistical analyses, researchers test the null hypothesis that there is no relationship among a set of variables or that there are no statistically significant differences among a set of variables. Generally speaking, we want to reject the null hypothesis and conclude that there are statistically significant relationships or differences. In SEM, the logic is reversed. We test the hypothesis that the population covariance of observed variables equals the covariance matrix implied by a particular model.[4] Assuming that the distributional assumptions (normality, etc.) for the data are satisfied, one can use a test statistic with a chi-square (χ^2) distribution to test the null hypothesis that the specified model exactly reproduces the population covariance matrix of observed variables (Bollen & Long, 1993). Therefore, analysts can evaluate exact model fit by comparing the chi-square (χ^2) of the specified model to the critical value for chi-square for its degrees of freedom. However, using this approach poses several problems. First, χ^2 is very sensitive to sample size; therefore, almost any model with a large sample size will be rejected if there is even a miniscule amount of data misfit.[5] On the other hand, because of the estimation method used, it is important to have large sample sizes. Because a researcher wants to accept the null hypothesis, having large sample sizes works against the researcher. He or she inevitably must reject the null hypothesis that the model fits the data. Second, knowing that the model-implied covariance matrix does not exactly fit the population covariance matrix does not tell us about the degree to which the model does or does not fit the data. In scientific inquiry, we generally reward parsimony and simplicity. Generally, our models are simplifications of reality. We try to capture the essence of a system without completely recreating it. Therefore, it should come as no surprise that model-implied covariance matrices fail to reproduce population covariance matrices exactly.

Researchers wanted to be able to talk about the degree to which the model fits the data. Therefore, researchers have developed many fit indices (i.e., comparative fit index, Tucker Lewis Index, root mean square error of approximation, etc.) that provide an estimate of model-data fit (or misfit). These fit indices attempt to correct the problems that result from judging the fit of a model solely by examining the chi-square of the model. SEM programs provide many measures of model fit.

There are two basic types of fit indices: absolute fit indices and incremental fit indices. Absolute fit indices evaluate the degree to which the specified model reproduces the sample data. Some of the more commonly used absolute fit indices include the root mean square error of approximation (RMSEA) and the standardized roo mean square residual (SRMR). The RMSEA is a function of the degrees of freedom in the model (Browne & Cudeck, 1993), the χ^2 of the model, and the sample size. The RMSEA has become one of the most popular fit indices because, unlike the χ^2, the value of the RMSEA should not be influenced by the sample size (Raykov & Marcoulides, 2000). The SRMR represents a standardized summary measure of the model-implied covariance residuals. Covariance residuals are the differences between the observed covariances and the model-implied covariances (Kline, 1998). "As the average discrepancy between the observed and the predicted covariances increases, so does the value of the SRMR" (Kline, p. 129). The RMSEA and the SRMR approach zero as the fit of the model nears perfection. Hu and Bentler (1999) have suggested that SRMR values of approximately .08 or below and values of approximately .06 or below for the RMSEA indicate that there is a relatively good fit between the hypothesized model and the data.

Incremental fit indices measure the proportionate amount of improvement in fit when the specified model is compared with a nested baseline model (Hu & Bentler, 1998). Some of the most commonly used incremental fit indices include the nonnormed fit index (NNFI), also known as the Tucker Lewis Index (TLI); the comparative fit index (CFI); and the relative noncentrality index (RNI). These three indices approach 1.00 as the model-data fit improves, and the TLI can actually be greater than 1.00 when the fit of the data to the model is close to perfect. Generally speaking, TLI, CFI, and RNI values at or above .95 indicate that there is a relatively good fit between the hypothesized model and the data (Hu & Bentler, 1995, 1999). TLI, CFI, and RNI values below .90 generally indicate that the fit of the model to the data is less than satisfactory.

Many factors, such as sample size, model complexity, and the number of indicators, can affect fit indices differentially (Gribbons & Hocevar, 1998); therefore, a researcher should examine more than one measure of fit when evaluating a SEM model. Several previously popular fit indices, such as the goodness of fit index (GFI) and the normed fit index (NFI), have fallen out of favor as they have been shown to be unduly influenced by such factors as the number of observed variables in the structural equation model (Gerbing & Anderson, 1993). Because the vast array of fit indices can be overwhelming, most researchers focus on and report only a few. The most popular of the fit indices at the present time seem to be the RMSEA, the SRMR, and the CFI.

Respecification. If the data do not fit the model, how should the researcher proceed? Sometimes, a researcher might wish to change certain aspects of the model, a process called "respecification." SEM models can be time consuming. Unlike traditional statistical techniques, SEM usually requires running and evaluating several models before adopting a final model. First, the researcher checks to see whether all of the specified paths are statistically significant. As in multiple regression, each unstandardized path coefficient is divided by its standard error to compute a critical ratio. If the absolute value of this ratio is greater than or equal to 1.96, the path is generally considered statistically significant. If the ratio of the unstandardized path coefficient to its standard error is less than |1.96|, the path is considered nonsignificant. Generally, nonsignificant paths can be deleted from a model without affecting the fit or the predictive power of the model.

