

---

## RESEARCH METHODOLOGY

---

### How to Use Structural Equation Modeling in Medical Education Research: A Brief Guide

Claudio Violato

Kent G. Hecker

Medical Education Research Unit  
University of Calgary  
Calgary, Alberta, Canada

**Background:** Structural equation modeling (SEM) is a family of statistical techniques used for the analysis of multivariate data to measure latent variables and their interrelationships. SEM has potential to advance theory and research in medical education.

**Purpose:** The purpose of this article is to introduce SEM to medical education researchers and provide procedural information for applying SEM.

**Methods:** We outline the basic tenets of SEM, principles of model creation, identification, estimation, and model fit to data, and the use of SEM in medical education research.

**Results:** Although it is a powerful statistical research tool, SEM has had only limited use in medical education research. We explicate a five-step procedure for applying SEM to research problems and summarize an example of SEM to test a hypothetical model.

**Conclusions:** Notwithstanding some pitfalls, SEM does provide promise for testing complex, integrated theoretical models and advance research in medical education.

Teaching and Learning in Medicine, 19(4), 362-371

Copyright © 2007 Lawrence Erlbaum Associates, Inc.

Structural equation modeling (SEM) is a family of statistical techniques used for the systematic analysis of multivariate data to measure underlying hypothetical constructs (latent variables) and their interrelationships. It is a framework that allows researchers to translate theory into a testable model.<sup>1</sup> SEM builds on statistical techniques such as correlation, regression, and analysis of variance (ANOVA) and combines the strength of the confirmatory and data reducing ability of factor analysis, with the causal multiregression techniques of path analysis to explicate the direct and indirect relationships between measured and hypothetical (latent) variables.

Over the past 30 years the use of SEM has increased in several disciplines such as economics, education, psychology, and sociology. In psychology, for example, the citation frequency of SEM has reached the popularity of ANOVA (based on the citation frequency of SEM and M[ANOVA] of PsycINFO database 1970-2002).<sup>2-5</sup> This increase is attributable to the realiza-

tion that univariate and most multivariate techniques in social sciences and education research do not allow researchers to account for simultaneous causation and measurement error and that explanatory variables are frequently omitted in their data analysis. Furthermore, most non-SEM studies include only significant relationships between variables rather than formulating and estimating explanatory models.<sup>6,7</sup>

In this article we outline the basic tenets of SEM, the development of SEM as an integrated statistical theory, the central features of model creation, identification, estimation, model fit to data, and respecification. We then provide an overview of the published medical education research work that has used SEM, followed by a discussion of the strengths and weaknesses of SEM, and we provide an example of the use of SEM in medical education research. Finally, we outline areas where there are possible opportunities for the application of SEM in medical education research.

---

Correspondence may be sent to Claudio Violato, Health Sciences Centre, Medical Education and Research Unit, Department of Community Health Sciences, University of Calgary, 3330 Hospital Drive, NW, Calgary, Alberta, Canada, T2N 4N1. E-mail: violato@ucalgary.ca

**Basics of SEM**

SEM is a confirmatory approach that provides a mechanism to study hypothesized underlying structural relationships between latent variables or constructs. The development of SEM is based on integration of three key components: path analysis, factor analysis, and the development of estimation techniques for model fit.<sup>1</sup>

In 1921 Sewall Wright graphically expressed the interrelationships between variables and defined the "paths" as series of equations describing the inheritance patterns of guinea pigs. These conventions have been adopted in modern SEM applications. Path models represent the hypothesized causal connections among a set of variables and path analysis estimates the magnitude and statistical significance of each connection.<sup>8-10</sup> Path analysis moves beyond predicting whether independent variables predict a phenomenon—a central feature of regression analysis as well—to examining the interrelationships between the variables.

In 1904, Spearman first described factor analysis in an attempt to identify underlying constructs, which gave rise to observed variables through measurable characteristics.<sup>6</sup> His original work has led to modern exploratory factor analysis (EFA), a data reduction technique that identifies the relationships between measured variables and groups them into factors. Confirmatory factor analysis (CFA) is an extension of EFA and allows a researcher to model a priori how measured variables identify latent constructs. CFA provides a means by which researchers can model "error free" latent variables; this method identifies any variability (systematic and error variance) in a measured variable that is not associated with the latent construct.<sup>11</sup>

Integration of these statistical methods, which are referred to as measurement (factor analysis) and struc-

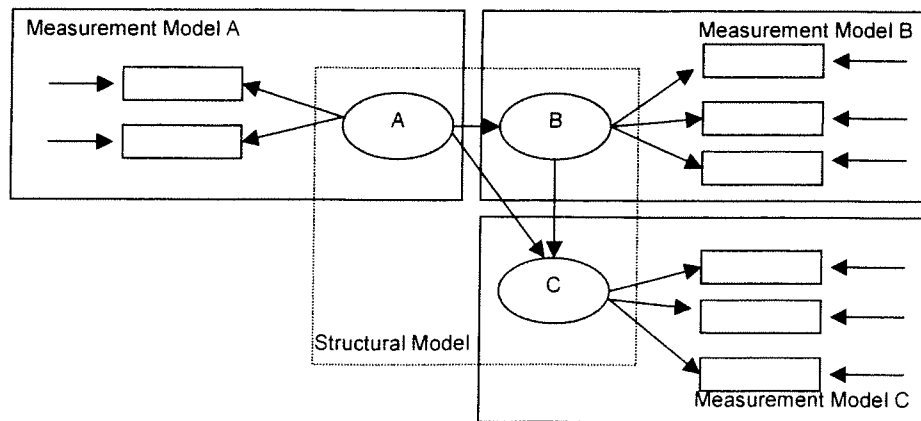
tural (path analysis) models, gave rise to the development of SEM first described by Joreskog, Keesing, and Wiley in the 1970s and is now known as the LISREL (linear structural relations) model.<sup>6</sup> The structural relations are those between the latent variables and are the core of SEM.

