Multivariate Behavioral Research

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hmbr20

Have Multilevel Models Been Structural Equation Models All Along?

Patrick J. Curran

Published online: 10 Jun 2010.

Multivariate Behavioral Research, *38* (4), 529-569

2002 Cattell Award

Have Multilevel Models Been Structural Equation Models All Along?

Patrick J. Curran
University of North Carolina at Chapel Hill

A core assumption of the standard multiple regression model is independence of residuals, the violation of which results in biased standard errors and test statistics. The structural equation model (SEM) generalizes the regression model in several key ways, but the SEM also assumes independence of residuals. The multilevel model (MLM) was developed to extend the regression model to dependent data structures. Attempts have been made to extend the SEM in similar ways, but several complications currently limit the general application of these techniques in practice. Interestingly, it is well known that under a broad set of conditions SEM and MLM longitudinal "growth curve" models are analytically and empirically identical. This is intriguing given the clear violation of independence in growth modeling that does not detrimentally affect the standard SEM. Better understanding the source and potential implications of this isomorphism is my focus here. I begin by exploring why SEM and MLM are analytically equivalent methods in the presence of nesting due to repeated observations over time. I then capitalize on this equivalency to allow for the extension of SEMs to a general class of nested data structures. I conclude with a description of potential opportunities for multilevel SEMs and directions for future developments.

The structural equation model (SEM) is a flexible and powerful analytical method that has become a mainstay in many areas of social science research. The generality of this approach is evidenced in the ability to parameterize the SEM to estimate well known members of the general linear modeling (GLM) family including the *t*-test, ANOVA, ANCOVA, MANOVA, MANCOVA, and the multiple regression model. However, the

This article is based on the Cattell Award presentation given by the author at the October, 2002, meeting of the Society for Multivariate Experimental Psychology in Charlottesville, Virginia. Correspondence should be addressed to Patrick J. Curran, Department of Psychology, University of North Carolina, Chapel Hill, NC, 27599-3270. Electronic mail can be sent to curran@unc.edu. This work was funded in part by grant DA13148 awarded to Patrick Curran and Kenneth Bollen. I would like to thank Dan Bauer, Ken Bollen, Robert MacCallum, Nathan Vandergrift, R. J. Wirth and the members of the Carolina Structural Equation Modeling Group for their valuable input throughout this project.

SEM can be expanded to allow for the estimation of measurement error through the use of multiple indicator latent factors, the testing of complex mediational mechanisms through the decomposition of effects, and the testing of moderational mechanisms through the estimation of multiple group analysis, just to name a few. Unquestionably, the SEM is a significant and indispensable tool to empirical researchers.

Of course the SEM is not without limitations. The most widely used method of estimation is maximum likelihood (ML). Under the general assumptions of sufficiently large sample size, proper model specification, and residuals that are independent and normally distributed, ML provides asymptotically unbiased, consistent, and efficient parameter estimates and standard errors (Bollen, 1989). However, these asymptotic properties are commonly violated in practice, and recent developments have led to improved estimators in the presence of non-normality (e.g., ADF estimation; Browne, 1984), categorical dependent variables (e.g., WLS; Muthén, 1984), misspecified models (2SLS, Bollen, 1995, 1996) and small sample size (e.g., Yuan & Bentler, 2001). Despite these improvements, one of the most vexing challenges remains proper model estimation of SEMs in the presence of dependent data.

The assumption of independence implies that the residuals for a given dependent variable are mutually uncorrelated. However, there are many situations in which this assumption is violated in practice, whether it be explicitly introduced by design (e.g., children nested within classrooms) or may arise more subtly during data collection (e.g., interviewees nested within interviewer). Regardless of source, it has long been known that violation of the assumption of independence leads to biased test statistics, standard errors, and even parameter estimates due to the inappropriate aggregation across levels of analysis (Hox, 1998; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999). Although much work has focused on expanding the SEM for application with dependent data structures using both true ML (Bentler & Liang, 2003; du Toit & du Toit, in press; McDonald & Goldstein, 1989) and pseudo-ML (e.g., Muthén, 1989, 1994) estimation, there remain limitations with each approach.