Theorists begin by specifying an a priori model based on previous literature and substantive hypotheses. Because theorists seek the most parsimonious explanation for a given phenomenon, analysts delete or trim nonsignificant paths. They can then test the fit of the new more parsimonious model (with greater degrees of freedom) against the original model using the χ^2 difference test. For the χ^2 difference test, we subtract χ^2 of the simpler model (χ^2_2) from the χ^2 of the more complex model (χ^2_1).[6] We then subtract the degrees of freedom of the less parsimonious model (df_2) from the degrees of freedom for the more parsimonious model (df_1). We compare this χ^2 difference (χ^2_1 - χ^2_2) to the critical value of χ^2 with df_1 - df_2 degrees of freedom. If this value is greater than the critical value of χ^2 with df_1 - df_2 degrees of freedom, we conclude that deleting the paths in question has significantly worsened the fit of the model. If the value of χ^2_2 - χ^2_1 is less than the critical value of χ^2 with df_1 - df_2 degrees of freedom, then we conclude that deleting the paths has not significantly worsened the fit of the model. When deleting paths does not significantly worsen the fit of the model, we choose the more parsimonious model (the one that has fewer paths and more degrees of freedom) as the better model.

The χ^2 difference test merits further discussion. First, the χ^2 difference test can be used to compare any two hierarchically nested models. Two models are hierarchical (or nested) models if one model is a subset of the other. For example, if a path is removed or added between two variables, the two models are hierarchical (or nested) models (Kline, 1998). However, if a new variable is added or removed, the models are not hierarchical models. Second, the χ^2 difference test can only be used to compare hierarchically related models. It is inappropriate to use the χ^2 difference test to compare nonhierarchical models. Third, because χ^2 is affected by sample size, the χ^2 difference test will also be affected by sample size. Therefore, it will be much easier to find a significant χ^2 difference between two hierarchical models with a large sample than it will be with a small sample. Therefore, any results should be viewed as a function of the power of the test, as well as a test of the competing models.

In addition, the SEM output usually includes modification indices (sometimes called "Lagrange Multiplier tests"). The modification indices suggest which parameters might be added to the model to improve model fit. It is tempting to use these modification indices to make changes to improve the fit of the model. Proceed cautiously! Respecification of SEM models should be guided by theory, not simply by a desire to improve measures of model fit. Analysts may consider making model modifications that are conceptually consistent with the research hypotheses. Other suggested model modifications may make no conceptual sense. Sometimes the modifications suggested by the SEM program are downright illogical and indefensible. For example, the modification index might suggest that a measure might cause a latent variable. A good analyst uses modification indices very cautiously. Because SEM models are so open to modification and manipulation, SEM allows for a great deal of artistic license on the part of the analyst. Structural equation modeling is as much an art as a science. It is this freedom that makes SEM so powerful and so appealing, but also so prone to misuse.

When SEM Gives You Inadmissible Results

In addition to examining the parameter estimates, tests of significance, and fit indices, it is very important to examine several other areas of the output to ensure that the program ran correctly. The variances of the exogenous variables should be positive and statistically significant. The variances of the error terms and the disturbances should also be positive and statistically significant. As in multiple regression, the standardized path coefficients should be between –1.00 and +1.00. Further, the standardized error terms and disturbances should fall in the range of 0.00 to 1.00. Negative error variances and standardized regression weights above 1 are called "Heywood cases," and they indicate the presence of an *inadmissible solution*. Heywood cases can be caused by specification errors, outliers that distort the solution, a combination of small sample sizes and only one or two indicators per factor, or extremely high or low population correlations that will result in empirical underidentification (Kline, 1998). When any of these problems occur, the SEM output cannot be trusted. The analyst should try to find the cause of the Heywood case, respecify the model to fix the problem, and run the data again. It is never advisable to interpret output that contains any Heywood cases or inadmissible solutions.

Another possible problem is that the SEM program will run out of iterations before it finds a maximum likelihood solution that minimizes the distance between the observed and model-implied covariance matrices. This problem is known as *lack of convergence*. When this happens, the output should not be trusted. Large or infinite numbers of iterations can be signs of a problem, such as an underidentified model, an empirically underidentified model, bad start values, extreme multicollinearity, a tolerance value approaching zero, or other specification error (Kline, 1998). Some computer programs are set for a maximum number of iterations. Once the program reaches this limit, it produces output based on the last iteration. Again, do not trust the output. If the program fails to converge, it is necessary to inspect the output for possible errors or clues to the reason for the nonconvergence, but analysts should not interpret output if the computer fails to converge upon a proper solution.