Figure 1 is a schematic summary of the measurement and structural parts of SEM. The measurement model represents how the latent variables are measured by indicator variables and describes the measurement properties of the indicator variables.<sup>5</sup> The structural model defines the relationships between latent variables (and possibly observed variables that are not indicators of latent variables) and allows for the determination of the extent of association between these variables.

**Advantages of SEM Over Other Statistical Methods**

SEM is a confirmatory approach that allows for the analyses of hypothesized interrelationships between latent constructs. It builds on the statistical tools that are based on regression, ANOVA, and correlation.<sup>11</sup> Although regression and ANOVAs define the degree of significant relationships between variables, it is difficult for researchers to model underlying constructs that independent variables might load upon.<sup>6</sup>

As we have seen, SEM is built on the multivariate techniques of factor and path analysis. Although these methods are quite strong independently, SEM subsumes and allows for a higher level of abstraction, through the development of structural models of hypothesized constructs. In path analysis, we analyze structural models between observed variables and depict direct and indirect effects based on hypotheses of causal effects.<sup>10</sup> Each variable is a single indicator,



**Figure 1.** A structural equation model depicting three latent variables (A, B, and C) and the relationship between measurement models and a structural model. *Note.* The path analytic model contains three variables. Variable A has both a direct and indirect on C. Modified from Nachtigall et al.<sup>2</sup>

however, and though it is generally understood that these variables contain measurement error, the analysis proceeds as if the variables are error free.<sup>12,13</sup> The strength of SEM over path analysis is that SEM allows for a structural model to be created between latent variables, or a combination of measured and latent variables. The path coefficients in a SEM between latent factors are corrected for measurement error (observed variables have measurement errors), because the hypothetical constructs are measured by multiple indicators where measurement error is "averaged out."

Factor analysis has resulted in the major applications of exploratory and confirmatory models. EFA does not require a priori hypotheses regarding the number of underlying factors or the relationships between measured variables and factors. Because of its atheoretical nature, it is not typically used as a SEM procedure.<sup>14</sup> CFA, on the other hand, requires a priori hypotheses and a proposed model (like SEM) about the number of factors and the nature of the relationships between measured and hypothetical constructs. Unlike SEM, we cannot model causality or the temporal relationships between variables in CFA. SEM subsumes CFA in that CFA is the measurement model of a SEM, outlining the relationships between indicators and underlying hypothetical constructs.

### An Integrated Research Method and Statistical Theory

SEM is a flexible and powerful statistical tool for the development, refinement, and validation of theories and hypothesized relationships between variables.<sup>1,13</sup> SEM requires fundamental understanding of statistical methods and concepts. It is considered to be primarily confirmatory; models are specified a priori based on theory and previous exploratory work. Nonetheless, it can also be exploratory in the sense that it provides a mechanism whereby competing models can be tested or models can be respecified (i.e., redrawn to improve fit). The relationships among the variables should be justifiable, based on theory, or based on previous research findings. Researchers should be cautious in respecifying models because simply adding more parameters will almost always improve fit but may not be theoretically relevant.<sup>13,15</sup> A researcher must specify a model, or competing models, a priori, before any analysis are performed. Analyses and interpretation of a SEM, especially the underlying hypothetical constructs represented as the latent variables, depend on how those variables were measured. Therefore, SEM not only subsumes many statistical concepts (correlation, ANOVA, multiple regression), it requires that the researcher understand measurement theory, meaning that the psychometric properties (reliability and validity) for each instrument used to measure an underlying

construct must be elucidated. It is important to understand that SEM tests the *whole* model for goodness of fit and provides information as to the relevance of various measurement components as well as relationships between variables that can be reviewed based on the theory that was used to create the model.<sup>6,7</sup>

At this point we introduce an example of a SEM application in medical education. Based on the results of previous theory and research on the predictive validity of the Medical College Admission Test (MCAT) and undergraduate grade point average (GPA) on the United States Medical Licensing Exams (USMLE), Collin and Violato<sup>16</sup> constructed a latent variable path model (which is the most fully developed SEM) of general achievement, aptitude for medicine, and performance in medicine (Figure 2).

To test the model depicted in Figure 2, data were obtained from the Association of American Medical Colleges and the National Board of Medical Examiners. MCAT student data consisting of all medical students in the United States from 1992 to 2001 and their corresponding USMLE data were employed. The model was fit to a random sample of 24,872 participants.