Yet here lies a curiosity. A classic source of dependency in data structures arises from the repeated assessments of individuals over time. Despite this clear two-level nesting of time within individual, it has long been known that two-level random effects growth models can be fully estimated within the standard SEM framework (MacCallum, Kim, Malarkey, & Kielcolt-Glaser, 1997; Meredith & Tisak, 1984, 1990; Willett & Sayer, 1994). Indeed, under many broad conditions the SEM growth model is analytically equivalent to that of the traditional multilevel model (MLM), an approach that is explicitly designed for evaluating nested data structures. This raises two interesting questions. First, exactly what characteristics of the two-level random effects growth model allow for the equivalent estimation in SEM and MLM? Second, might we be able to capitalize upon these characteristics to extend the standard SEM for novel uses in more general complex data structures? Exploring these two questions is my goal here.

I will show that the general rule for using the SEM in the analysis of longitudinal data is that the values of the level-1 predictor variables are incorporated into the SEM via the factor loading matrix. This is of course well known given the coding of the measure of time in the SEM, but I will demonstrate that this strategy can be extended for any number of level-1 predictors. I will then demonstrate that the standard SEM can be used to evaluate a variety of models in the presence of nested data structures that arise from sources other than longitudinal data. Finally, I will argue that for some types of questions the SEM approach to nested data will provide certain advantages over the standard MLM approach, whereas in some situations the opposite will hold.

Prior Explorations into the Intersection between MLM and SEM

I am far from the first person to explore the increasingly porous boundaries between multilevel and structural equation models. Of course Meredith and Tisak (1984, 1990) had early insights into using the SEM framework to fit what was to become more widely known as multilevel models. McArdle and Hamagami (1996) used a multiple group SEM to estimate a particular subset of MLMs. Rovine and Molenaar (1998, 2000, 2001) explored the intersection between SEM and MLM using separate structures for the fixed and random effects to stay maximally consistent with the Laird and Ware (1982) expressions. Newsom (2002) explored a creative variant of these models to use the SEM framework to estimate a MLM for dyadic data. When discussing potential applications of SEMs in nested data structures using Mx, Neale, Boker, Xie, and Maes (1999) noted "...this type of modeling is equivalent to Hierarchical Linear Modeling (HLM) as specified by Bryk and Raudenbush (1992) and others. This aspect of Mx has not received much attention..." (p. 33). And most recently, Bauer (in press) expanded upon an idea presented in Bauer and Curran (2002) to pursue a rigorous development and application of multilevel SEMs to test complex factorial measurement in three-level nested data structures. Thus, although my goal is to make several unique contributions here, I have drawn upon the creativity and insights of this body of work in the development and articulation of my own thoughts on this matter. I am particularly inspired by the work of my colleague Dan Bauer, and much of his original thinking is reflected in my work here.

The Standard Structural Equation Model

There are two fundamental equations that define the general SEM: the measurement equation and the structural equation. The measurement equation is given as

$$\mathbf{y} = \boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{y}$ is a $p \times 1$ vector of p-observed variables, $\boldsymbol{\nu}$ is a $p \times 1$ matrix of measurement intercepts, $\boldsymbol{\Lambda}$ is a $p \times k$ matrix of factor loadings relating the p-observed variables to the k-latent factors, $\boldsymbol{\eta}$ is a $k \times 1$ matrix of latent factor scores, and $\boldsymbol{\varepsilon}$ is a $p \times 1$ vector of residuals. The structural equation is then defined as

$$\boldsymbol{\eta} = \boldsymbol{\mu} + \boldsymbol{\beta}\boldsymbol{\eta} + \boldsymbol{\zeta}, \tag{2}$$

where $\boldsymbol{\eta}$ is defined as before, $\boldsymbol{\mu}$ is a $k \times 1$ vector of latent factor means and intercepts, $\boldsymbol{\beta}$ is a $k \times k$ matrix of regression coefficients among the latent factors, and $\boldsymbol{\zeta}$ is a $k \times 1$ vector of residuals.