Assumptions and Requirements of SEM

Normality. Many of the assumptions of SEM are similar to the assumptions of multiple linear regression. Namely, SEM techniques assume that the variables of interest are drawn from a multivariate normal population (Kaplan, 2000). Maximum likelihood estimation performs optimally when the data are continuous and normally distributed (Kaplan). Much has been written on the effects of violating the assumption of normality. Generally, SEM is fairly robust to small violations of the normality assumption; however, extreme nonnormality can cause problems. (For more information about dealing with nonnormal data, see Curran, West, & Finch, 1996, and West, Finch, & Curran, 1995.)

Linearity. As in multiple regression, SEM assumes that the variables of interest are linearly related to each other. In addition, there are specialized techniques to examine nonlinear effects using SEM; however, reviewing these techniques is beyond the scope of this paper. Interested readers should consult Schumacker and Marcoulides (1998) for a detailed treatment of interaction and nonlinear effects in structural equation modeling.

Sampling. Maximum likelihood estimation assumes that the data represent a simple random sample from the population. In reality, this is rarely the case (Kaplan, 2000). The effects of nonindependence of observations can bias the results of the analysis. New techniques are becoming available to combine multilevel modeling and structural equation modeling techniques to use SEM to analyze data that has been collected using multistage or cluster sampling techniques (Kaplan). (For more information about multilevel structural equation modeling, see Heck & Thomas, 2000; Kaplan, 2000; and Marcoulides & Schumacker, 2001).

Because SEM uses maximum likelihood estimation[7] to minimize the discrepancy between the observed covariance matrix and the model-implied covariance matrix, SEM is a large-sample technique. Although there are no definitive rules for a minimum sample size, there are several rules of thumb. Generally speaking, under most circumstances, sample sizes below 100 are considered too small to use SEM techniques (Kline, 1998, Schumacker & Lomax, 1996). Schumacker and Lomax's examination of the published SEM literature revealed that many SEM articles used sample sizes of 250–500. Generally speaking, sample sizes of 200 or more are considered sufficient for estimating most types of SEM models, especially if the variables are normally distributed and obtained from a random sample of subjects. However, very complex models may require larger sample sizes. As the ratio of the number of cases to the number of parameters declines, the estimates generated by SEM become more unstable (Kline, 1998). Kline recommended having at least 10 cases for each estimated parameter.

Range of Values. Because SEM is essentially a correlational technique, anything that affects the magnitudes of the covariances among the variables in the model will impact the SEM analysis. For example, restriction of range in one variable will attenuate the covariance between that variable and any other variables in the model. This will result in small path coefficients leading to and from that variable. Therefore, researchers must exercise caution when framing research questions. For example, a researcher in the field of gifted education might wish to study the relationship between intelligence and certain personality characteristics. Using any correlational techniques (including SEM) with a sample of students with high measured intelligence (say, IQ > 130) is likely to result in the conclusion that intelligence does not relate to the personality characteristics of interest simply because there is inadequate variability in the predictor variable (IQ). In addition, researchers should be cautious about using such techniques as mean substitution as an imputation technique with SEM, as mean substitution will decrease the variability within any variables that have missing data. If the amount of missing data is substantial, using mean substitution techniques could lead to erroneous conclusions. Luckily, some SEM programs (such as AMOS) allow for missing data.

Common Mistakes and Misunderstandings in SEM

Correlation Versus Causation

The purpose of path analysis is to determine if the causal inferences posited by a researcher are consistent with the data (Bollen, 1989). Therefore, SEM techniques inform us about the degree of model data fit (or misfit). However, knowing that a specified model is consistent with the data does not prove that the model is "correct." There are several reasons that it is inappropriate to view a good-fitting structural equation model as correct or indicative of causality. First, for every SEM that a researcher specifies, there are other equivalent models that will result in the sample chi-square and fit indices (e.g., specifying that X > Y > Z is equivalent to specifying that Z < Y < X). For complex models, there are often several (if not dozens!) of functionally equivalent models that the researcher has not tested. Therefore, the model that is tested provides good fit to the data; however, many other equivalent models will provide equally good fit to the data; and, moreover, the possibility even exists that an untested model will provide even better fit to the data.