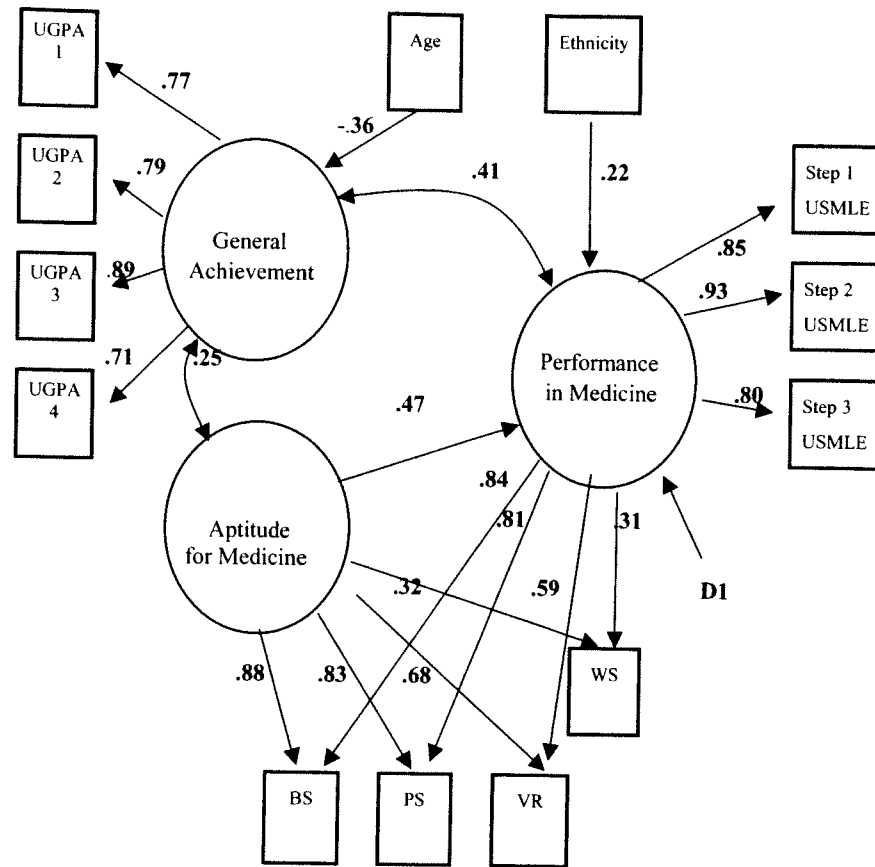
The model summarized in Figure 2 consists of 11 variables: GPA Year 1, GPA Year 2, GPA Year 3, GPA Year 4, the MCAT subtests (Biological Sciences, Physical Sciences, Verbal Reasoning, Written Sample), and the USMLE Steps (1, 2, and 3) were included as identifying the three latent variables (General Achievement, Aptitude for Medicine, Performance in Medicine).

The development and testing of a model occurs systematically. Five steps characterize most applications of SEM<sup>15</sup>: model specification, identification, estimation, testing fit, and respecification, as summarized in Figure 3 and discussed next. These operational steps are necessary to increase the likelihood that the predicted model will fit the observed data. These steps have also been discussed by Kline,<sup>13</sup> Bollen and Long,<sup>15</sup> and Boomsma.<sup>17</sup>

#### Step 1

A research problem is outlined and a model is specified. This problem and subsequent questions influence and are influenced by the underlying theory that has buttressed the work in that area. The theory underlying the model should be supported by the presence of preliminary empirical evidence gathered by reliable and valid psychometric tools and analyzed by the appropriate univariate and multivariate techniques.<sup>13</sup> As SEM relies on a priori hypotheses, the researcher must be able to define the theoretical underpinnings and how the research question was developed. The population under study also has an impact on the development of theory and research question(s) that are of interest.

A testable model, or competing testable models, is then developed based on the research question and



**Figure 2.** Latent variable path analysis model of UGPA, MCAT, and USMLE (Steps 1–3) latent variables employing ML estimation ( $n = 24,872$ ). *Note.* Fit indexes:  $\chi^2(55) = 11726.28, p < .001$  (CFI = .928, RMSEA = .025). UGPA 1–4 = Undergraduate GPA Year 1–4; BS = Biological Sciences MCAT Subtest; PS = Physical Sciences MCAT Subtest; VR = Verbal Reasoning MCAT Subtest; WS = Writing Sample MCAT Subtest; Step 1–3 USMLE = United States Medical Licensing Exam Step 1–3.

theory. The model is specified—the relationships between the variables must be explained and the measurement and the structural models are explicitly defined. Here the two component parts, the measurement model and the structural model, need to be explicated.

**Step 2**

The model is reviewed for identification—whether it is possible to find unique values for the parameters of the specified model. For models to be properly empirically assessed they should be identified or overidentified, meaning that the information in the data (that are the known values such as variances and covariances) is equal to or exceeds the information being estimated (unknown values such as parameter estimations, measurement error, etc.). If the unknowns exceed the knowns, then the model is underidentified.<sup>15</sup> Identification (under, over, or just) is determined by the number of manifest variables and the number of paths to estimate.

In the measurement model, the data should contain multiple indicators of each latent variable. Models with one latent construct should have three observed

variables for identification, whereas models with two or more latent constructs should have two or more observed variables for identification.<sup>13</sup> One indicator alone of a hypothesized latent construct results in a biased measurement if the error is unknown.<sup>7,15</sup> Although some suggest that at least three indicators should be used to account for as much of the variance as possible in the latent variable, typically most SEM models use two indicators. This may be adequate if the indicators are tests or scales. If they are single items or item “parcels,” however, several items should be used at least to achieve adequate reliability.<sup>18</sup> The indicators must be selected carefully, and the reliability and validity of the psychometric instruments used for the measured variables should be well described.