Finally, we can substitute Equation 2 into Equation 1 to express the reduce form expression for $\mathbf{y}$ such that

$$\mathbf{y} = \boldsymbol{\nu} + \boldsymbol{\Lambda}(\boldsymbol{\mu} + \boldsymbol{\beta}\boldsymbol{\eta} + \boldsymbol{\zeta}) + \boldsymbol{\varepsilon}, \tag{3}$$

and with simple rearrangement

$$\mathbf{y} = (\boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\mu}) + (\boldsymbol{\Lambda}\boldsymbol{\beta}\boldsymbol{\eta}) + (\boldsymbol{\Lambda}\boldsymbol{\zeta} + \boldsymbol{\varepsilon}), \tag{4}$$

which highlights the parameterization of the means and intercepts, the factor loadings and factor regressions, and the disturbances and residuals.

Importantly, the covariance and mean structure of $\mathbf{y}$ can be expressed in terms of the model parameters. The covariance structure implied by Equation 4 is

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}(\mathbf{I} - \boldsymbol{\beta})^{-1}\boldsymbol{\Psi}(\mathbf{I} - \boldsymbol{\beta})^{-1\prime}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}_{\boldsymbol{\varepsilon}}, \tag{5}$$

where $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ represents the $p \times p$ covariance matrix of $\mathbf{y}$ expressed as a function of the model parameters in $\boldsymbol{\theta}$, $\boldsymbol{\Psi}$ is the $k \times k$ covariance matrix among latent factors, $\boldsymbol{\Theta}_{\boldsymbol{\varepsilon}}$ is the $p \times p$ covariance matrix of residuals, $\mathbf{I}$ is a $k \times k$ identity matrix, and all else is defined as earlier. Further, the mean structure implied by Equation 4 is

$$\text{(6)} \qquad \boldsymbol{\mu}(\boldsymbol{\theta}) = \boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\beta}\boldsymbol{\mu},$$

where $\boldsymbol{\mu}(\boldsymbol{\theta})$ represents the $p \times 1$ vector of means of $\mathbf{y}$ expressed as a function of the model parameters in $\boldsymbol{\theta}$, and all else is defined as earlier.

The analytic goal is to estimate the model parameters in $\boldsymbol{\theta}$ via the minimization of a suitable discrepancy function F. Although there are a variety of methods available, the most common is the maximum likelihood (ML) estimator. This fit function is denoted F_{ML} and can be expressed as

$$\text{(7)}\ F_{\mathrm{ML}} = \{\ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})| - \ln|\mathbf{S}| + tr[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{S}] - p\} - \{[\bar{\mathbf{y}} - \boldsymbol{\mu}(\boldsymbol{\theta})]'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})[\bar{\mathbf{y}} - \boldsymbol{\mu}(\boldsymbol{\theta})]\},$$

where $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and $\boldsymbol{\mu}(\boldsymbol{\theta})$ are defined as before, $\mathbf{S}$ is the sample covariance matrix and $\bar{\mathbf{y}}$ is the observed mean vector. Under the assumptions of sufficient sample size, correct model specification, no excess multivariate kurtosis, and independence of residuals, the parameter estimates in $\hat{\boldsymbol{\theta}}$ are asymptotically unbiased, efficient and consistent (Bollen, 1989). Further, under these same assumptions, the test statistic defined as $T = \hat{F}_{\mathrm{ML}}(N_T - 1)$ (where $\hat{F}_{\mathrm{ML}}$ is the sample minimum of the discrepancy function and N_T is the total sample size) is distributed as a central-χ^2 with $df = \{[p(p+1)/2] + p\} - t$ where t is the number of free parameters in $\boldsymbol{\theta}$. The test of T provides an evaluation of the null hypothesis that the covariance matrix and mean vector in the population are equal to those implied by the model (e.g., as defined in Equations 5 and 6). The implications of violating each of the assumptions underlying ML have been studied extensively. Further, post hoc adjustments and alternative methods of estimation have been developed that are more robust to many types of violations. However, the assumption of independence of observations poses a particularly salient challenge when estimating SEMs under maximum likelihood.

The Assumption of Independence of Observations in SEM

Independence is a standard assumption within the entire general linear model (GLM) and thus of course applies directly to the SEM as well. It is well known that violation of the independence assumption introduces predictable sources of bias into the estimation of the SEM. Specifically, standard errors tend to be too small, test statistics tend to be too large, and biased coefficients may result from inappropriate aggregation across levels of nesting (Hox, 1998;

Raudenbush & Bryk, 2002; Snijders & Bosker, 1999). Further, failure to explicitly model the nested structure of the data may significantly preclude our ability to test certain questions of interest (e.g., simultaneous disaggregation of child effects from teacher effects from school effects). Given the many advantages associated with the SEM in general, attempts have been made to incorporate nested data structures into the SEM.