To use correlational techniques to infer that X causes Y, a researcher must meet three criteria (Kline, 1998). First, X must precede Y temporally. In other words, if the two variables are measured at the same time, it is impossible to infer causality from SEM models. Therefore, it is impossible to infer causality from any cross-sectional SEM models. Second, the direction of the causal relation must be correctly specified. The measurement of X before Y is necessary but not sufficient to prove a causal relation. For example, if high academic motivation results in high academic achievement, but the measures of academic achievement are taken prior to the measures of motivation, one could erroneously conclude that high academic achievement causes high academic motivation. In practice, it is often difficult or impossible to meet criterion 2, as it is often difficult or impossible to develop theories and gather data in a way that allows a researcher to know that the direction of the causal relation is correctly specified. Finally, the relationship between the causal variable and the criterion variable must not vanish when external causes, such as common causes of both variables, are *partialed* out (Kline, 1998). This is often referred to as the "omitted variable problem." A researcher may conclude that X causes Y when in reality, Z, a variable that was not included in the model, causes both X and Y. For example, a researcher may erroneously conclude that having high academic self-concept causes high academic achievement, when in fact, having high academic ability causes

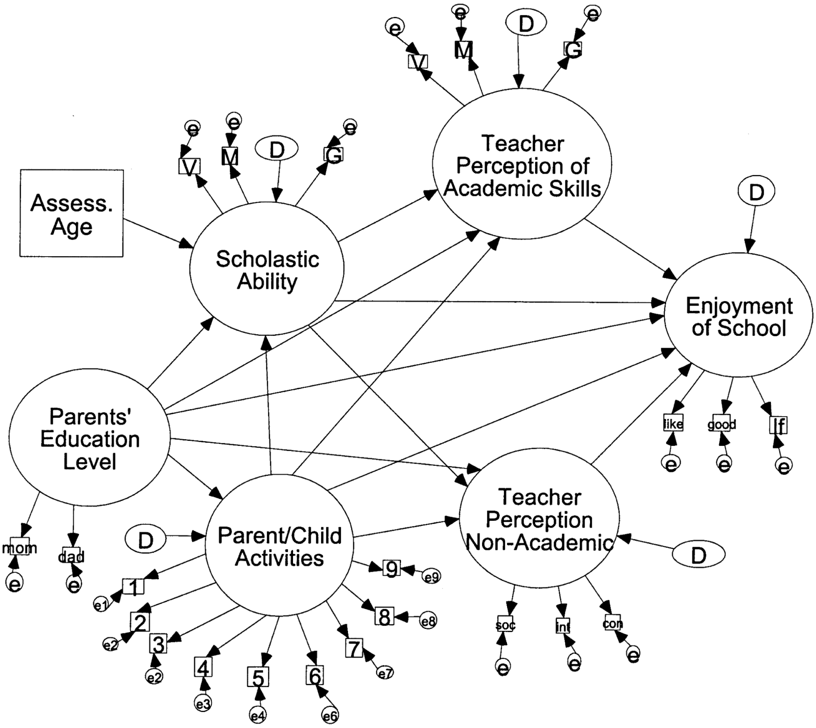

Figure 3. Full originally specified structural equation model for ECLS-K example.

both high academic self-concept and high academic achievement. Satisfying these three criteria is extremely difficult. Therefore, the ability to infer causality using SEM represents the very rare exception, rather than the rule.

An Example Using SEM

Figure 3 shows a hypothesized SEM model of kindergarten students' educational experiences. The data for this model came from the Early Childhood Longitudinal Study–Kindergarten Cohort (ECLS-K), a federally sponsored longitudinal study of almost 20,000 school children who were kindergarteners during the 1998–1999 school year. For this illustration, I randomly chose a subsample of the students who were first-time kindergarteners and who were not missing data on any of the key variables. The sample size for this analysis was 3,853 cases. Technically, because students were nested within schools, this type of analysis is best approached using multilevel modeling or multilevel structural equation modeling techniques. Since this example is being presented to illustrate the basic components of a structural equation model, we will ignore this issue. However, remember that, when using nonrandom samples, nonindependence of data must be carefully considered.

In this model, there are six latent variables. Parents' education level is a latent variable that was estimated using two observed variables: mother's education level and father's education level. Scholastic ability is a latent variable that was estimated using three observed variables: the child's scale scores on the reading, math, and general knowledge tests given at the beginning of the kindergarten year. The parent/child activities latent variable consists of nine questions in which parents reported the frequency with which they engaged in activities with their children, such as reading books, playing games, and singing songs. The teacher perception of academic skills latent variable was estimated using three observed variables: the teacher's rating of the student's language and literacy skills, the teacher's rating of the student's math skills, and the teacher's perception of the student's general knowledge. The teacher perception of nonacademic skills latent variable was estimated using three observed variables: the teacher's rating of the student's sociability and extroversion, the teacher's rating of the student's self-control, and the teacher's rating of the student's interpersonal skills. The enjoyment of school latent variable was estimated using three observed variables. Each parent reported the degree to which his or her child liked school, looked forward to going to school, and felt good about school. The model also contained one manifest (observed) variable: the student's age at the beginning of kindergarten.

Using the two-step hypothesis procedure, I first estimated the measurement model using EQS 5.7 (Bentler, 1998). The measurement model specified that each indicator loaded on only one of the 6 latent variables and allowed the 6 factors to correlate with one another. Figure 4 illustrates the initial specification of the measurement component of the model. The χ^2 of the measurement model was 2,184.64 with 215 degrees of freedom. The χ^2 was significant, but, given the large sample size, it would be virtually impossible to obtain a nonsignificant χ^2. The overall fit of the measurement model was adequate (NNFI/TLI = .923, CFI = .934; RMSEA = .049; SRMR = .051).