**Step 3**

The model is estimated. Estimation techniques were developed to determine how a model fits the observed data based on the extent to which the model implied covariance matrix is equivalent to the data derived covariance matrix.<sup>15</sup> This is formally described in the simple equation,  $\Sigma = \Sigma(\theta)$ , where  $\Sigma$  (sigma) is the population

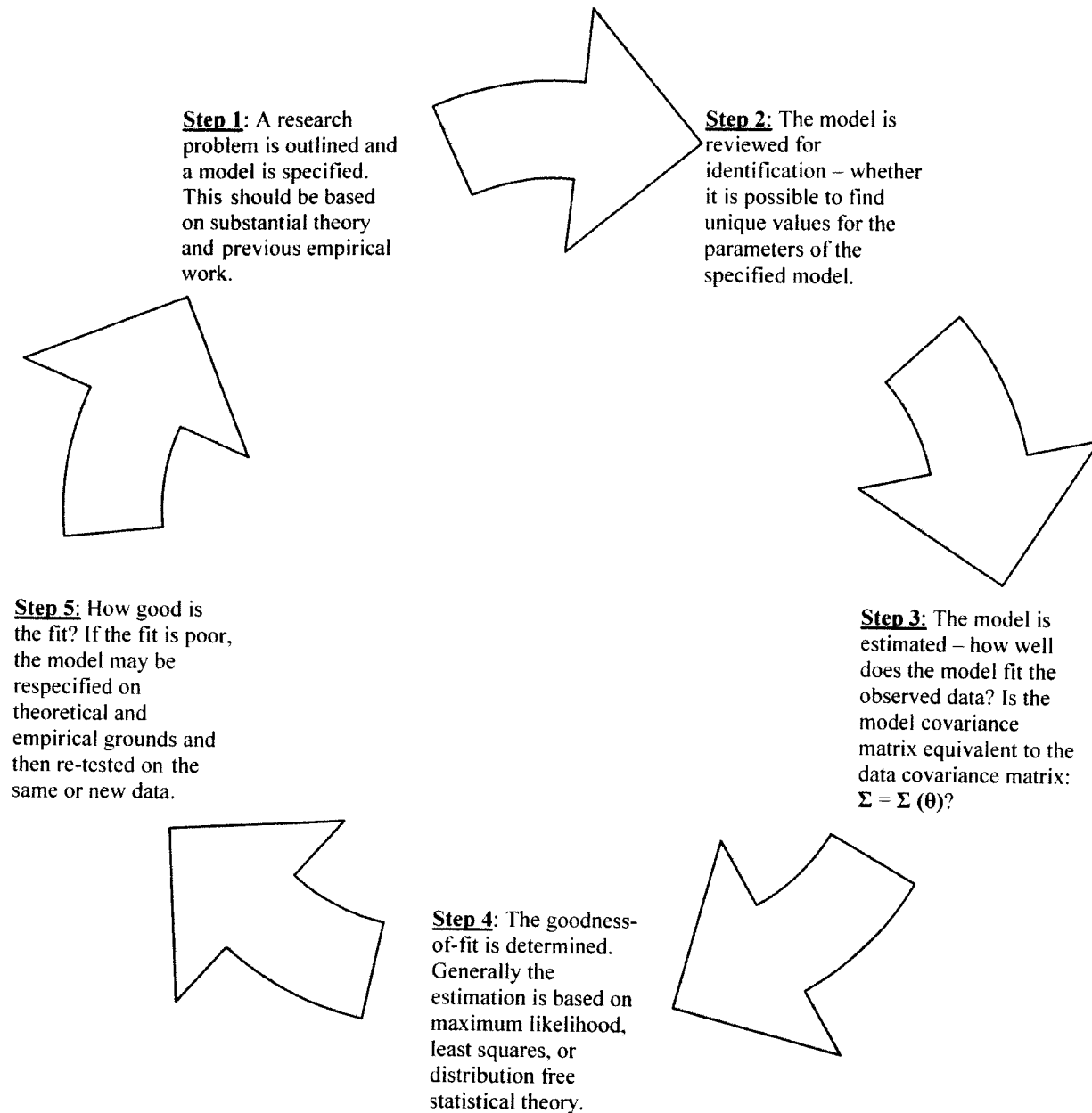


Figure 3. Iterative procedures and steps for developing and testing a structural equation model.

covariance matrix of observed variances,  $\theta$  (theta) is a vector that contains the population parameters, and  $\Sigma(\theta)$  is the covariance matrix written as a function of  $\theta$ .<sup>\*</sup> Both the simplicity and generality of this equation belies its elegance—it provides a unified way of including most statistical techniques in our field.<sup>1</sup>

The most common methods for estimation are based on the statistical theory that forms the basis of nor-

mal, elliptical, and arbitrary distribution estimators. They are maximum likelihood (ML), generalized least squares (GLS), and asymptotic distribution free (ADF) approaches respectively. These estimation procedures are iterative—the calculations performed are iterated until the best parameter estimation is obtained.<sup>\*\*</sup>

<sup>\*</sup>A hand-in-glove metaphor is useful in understanding this technique. A glove is the model ( $\Sigma(\theta)$ ) whereas a hand is the data ( $\Sigma$ ). We attempt to derive a perfect “glove” (i.e., model or theory) to fit over the hand (the data). The lack of fit of the glove (e.g., too few fingers [or too many], parts that are too large or too small, etc.) to the hand is indicated in the  $\theta$  vector.