The first approach was to incorporate a true full information ML (FIML) estimator for multilevel SEMs. Seminal work in this area was presented in Goldstein and McDonald (1988), McDonald and Goldstein (1989), and McDonald (1993, 1994). Although analytically elegant and ahead of its time, this FIML approach was limited in application given the need to invert a large number of high dimension matrices and the reliance on specialized software packages. Recent efforts to overcome these challenges have been presented by Bentler and Liang (2003) and du Toit and du Toit (in press), but these new methods have not yet been closely studied. An alternative to the FIML approach is a pseudo-ML estimator that is a drastically simplified version of FIML given the imposed assumption of a fully balanced design (e.g., equal numbers of level-1 observations nested within all level-2 units). This method has primarily been described by Muthén (1989, 1994) who drew on the developments of McDonald and Goldstein (1989) and McDonald (1993) and referred to this as Muthén's ML (MUML) estimator. MUML is not a true ML estimator under the realistic condition of unbalanced designs and is also limited to the estimation of random intercepts only. Although both FIML and MUML estimators provide certain advantages, there are still a number of limitations that reflect the difficulty of extending the general SEM for application with nested data structures.

Multilevel Modeling

Through the seminal work of Burstein (1980), Goldstein (1986), Laird and Ware (1982), Longford (1987), Mason, Wong and Entwisle (1983), and many others, the general linear model has been expanded to allow for complex nested data structures. The multilevel model (MLM) can heuristically be expressed as a set of equations operating at two levels. For a continuous measure y assessed on individual i nested within group j, the level-1 equation can be expressed as

(8) $$y_{ij} = \beta_{0j} + \sum_{p=1}^{P} \beta_{pj} x_{pij} + r_{ij} ,$$

where β_{0j} is the level-1 intercept within group j, β_{pj} is the regression of y_{ij} on the p^{th} variable x within group j, and r_{ij} is the residual for individual i within group j. Given that the intercept and slope coefficients vary randomly over group, these can be regressed upon one or more level-2 variables denoted w such that

(9) $$\beta_{0j} = \gamma_{00} + \sum_{q=1}^{Q} \gamma_{0q} w_{qj} + u_{0j} ,$$

(10) $$\beta_{pj} = \gamma_{p0} + \sum_{q=1}^{Q} \gamma_{pq} w_{qj} + u_{pj} ,$$

where the γs represent the fixed coefficients for the regression of the random intercepts and slopes from the level-1 equation (e.g., β_{0j} and β_{pj}) on the level-2 predictor w_j, and u_{0j} and u_{pj} represent the associated level-2 residuals.

This two-level expression is for heuristic purposes only, and the level-2 equation may be substituted into the level-1 equation to create the reduced form expression. In matrix terms we can generally express the level-1 and level-2 equations as

(11) $$\mathbf{y}_{\text{j}} = \mathbf{X}_{\text{j}}\boldsymbol{\beta}_{\text{j}} + \mathbf{r}_{\text{j}},$$

(12) $$\boldsymbol{\beta}_{\text{j}} = \mathbf{W}_{\text{j}}\boldsymbol{\Gamma} + \mathbf{u}_{\text{j}},$$

with reduced form

(13) $$\mathbf{y}_{\text{j}} = \mathbf{X}_{\text{j}}\mathbf{W}_{\text{j}}\boldsymbol{\Gamma} + \mathbf{X}_{\text{j}}\mathbf{u}_{\text{j}} + \mathbf{r}_{\text{j}},$$

where $\mathbf{y}_{\text{j}}$ is the response vector for group $j = 1, 2, ..., J$, $\mathbf{X}_{\text{j}}$ is the design matrix for the set of level-1 predictors (including a column vector of 1s for the intercept), $\mathbf{W}_{\text{j}}$ is the design matrix for the set of level-2 predictors (also including a column vector of 1s for the intercept), $\boldsymbol{\Gamma}$ is the vector of fixed regression coefficients, and $\mathbf{u}_{\text{j}}$ and $\mathbf{r}_{\text{j}}$ are the vectors of level-2 and level-1 residuals, respectively (see Raudenbush & Bryk, 2002, pp. 42-45 for more detail). Importantly, it is assumed that the random effects and residuals are independent and multivariate normally distributed as