Next, I estimated the parameters for the entire latent variable model. Figure 4 illustrates the initial specification of the structural model. The χ^2 of the initial model was 2,435.53 with 238 degrees of freedom. The

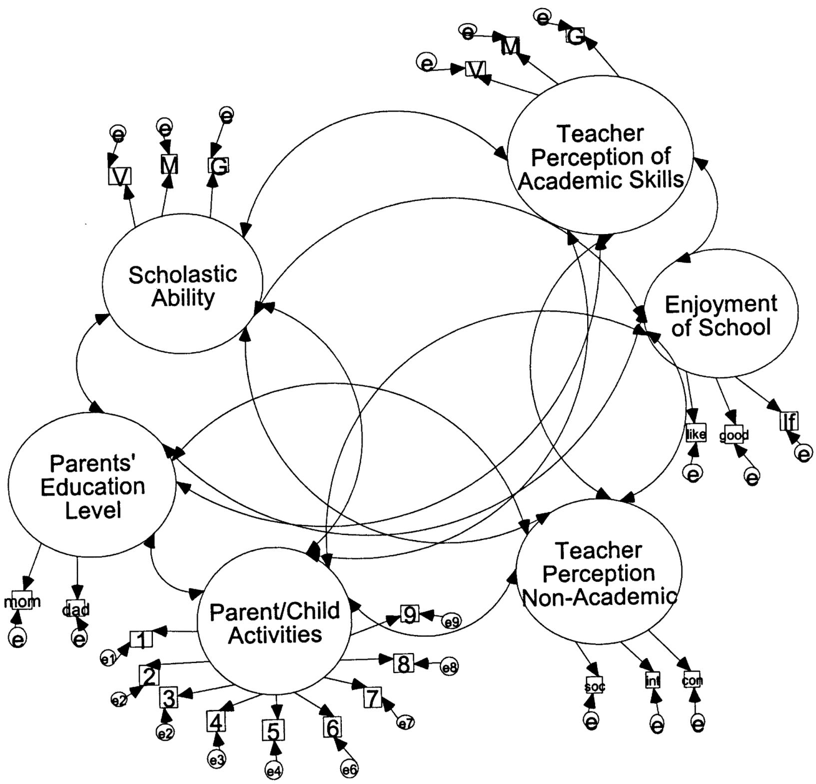

Figure 4. Specifications for measurement model of ECLS-K data.

χ^2 was significant, but, given the large sample size, it would be virtually impossible to obtain a nonsignificant χ^2. The overall fit of the initial model was adequate (BBNI/TLI = .924, CFI = .934; RMSEA = .049; SRMR = .046). However, several of the originally specified paths were nonsignificant. Therefore, I eliminated these paths from the model and reestimated the model. The new model had a χ^2 of 2,443.53 with 243 degrees of freedom. Because these models are nested, we can compare them using the χ^2 difference test. The χ^2 difference between the two models was 7.7 with 5 degrees of freedom. This suggests that dropping the five statistically insignificant paths did not significantly worsen the fit of the model. Furthermore, eliminating paths increases the parsimony of the model. Therefore, the second simpler model is determined to be the preferred model. Figure 5 illustrates the structural component of the final model. The fit indices for the final model indicate that it exhibits adequate fit to the observed data. The NNFI/TLI is .925, the CFI is .934, the RMSEA is .048, and the SRMR is .046.

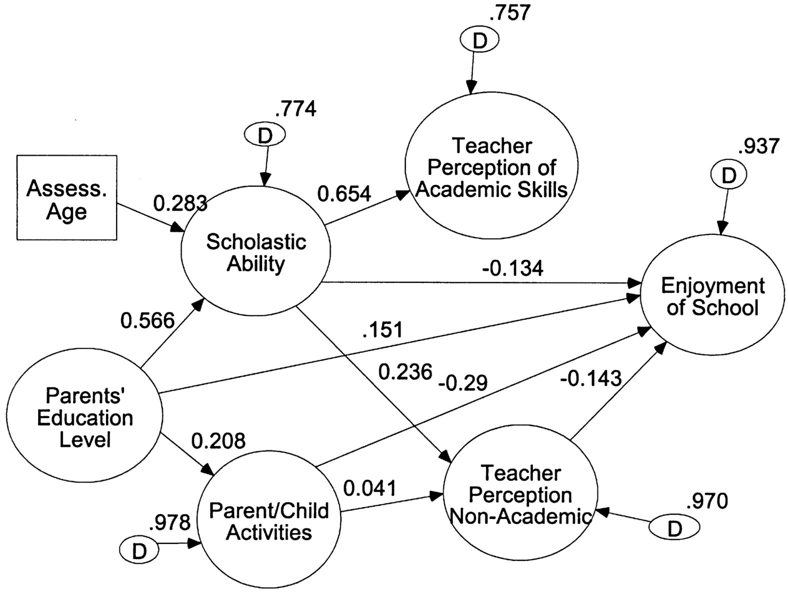

Figure 5. Final structural model of ECLS-K example.