<sup>\*\*</sup>ML is the default estimation technique in most SEM software and is the most widely used. ML assumes that variables are multivariate normal (i.e. distribution of the variables is multivariate normal) and requires large sample sizes. ML estimates parameters that maximize the likelihood (the probability) that the predicted model fits the observed model based on the covariance matrix.<sup>1,19</sup> The covariance, defined as  $cov_{xy} = r_{xy}SD_xSD_y$  (Pearson product moment correlation between variables  $X$  and  $Y$  multiplied by the standard deviation

The sample data are collected from the population under study. Descriptive and univariate analyses are performed to determine whether the data meets the assumptions for SEM, which are similar to most statistical methods dealing with parametric data. SEM is based on assumptions of multivariate normality, independence of observations (assumption of local statistical independence), and homoscedasticity (uniform variance across measured variables).

#### Step 4

The goodness of fit is determined. Generally the estimation is based on testing the null hypothesis of  $\Sigma = \Sigma(\theta)$ . A test statistic allows us to test the null hypothesis that the specified model leads to a reproduction of the population covariance matrix of the observed variables.

To assess goodness of fit of the observed data to the model, various fit indexes have been developed. There are no fixed rules as to which one to use, or which combination to use, and there is still little agreement on what represents the best fit.<sup>13</sup>

Fit criteria indicate the extent to which the model fits the data. A commonly used goodness-of-fit statistic, chi-square, provides a significance test that assesses whether the calculated covariance matrix is equal to the model implied covariance matrix or the null hypothesis that the differences between the two are zero. Other fit indexes are descriptive and are broken down into three categories—measures of overall model fit, measures based upon model comparisons, and measures of model parsimony.<sup>15</sup>

The chi-square statistic, however, should not be used as the sole criterion for fit because it has a number of technical problems (sensitivity to sample size, not interpretable in a standardized way, inflated Type I error rate for model rejection) as used in SEM. Descriptive goodness-of-fit measures are used in conjunction with the chi-square values to determine overall fit of the model to the data.

(SD) for  $X$  and  $Y$ ), is the statistic primarily used in SEM.<sup>15</sup> The covariance matrix used in SEM is meant to explicate patterns of correlation among a set of variables and to explain as much of their variance as possible within the model specified by the researcher.<sup>20</sup> GLS is based on the same assumptions as ML and used under the same conditions (it performs less well with smaller sample sizes; therefore ML is recommended).<sup>20,21</sup> GLS reduces the sum of the squared deviations (or variances) between the predicted model and observed model and is more popular in regression analyses.<sup>20</sup> ADF estimation techniques (such as arbitrary distribution least squares) may be used if some measured variables are dichotomous and others are continuous, and therefore multivariate normality cannot be assumed or if the distributions of the continuous variables are non-normal. These estimation techniques are not as widely used as ML and GLS. ML, GLS, and ADF estimation methods are based on some rather complex algorithms and further accessible explanations can be found in Kline,<sup>13</sup> Schermelleh-Engel et al.<sup>19</sup> and Myung.<sup>22</sup>

Descriptive measures based on model comparisons assess the fit of a model compared to the fit of a baseline model. Baseline models are either an independence model where it is assumed that the observed variables are measured without error or a null model where all parameters are fixed to zero. Descriptive measures of model parsimony are used primarily when competing models are being compared.<sup>13</sup> Typically values of these measures range from zero (no fit) to one (perfect fit), whereas some provide a badness-of-fit with parameters of zero to one.

Common fit indexes include root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), goodness of fit index (GFI), Akaike information criterion (AIC), the comparative fit index (CFI), and a number of others. The most widely reported index is the CFI, which compares a predicted model with a baseline model.

How should fit be interpreted? Most software packages that are specialized for SEM analyses (e.g., EQS, LISREL, AMOS) provide several indexes that represent the different classes of fit criteria. Researchers should compare and contrast the chi-square value and its  $p$  value,  $\chi^2/df$ , RMSEA, and its confidence interval (CI), SRMR, nonnormed fit index, and CFI. Usually the chi-square value, CFI, SRMR, and RMSEA are presented as indicators of goodness of fit.<sup>20,23</sup> For model comparisons, the chi-square difference test and AIC should be examined. Results based on sample sizes of 250 or less should be treated with caution. We recommend reporting multiple fit indexes such as the CFI in combination with SRMR and RMSEA, as this combination tends not to reject models under nonrobust conditions (e.g., violations of multinormality).<sup>21</sup> Generally models with CFIs  $\geq .90$ , SRMR  $< .01$ , and RMSEA  $< .10$  are considered a good fit.

Finally, model selection should be guided by the principles of parsimony. If several models fit the data, the simplest should be selected.<sup>15</sup> A model can only be rejected, it can never be *proven* to be valid. If the fit is acceptable the proposed relationships in the measurement model between the latent and the observed variables and the structural models between the latent variables is supported by the data. If the fit is considered poor (e.g., CFI  $< .90$ ), the model can be respecified.

#### Step 5

The model may be respecified on theoretical grounds. The first step is to respecify the measurement model, which is then analyzed to assess fit. If the fit is acceptable, then the SEM model can be analyzed. The model is then assessed and respecified if necessary, based on statistical output and theoretical relevance. The respecified model should also be tested on new data. The theory can be extended and refined and the results replicated by other researchers as well.<sup>24</sup>

Finally, conclusions can be drawn from analyses and comparison to theory. This is the iterative process of model development, identification and testing as summarized in Figure 3.