(14) $$\mathbf{r}_{\text{j}} \sim N\left(0, \boldsymbol{\Sigma}_{\mathbf{r}_{\text{j}}}\right),$$

$$\mathbf{u}_j \sim N(\mathbf{0}, \mathbf{T}). \tag{15}$$

The covariance matrix of the random effects $\mathbf{T}$ is typically unstructured while the residuals are constrained to be homoscedastic and independent (i.e., $\boldsymbol{\Sigma}_{\mathbf{r}_j} = \sigma^2 \mathbf{I}_{N_j}$), although these specific forms of $\mathbf{T}$ and $\boldsymbol{\Sigma}_{\mathbf{r}_j}$ are often for convenience and are typically not required.

The MLM provides a powerful and flexible analytic framework for testing a variety of interesting questions in the social sciences. Because the nesting in the data is explicitly modeled through the disaggregation of the level-1 and level-2 covariance structures, under general assumptions the model results in accurate standard errors and unbiased coefficients. Importantly, disaggregated effects may be estimated by including predictors in either the level-1 or level-2 parts of the model. These advantages combine to make the MLM an important analytic tool for the applied researcher.

SEM and MLM Approaches to Growth Curve Analysis

A comparison of the reduced form expressions and corresponding assumptions of the multilevel and structural equation models leads one to believe that these are fundamentally different methods of analysis. The assumption of independence of observations is highlighted in the standard estimation of the SEM in that the discrepancy function is based on a single aggregate sample covariance matrix that allows for covariance structures only at a single level of analysis; the covariance structure within any other level of nesting is assumed to be null. In contrast, the estimation of the MLM explicitly incorporates complex data structures through the simultaneous disaggregation of covariance structures among lower (e.g., individuals) and higher (e.g., classrooms) levels of data hierarchy. Accordingly, whereas nested data structures pose a significant problem to standard ML estimation in SEMs, the estimation of the MLM explicitly allows for these dependent data structures.

Despite the seemingly radical differences between the SEM and MLM, we are left with the curious isomorphism between these approaches when estimating a broad class of random effects growth curve models. More specifically, when a two-level data structure arises from the repeated observations of a set of individuals over time (such that time is hierarchically nested within individual), under a broad set of conditions the SEM is analytically equivalent to the MLM (MacCallum et al., 1997; Raudenbush, 2001; Willett & Sayer, 1994). Thus, despite the key differences between the estimation procedures in the SEM and MLM, these two approaches provide analytically identical solutions within the two-level growth model. It is helpful to consider the specific parameterization and estimation of the MLM and SEM growth models to better see the subsequent relations between the two approaches.

Multilevel Growth Models. Bryk and Raudenbush (1987) demonstrated that the MLM provides a powerful method for estimating random effects growth models by approaching the problem from a nested data perspective. That is, if the data were considered to be dependent such that the repeated measures are nested within individuals, the level-1 model captured the within-person relations over time (intra-individual change) and the level-2 model captured the between-person relations (inter-individual differences in intra-individual change).

The level-1 equation for a linear trajectory is given as

$$y_{ti} = \beta_{0i} + \beta_{1i}x_{ti} + r_{ti}, \tag{16}$$

where y_{ti} is the measure of construct y at time t for for individual i, x_{ti} is the measure of time for individual i, β_{0i} and β_{1i} are the intercept and linear slope for individual i, respectively, and r_{ti} is the time-specific and individual-specific residual. Whereas earlier I used subscripts to denote individual i in group j, I now modify this to denote timepoint t nested within individual i. Given this expression, we can write an equation for the random intercept and slope parameters such that

$$\begin{aligned} \beta_{0i} &= \gamma_{00} + u_{0i} \\ \beta_{1i} &= \gamma_{10} + u_{1i} \end{aligned}, \tag{17}$$

where γ_{00} and γ_{10} are the mean intercept and slope values, and u_{0i} and u_{1i} are the individual deviations of each observation from these mean values, respectively. One or more predictors can be incorporated into either the level-1 or level-2 equations to evaluate time-varying and time-invariant predictors of y. As can be seen, the random effects growth model is simply a specific parameterization of the general MLM presented in Equation 13.