Note. This figure shows the structural model only.

Parent's education level (b = .566) and assessment age (b = .283) are both positively related to students' scholastic ability at the beginning of kindergarten. The path coefficients (standardized regression weights) in a path diagram are interpreted similarly to the beta weights in a multiple regression. As a general rule of thumb, path coefficients in the .10–.20 range indicate a small effect, path coefficients in the .30–.40 range indicate a medium effect, and path coefficients larger than .50 indicate a large effect (Kline, 1998). Therefore, after controlling for parents' education level, a child's age at kindergarten entry does have a small direct effect on his or her scholastic ability. The path from parents' education to scholastic ability is quite large. Therefore, after controlling for a child's assessment age, the parents' level of education has a large direct effect on the student's scholastic ability at the beginning of kindergarten. In fact, the combination of parent's education level and student's assessment age explains more than 40% of the variance in a student's cognitive ability at the beginning of kindergarten. Perhaps surprisingly, after controlling for parents' education level and student's age at kindergarten entry, the parent/child activity factor is not a significant predictor of the child's scholastic ability.

How does parental education level impact the teacher's perception of a student's academic skills? First, we see that a student's ability has a large direct effect on the teacher's perceptions of the student's academic skills. None of the other variables in the model directly affects teacher's perceptions of a student's academic skills. After controlling for a student's scholastic ability, parental education has no direct influence on teacher's perceptions of a student's academic skills. However, parental education level does exert an *indirect* effect on a teacher's perceptions of the student's academic skills. An indirect effect between two variables occurs when no single straight line or arrow directly connects them, but when the first variable may be reached through one or more variables via their paths (Schumacker & Lomax, 1996). Parent education level predicts the child's scholastic ability, which in turn predicts the teacher's rating of the child. The effect of parent's education on a teacher's rating is completely mediated by the scholastic ability variable. We can estimate the indirect effect of parental education on teacher's perceptions of a student's academic skills by multiplying the two path coefficients: .566 multiplied by .654 equals .37. Therefore, there is a medium-sized indirect effect of parental education on teacher's perceptions of a student's academic skills; however, this effect is completely mediated by the student's scholastic ability at kindergarten entry.

Let's turn to the most complex network of direct and indirect effects: those for the student enjoyment of school latent variable. Several variables exert direct and indirect influences on a child's enjoyment of school. The direct effect of a student's scholastic ability on his or her enjoyment of school is -.134. This suggests that, after controlling for the other variables in the model, the higher a student's scholastic ability, the lower his or her enjoyment of school will be. However, -.134 is a small effect. In addition, after controlling for the other variables in the model, the parent/child activities factor is negatively related to student's enjoyment of school. Therefore, after controlling for the other variables in the model, the greater the parent/child activity level, the less the parent reports the child enjoys school. Again, this is a relatively small direct effect. After controlling for the other variables in the model, teacher's ratings of a student's nonacademic skills is negatively related to the student's enjoyment of school. Again, this is a small direct effect (b = -.143). Finally, after controlling for the other variables in the model, parents' education level is positively related to students' enjoyment of school; however, this is again a small direct effect (b = .151).

Several variables in the model exert indirect effects on students' enjoyment of school. To determine the indirect effects of variables, we must trace all continuous lines of arrows on the path diagram from any exogenous or predictor variables to the enjoyment of school factor, making sure to follow the direction of the arrows. To compute the indirect effect through a particular pathway, we take the product of all of the path coefficients in the series from the predictor variable to the criterion variable. To compute the total indirect effect of a variable on another variable, we sum the products from all possible tracings of indirect effects from the predictor variable to the criterion variable.

For example, the indirect effect of parents' education on a student's enjoyment of school can be traced through the scholastic ability factor (.566 * -.134 = -.076), the parent activities factor (.208 * -.291 = -.060), the parent activities factor to the nonacademic factor (.208 * .041 * -.143 = -.001), and the scholastic ability factor to the nonacademic factor (.566 * .236 * -.143 = -.019). Thus, the indirect effect of parents' education level on a student's enjoyment of school is (-.076 + -.060 + -.001 + -.019 = -.156).

To compute the total effect of a predictor variable on a criterion variable, we compute the sum of all indirect effects (in this case -.156) and add the total of the indirect effects to the direct effect of the variable. The direct effect of parent's education level on student's enjoyment of school is .151; the indirect effect through all the other variables in the model is -.156. Therefore, the total effect of parent's education level on student's enjoyment of school is -.005. This value is very close to 0. Therefore, we can conclude that parent's education level really does not influence student's enjoyment of school after controlling for all of the other variables in the model.