### Sample Size and Model Fit

Sample size is important for goodness of fit as the estimation procedures used to calculate the model parameters require a sample size large enough to obtain meaningful parameter estimates.<sup>21</sup> Some have suggested that sample size needs to be more than 25 times the number of parameters with a minimum subject to parameter ratio of 10:1, but with the proviso that the lower bound of the total sample size should be approximately 100 to 200.<sup>1,21</sup> Bentler proposed that when sample sizes are small, multiple competing models should be tested. If some of the models are rejected using the fit indexes, then the sample size is probably large enough because there is enough power to reject competing models.<sup>21</sup>

### Strengths and Weaknesses of SEM

#### Strengths

SEM is a theory strong approach. Researchers employing this technique model hypothesized (latent) variables and, more important, model the indirect and direct relationships between the latent and observed variables. It supersedes other multivariate analyses because it can model the relationships between "error free" latent variables by partialing out measurement error from multiple, imperfectly reliable indicators.<sup>6</sup> Therefore, SEM can be used for a number of research designs.<sup>13</sup>

Path analysis has sometimes been referred to as causal modeling. Because path analysis forms the structural model in SEM, it can provide evidence for causality among variables. Three conditions of causation must be met, however: There should be an observed and measured relationship between  $x$  and  $y$  (should be correlated: association),  $x$  should precede  $y$  in time (temporal order: direction of influence), and  $x$  and  $y$  should have a nonspurious relationship—the observed, measurable and temporal relationship will not disappear when other variables on this relationship are controlled (isolation).<sup>1,13</sup>

Theoreticians caution that causation can never be *proven* and argue that the structural model can only determine if the researcher's hypothesized causal inferences are consistent with the data.<sup>1,6</sup> A good fit does not imply a strong effect on the dependent variable—even a high proportion of explained variance does not confirm causality. The best we can expect from SEM

is evidence against a poor model but never *proof* of a good one.<sup>25</sup> Well-designed, theory strong studies, however, can provide powerful evidence for the effect of one variable on another.

#### Weaknesses

SEM does have a number of limitations. It requires a very well-developed theory and empirical evidence and therefore is limited in its applicability. When SEM is used for exploratory purposes, model fit may be more a function of statistical fit than theoretical fit. Furthermore, if the model is misspecified due to weak theory, unclear hypotheses or poor study design, the causal relationships between the variables will be misinterpreted.

SEM, like other analytic procedures, cannot compensate for unreliable measures.<sup>6</sup> If measuring instruments with poor reliability are employed, the SEM will be laden with error. The use of only one measure to identify a latent variable is also a weakness in that it reduces the amount of variability that can be identified in the latent variable, thus producing a biased measurement.<sup>6</sup> SEM cannot compensate for instruments with poor reliability and validity, poorly specified theoretical models, inadequate sampling, or misinterpretation of the fit indexes. SEM generally requires large samples that may be difficult to get. Finally, SEM requires considerable theoretical and statistical sophistication by researchers.

### SEM and Medical Education Research

Although identified as a potentially powerful research tool, SEM has not been used extensively in medical education research. A PubMed search using *structural equation model* and *medical education* resulted in only 8 citations in English language journals (as of May 24, 2006). Using *confirmatory factor analysis* or *path analysis* with *medical education* provided 16 and 8 citations, respectively. Assuming that research with these types of analyses might not be reported in biomedical literature but may be in the education and psychological literature, we searched ERIC and PsycINFO using the same search terms and found 6 additional empirical reports.

The majority of the reported use of SEM in medical education research has in fact been CFAs and path analyses. These techniques have been used to test competing hypothetical models of how basic science knowledge and clinical knowledge are used in diagnostic reasoning,<sup>26,27</sup> to assess the predictive validity of various standard and nonstandard selection criteria on medical school performance,<sup>28</sup> to assess students' motivation for choosing a specialization,<sup>29,30</sup> to test a causal model of the influence of educational

interventions and motivation in problem-based learning,<sup>31,32</sup> to confirm the factor structure of a psychometric tool measuring readiness to engage self-directed learning,<sup>33</sup> to determine the stability of communication skills in medical students as measured in objective structured clinical examinations measured at different times,<sup>34</sup> to explore two types of formative assessment techniques on performance,<sup>35</sup> and to evaluate clerkship experiences on clinical competence.<sup>36</sup> Very few of these have employed the full SEM models of latent variable path analysis (LVPA) identifying latent variables, their intercorrelations, and directional influence.

De Bruin et al.<sup>26</sup> tested four theories on the role of basic science knowledge and clinical knowledge in diagnostic reasoning of 59 family physicians and 184 second- to sixth-year medical students at Maastricht University. SEM was used to analyze the data. In the first model, only basic science knowledge was involved in diagnostic reasoning; in the second model, only clinical knowledge was related to diagnostic reasoning; in the third model, clinical knowledge was related to diagnostic reasoning, but basic science knowledge was integrated in clinical knowledge; and in the fourth model, both basic science knowledge and clinical knowledge independently influence diagnostic reasoning. The results indicated that the third model, which is based on the knowledge encapsulation theory, provided the best fit to the data, whereas the models that had directly related basic science knowledge with diagnostic performance did not fit the data adequately. Thus the results strongly supported the Schmidt and Boshuizen<sup>37</sup> theory of knowledge encapsulation that basic science knowledge is activated in expert diagnostic reasoning through its relation with clinical knowledge.