It is quite easy to determine the percentage of variance in a given variable that is explained by the model. The square of the disturbance (or error) represents the percentage of unexplained variance in the model. Therefore, R^2 is simply 1 minus the square of the disturbance term. (Occasionally, the disturbance is presented as a path coefficient. In this case, R^2 equals 1 minus the disturbance path coefficient.) The percentage of variance in student's school enjoyment that is explained by the entire system of variables within the model is 12.3%. This is equal to $1 - .937^2$. Fortunately, the output provided in SEM programs includes the computed direct, indirect, and total effects and percentage of variance explained by the model for each of the variables in a given structural equation model.

Research Questions That Can Be Answered With SEM

Clearly, given the versatility and richness of SEM, researchers can utilize this technique to answer a broad array of research questions within the field of gifted education. They can use SEM techniques to systematically study interrelationships among a variety of factors. SEM can be used to analyze cross-sectional or longitudinal studies. SEM can also be used to analyze experimental studies. In addition, SEM is an extremely powerful technique for analyzing nonexperimental longitudinal data or for modeling students' growth over time.

In the field of gifted education, many possible mediational models abound. SEM provides a way to determine the direct, indirect, and total effects of multiple predictor variables on multiple criterion variables. In this way, we can begin to paint richer and more complex pictures of the patterns among such variables as creativity, intelligence, motivation, and achievement. For example, what is the relationship among creativity, IQ, and scholastic achievement and adult achievement? It seems reasonable to hypothesize that academic achievement mediates the relationship between IQ and later success in life.

In addition, researchers can use SEM to analyze two-step experimental or intervention procedures. For example, a researcher who studies gifted underachievers might hypothesize that gifted underachievers suffer from low self-efficacy and that developing interventions to increase their self-efficacy will lead to increased achievement. Using ANOVA techniques to determine whether the intervention leads to increased academic achievement ignores an important piece of the study. Perhaps the intervention does increase students' self-efficacy; however, the increase in self-efficacy does not translate into increased achievement. Or, perhaps, the opposite result occurs. Perhaps the intervention increases students' achievement, but not their self-efficacy. SEM techniques allow the researcher to examine the mediational effects of increasing self-efficacy on increasing underachievers' academic achievement.

Several specialized SEM techniques further expand the variety of research questions that can be answered using SEM. For instance, multisample or multiple-groups SEM analysis involves the specification of a theoretical model with more than one sample simultaneously. Multiple-groups analysis allows the researcher to compare the patterns of interrelationships among latent variables across multiple samples. Researchers can use this approach to analyze experimental, cross-sectional, or longitudinal data (Schumacker & Lomax, 1996). This approach is particularly well-suited for researchers within the field of gifted education. Often, researchers conduct studies comparing gifted students to nongifted students on a number of key variables. Multiple-groups SEM allows researchers to determine whether the pattern of interrelationships among those variables are similar in the gifted and nongifted groups. For example, I recently examined the relationship between academic self-perceptions and academic achievement among a convenient sample of gifted high school students and a convenient sample from a general population of high school students. I found that, although gifted students exhibited higher mean academic self-perceptions and higher academic achievement than the general population of high school students, the relationship between academic self-perceptions and academic achievement was similar for both groups of students (McCoach & Siegle, 2003). Traditionally, our field has examined the differences between gifted and nongifted students in terms of mean differences. However, exploring the differences in the relationships among key variables is a fertile and unexamined area for future research.

Finally, until this point, we have talked only about modeling the covariances among variables. However, structural equation models can also include means. A mean-structure analysis includes the variance/covariance matrix, as well as the means of the observed variables. Using mean structure analysis allows researchers to model and test hypotheses about the means of latent variables (Kline, 1998). These models provide a flexible alternative to traditional analysis of variance models (Raykov & Marcoulides, 2000).

Latent change (or latent growth) models are a special class of mean structure models that allow researchers to model growth and change over time (Duncan et al., 1999). Using latent change analysis, researchers can describe the growth patterns of gifted children. In addition, researchers can compare the growth of gifted, nongifted, or both groups of children on a variety of factors. Do children who come to kindergarten reading exhibit faster or slower rates of reading growth than students who cannot read when they enter kindergarten? Do gifted children really learn new material in a given subject area at a faster pace than other children? Do different forms of instruction result in increased rates of learning for gifted students? These are a few of the many questions that can be explored using latent change analysis.

Structural equation modeling provides an extremely versatile method to model very specific hypotheses involving systems of latent variables. Researchers can use SEM to study patterns of interrelationships among variables, to compare different groups to each other, to model latent means, to study change over time, and to do many other types of sophisticated analyses. It is my hope that structural equation modeling techniques will come to play a larger role in research within the field of gifted education as researchers realize the power and flexibility this method of analysis offers.

References

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Psychological Bulletin, 51*, 1173–1182.

Bentler, P. M. (1998). *EQS structural equations model program* (Version 5.7) [Computer software]. Encino, CA: Multivariate Software.

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

Bollen, K. A., & Long, J. S. (1993). *Testing structural equation models*. New York: Wiley.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Thousand Oaks, CA: Sage.

Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artifacts*. New York: Guilford.

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

Diamantopoulos, A., & Siguaw, J. A. (2000). *Introducing Lisrel*. Thousand Oaks, CA: Sage.

Duncan, T. E., Duncan, S. C., Strycker, L. A., Li, F., & Alpert, A. (1999). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications*. Mahwah, NJ: Erlbaum.

Gerbing, D. W., & Anderson, J. C. (1993). Monte Carlo evaluations of goodness of fit indices for structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 40–65). Thousand Oaks, CA: Sage.

Gribbons, B. C., & Hocevar, D. (1998). Levels of aggregation in higher level confirmatory factor analysis: Applications for academic self-concept. *Structural Equation Modeling, 5*, 377–390.

Heck, R. H., & Thomas, S. L. (2000). *An introduction to multilevel modeling techniques*. Mahwah, NJ: Erlbaum.

Hoyle, R. H. (1995). The structural equation modeling approach: Basic concepts and fundamental issues. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 158–176). Thousand Oaks, CA: Sage.

Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation models. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 158–176). Thousand Oaks, CA: Sage.

Hu, L., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76–99). Thousand Oaks, CA: Sage.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to unparameterized model misspecification. *Psychological Methods, 3*, 424–453.

Hu, L., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.

Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions*. Thousand Oaks, CA: Sage.

Kelloway, E. K. (1998). *Using LISREL for structural equation modeling: A researcher's guide*. Thousand Oaks, CA: Sage

Kenny, D. A. (1979). *Correlation and causality*. New York: Wiley.

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., pp. 252–258). Boston: McGraw-Hill.

Kline, R. B. (1998). *Principles and practice of structural equation modeling.* New York: Guilford.

Loehlin, J. C. (1998). *Latent variable models. Factor, path, and structural analysis* (3rd ed.). Mahwah, NJ: Erlbaum.

Marcoulides, G. A., & Schumacker, R. E. (1996). *Advanced structural equation modeling: Issues and techniques*. Mahwah, NJ: Erlbaum.

Marcoulides, G. A., & Schumacker, R. E. (2001). *New developments in structural equation modeling*. Mahwah, NJ: Erlbaum.

Maruyama, G. M. (1998). *Basics of structural equation modeling.* Thousand Oaks, CA: Sage.

McCoach, D. B., & Siegle, D. (2003). The structure and function of academic self-concept in gifted and general education students. *Roeper Review, 25*, 61–65.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Raykov, T., & Marcoulides, G. A. (2000). *A first course in structural equation modeling*. Mahwah, NJ: Erlbaum.

Schumacker, R. E., & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling*. Mahwah, NJ: Erlbaum.

Schumacker, R. E., & Marcoulides, G. A. (1998). *Interaction and nonlinear effects in structural equation modeling*. Mahwah, NJ: Erlbaum.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R.

H. Hoyle (Ed.), *Structural equation modeling* (pp. 56–75). Thousand Oaks, CA: Sage.

Author Notes

1. The research presented in this document is supported under the Educational Research and Development Centers Program, PR/Award number R206R50001, as administered by the Office of Educational Research and Improvement (OERI), U.S. Department of Education. The opinions expressed in this document do not necessarily reflect the positions or policies of OERI or the Department of Education.

2. The author wishes to thank Dr. David A. Kenny for his thoughtful comments on a previous version of this paper.

Endnotes

[1]A researcher who is conducting a mean structure analysis or a growth curve analysis would need the means for all of the observed variables, as well as the variance/covariance matrix. However, under normal circumstances, the variance/covariance matrix serves as the sufficient statistic for a SEM analysis.

[2]Technically, it is considered proper form to analyze a covariance matrix, but under a variety of conditions analyzing a correlation matrix will produce the same results, as a correlation matrix is simply a standardized version of a covariance matrix.

[3]Of course, they all must be "good" indicators.

[4]This hypothesis is commonly symbolized as Ho: S=S(q). See Bollen & Long (1993) for more detailed information about hypothesis testing in SEM.

[5]To correct for this problem, some researchers divide the χ^2 by the model degrees of freedom. The common rule of thumb is that this χ^2/df ratio should be less than 3 (Kline, 1998). This does not really solve the problem, as the degrees of freedom are related to model complexity and size, rather than sample size.

[6]Ideally, χ^2_1 should be nonsignificant.

[7]There are other estimation methods, but they are beyond the scope of this paper. For more information about alternative estimation methods, see Kaplan, 2000, or Hoyle, 1995.