In another interesting application of SEM in medical education, a British medical student cohort was studied over 5 years by Ferguson et al.,<sup>28</sup> who incorporated 18 different assessments into a structural equation model. They found that conscientiousness was positively related to A-level grades and preclinical performance but was negatively related to clinical grades. Other measures (e.g., teachers' reference, personal statements) had no predictive value in their model. The SEM analyses permitted Ferguson et al. to show the complex relationship of conscientiousness with preclinical and clinical grades while showing that more conventional measures lacked predictive validity.

### An Example of an LVPA

The Collin and Violato LVPA model depicted in Figure 2 was tested employing EQS 6.1.<sup>25</sup> The 11 measured variables and the 2 exogenous variables, age and ethnicity, were intercorrelated and converted to a variance-covariance matrix, and the model was fit using ML estimation.

The overall fit of the model to the data was good, resulting in a CFI of .928,  $\chi^2(55) = 11726.28$ ,  $p < .001$ . A CFI of .928 indicates that 92.8% of the variance and covariance in the data is accounted for by the proposed model. Further evidence of the model's fit comes from the RMSEA of .025. Nearly all of the residuals were zero. The standardized path coefficients and other parameters are summarized in Figure 2.

All three latent variables in Figure 2 are identified. General Achievement (path coefficients ranging .36–.89) is identified by four variables (GPA Year 1, GPA Year 2, GPA Year 3, GPA Year 4) that are all theoretically relevant to an achievement construct. The path coefficients on Aptitude for Medicine range from .32 to .84 (Bio Sc, Phys Sc, Verbal, Written) that are all subsets of the MCAT. Three of the measured variables (USMLE Step 1, USMLE Step 2, USMLE Step 3) also serve to identify Performance in Medicine (MCAT Bio Sc, MCAT Phys Sc, MCAT Verbal, MCAT Written have split loadings on this latent variable) that are all relevant to a performance in medicine construct. There is a direct path coefficient (0.47,  $p < .001$ ) from Aptitude for Medicine to Performance in Medicine that confirms the "causal" direction. The correlations between General Achievement ( $r = .41$ ,  $p < .001$ ) and Performance in Medicine and Aptitude for Medicine ( $r = .25$ ,  $p < .001$ ) all confirm the expected relationship.

The General Achievement factor has a significant path coefficient (0.36,  $p < .01$ ) from the exogenous variables age and ethnicity. For age, the path accounts for 12.7% of the variance indicating that age is related to undergraduate achievement (inversely— younger students perform better than older ones). Ethnicity has a small (.22,  $p < .05$ ) path coefficient on Performance in Medicine (5.0% of the variance) but nonetheless influences this variable.

All of the latent variables are related as predicted, and the crucial path coefficient from Aptitude for Medicine to Performance in Medicine is significant (path coefficient = 0.47,  $p < .001$ ). The intercorrelations among the other latent variables are all significant. These results support the overall model and the particulars of the predicted relationships as well.

This LVPA model tested the long-standing theory that three constructs, Aptitude for Medicine, General Achievement, and Performance in Medicine, can be identified and their interrelationships determined. No previous research has been published testing the entire model of aptitude-achievement-performance, identifying these as latent variables, however. Most research heretofore has employed simple correlation or regression methods. The Collin and Violato research, therefore, provides a considerable advance over previous work in this area and demonstrates how a complex

theoretical problem can be addressed employing SEM. The use of the large sample in the study provides confidence in the stability of the model (i.e., it should readily replicate across samples).

### Conclusion

SEM has a number of promises. It is an integrated statistical method that allows researchers to test proposed theories through quantifiable measures. It builds on statistical techniques such as correlation, EFA, regression, and ANOVA and combines the strength of CFA with path analysis to explicate the direct and indirect relationships between measured and hypothetical variables.

Models for SEM are composed of two basic parts—the measurement model, which is the theorized organization of indicator variables and how they identify the latent variables, and the structural model, which refers to the relationships between the latent variables. A good fit of the structural and measurement models to the data provides evidence that the hypothesized relationships in the model are consistent with the relationships in the observed data.

The research methods underpinning SEM makes it a theory strong approach. For models to be tested for fit to data, a series of sequential steps must be undertaken, including model specification based on theory and empirical evidence, identification of reliable and valid psychometric tools to be used as indicators for the hypothetical constructs, estimation procedures, model evaluation through the use of fit indexes, and model respecification.<sup>1,13,15,17,23,24</sup>

SEM is a useful research and statistical method for medical education research and has the potential for advances in the area. These may include advances in longitudinal measurement of student learning, verifying the predictive validity of the medical school selection process, predictive validity studies using cognitive and noncognitive factors as predictors for student success, testing aspects of clinical performance and test scale development, and testing the validity of the aptitude–achievement–performance complex.

SEM does have limitations. It requires a strong conceptual understanding of the theory relevant to the research question. Much of theory in medical education may currently lack this strong conceptual understanding. SEM also requires substantial prior empirical evidence. Again, basic empirical evidence is lacking in many aspects of medical education research. SEM also requires large samples that may be difficult to obtain. Finally, it requires considerable theoretical and statistical sophistication that may be lacking among medical education researchers. Notwithstanding these limitations, SEM does provide promise for testing complex, integrated theoretical models in medical education.

### References

1. Bollen KA. *Structural equations with latent variables*. New York: Wiley, 1989.
2. Tremblay PF, Gardner RC. On the growth of structural equation modeling in psychological journals. *Structural Equation Modeling* 1996;3:93–104.
3. Hershberger SL. The growth of structural equation modeling: 1994–2001. *Structural Equation Modeling* 2003;10:35–46.
4. MacCallum RC, Austin JT. Applications of structural equation modeling in psychological research. *Annual Review of Psychology* 2000;51:201–26.
5. Nachtigall C, Kroehne U, Funke F, Steyer R. (Why) should we use SEM? Pros and cons of structural equation modeling. *Methods of Psychological Research* 2003;8(2):1–22.
6. Joreskog KG. The LISREL approach to causal model-building in the social sciences. In KG Joreskog, H Wold (Eds.), *Systems under indirect observation causality-structure-prediction* (pp. 81–99). New York: North-Holland, 1982.
7. Anderson JG. Structural equation modeling in the social and behavioral sciences: Model building. *Child Development* 1987;58:49–64.
8. Duncan OD. *Introduction to structural equation models*. New York: Academic Press, 1975.
9. Li CC. *Path analysis: A primer*. Pacific Grove, CA: Boxwood, 1975.
10. Stage FK, Carter HC, Nora A. Path analysis: An introduction and analysis of a decade of research. *Journal of Educational Research* 2004;98:5–12.
11. Kline RB. Latent variable path analysis in clinical research: A beginner's tour guide. *Journal of Clinical Psychology* 1991;47:471–84.
12. Bentler PM. Linear systems with multiple levels and types of latent variables. In KG Joreskog, H Wold (Eds.), *Systems under indirect observation causality-structure-prediction* (pp. 101–3). New York: North-Holland, 1982.
13. Kline RB. *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford, 2005.
14. Fabrigar LR, Wegener DT, MacCullum RC, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. *Psychology Methods* 1999;4:272–99.
15. Bollen KA, Long JS (Eds.). *Testing structural equation models*. Newbury Park, CA: Sage, 1993.
16. Collin VT, Violato C. *Predictive validity study of the medical college admission test using multiple regression and latent variable path analysis*. In: O'Sullivan P, ed. RIME 44th Annual Conference on Research in Medical Education. Washington, DC, 2005, p. 65.
17. Boomsma A. Reporting analyses of covariance structures. *Structural Equation Modeling* 2000;7:461–83.
18. Rigdon EE. A necessary and sufficient identification rule for structural models in practice. *Multivariate Behavioral Research* 1995;30:359–83.
19. Schermelleh-Engel K, Moosbrugger H, Muller H. Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research* 2003;8:23–74.
20. Byrne BM, Crombie G. Modeling and testing change: An introduction to the latent growth curve model. *Understanding Statistics* 2003;2:177–203.
21. Bentler PM, Yuan K-H. Structural equation modeling with small samples: Test statistics. *Multivariate Behavioral Research* 1999;34:181–97.
22. Myung IJ. Tutorial on maximum likelihood estimation. *Journal of Math Psychology* 2003;47:90–100.
23. Hu L-T, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling* 1999;6:1.

24. Anderson JG, Gerbing DW. Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin* 1988;103:411-23.
25. Bentler PM. *EQS 6.1 structural equations program manual*. Encino, CA: Multivariate Software, Inc., 2004.
26. de Bruin AB, Schmidt HG, Rikers RM. The role of basic science knowledge and clinical knowledge in diagnostic reasoning: a structural equation modeling approach. *Academic Medicine* 2005;80:765-73.
27. Donnon T, Violato C. Medical students' clinical reasoning skills as a function of basic science achievement and clinical competency measures: A structural equation model. *Academic Medicine* 2006;81(10, Suppl.):51-4.
28. Ferguson E, James D, O'Hehir F, Sanders A. Pilot study of the roles of personality, references, and personal statements in relation to performance over the five years of a medical degree. *British Medical Journal* 2003;326:429-32.
29. Williams GC, Saizow R, Ross L, Deci EL. Motivation underlying career choice for internal medicine and surgery. *Social Science Medicine* 1997;45:1705-13.
30. Williams GC, Wiener MW, Markakis KM, Reeve J, Deci EL. Medical students' motivation for internal medicine. *Journal of General Intern Medicine* 1994;9:327-33.
31. Schmidt HG, Moust JH. What makes a tutor effective? A structural-equations modeling approach to learning in problem-based curricula. *Academic Medicine* 1995;70:708-14.
32. Berkel HJM, Schmidt HG. Motivation to commit oneself as a determinant of achievement in problem-based learning. *Higher Education* 2000;40:231-42.
33. Hoban JD, Lawson SR, Mazmanian PE, Best AM, Seibel HR. The Self-Directed Learning Readiness Scale: A factor analysis study. *Medical Education* 2005;39:370-9.
34. Humphris GM. Communication skills knowledge, understanding and OSCE performance in medical trainees: A multivariate prospective study using structural equation modeling. *Medical Education* 2002;36:842-52.
35. Krasne S, Wimmers PF, Relan A, Drake TA. Differential effects of two types of formative assessment in predicting performance of first-year medical students. *Advances in Health Sciences Education: Theory and Practice* 2006;11:155-71.
36. Wimmers PF, Schmidt HG, Splinter TAW. Influence of clerkship experiences on clinical competence. *Medical Education* 2006;40:450-58.
37. Schmidt HG, Boshuizen HP. On acquiring expertise in medicine. *Education and Psychology Review* 1993;5:205-21.

*Final revision received on May 4, 2007.*