Structural Equation Growth Models. Drawing on developments by Tucker (1958) and Rao (1958), Meredith and Tisak (1984, 1990) proposed a method for estimating random effects growth models with the structural equation modeling framework. Whereas the MLM approaches the analysis of growth from a nested data perspective, the SEM approaches this through the use of multiple indicator latent factors. For the linear growth model the measurement equation is

(18) $$y_{it} = \eta_{\alpha_i} + \eta_{\beta_i}\lambda_t + \varepsilon_{it},$$

and the structural equation is

(19) $$\begin{aligned} \eta_{\alpha_i} &= \mu_\alpha + \zeta_{a_i} \\ \eta_{\beta_i} &= \mu_\beta + \zeta_{\beta_i} \end{aligned},$$

Equations 18 and 19 are of the same form as the general SEM expressions given in Equations 1 and 2. Further, note the similarities between the SEM Equations 18 and 19 and the multilevel Equations 16 and 17. Under general conditions, both analytic strategies are approaching the same problem from a different perspective.

However, there remains the clear result that the estimation of the SEM is making an assumption of independence of observations as reflected in the analysis of a single aggregate covariance matrix, yet the repeated measures design naturally gives rise to a two-level data structure with time nested within individual. The key to overcoming this challenge in the estimation of the SEM is to incorporate the level-1 measure of *time* as fixed values within the factor loading matrix $\mathbf{\Lambda}$. So, whereas time is entered as a *predictor variable* in the multilevel model (e.g., x_{ti} in Equation 16), time is entered as the values of the *factor loadings* relating the repeated measures to the underlying latent factors (e.g., λ_t in Equation 18). This strategy allows for the disaggregation of the level-1 and level-2 covariance structures within a single partitioned covariance matrix **S** which is then used as the unit of analysis in the estimation of the SEM. Once defined in this fashion, the latent factors underlying the repeated measures reflect the fixed and random effects associated with stability and change of the repeated measures over time.

The Isomorphism between SEM and MLM Growth Models

I will begin with the MLM for the balanced condition (i.e., where all individuals are assessed at the same time), although we will expand this to the unbalanced condition shortly. In matrix terms, the level-1 equation for the unconditional linear multilevel growth model in Equations 16 and 17 is

(20) $$\mathbf{y}_i = \mathbf{X}\boldsymbol{\beta}_i + \mathbf{r}_i,$$

where the first column of the design matrix **X** contains a vector of 1s to define the intercept and the second column contains the individual-specific measure of time (e.g., $x = 0, 1, ..., T - 1$ for a linear trajectory for T repeated observations). Given the design is balanced, the design matrix **X** is not further subscripted. The level-2 equation is

$$\boldsymbol{\beta}_i = \boldsymbol{\Gamma} + \mathbf{u}_i, \tag{21}$$

in which there is no level-2 design matrix because we are not yet considering predictors of the random effects.[1] The associated reduced form expression is

$$\mathbf{y}_i = \mathbf{X}\boldsymbol{\Gamma} + \mathbf{X}\mathbf{u}_i + \mathbf{r}_i, \tag{22}$$

highlighting that the design is balanced (given no subscript on **X**) and that the same design matrix holds for both the fixed and random components of the model. The covariance structure implied by Equation 22 is

$$\boldsymbol{\Sigma}_{yy} = \mathbf{XTX}' + \boldsymbol{\Sigma}_r, \tag{23}$$

where $\boldsymbol{\Sigma}_r$ and **T** are the covariance matrices of the level-1 and level-2 random effects, respectively. The mean structure implied by Equation 22 is

$$\boldsymbol{\mu}_y = \mathbf{X}\boldsymbol{\Gamma}. \tag{24}$$

Now consider precisely the same model defined as a SEM (e.g., Equations 18 and 19). We again consider the balanced condition, but we will expand this shortly. Here, the measurement (or level-1) equation is given as

$$\mathbf{y}_i = \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i, \tag{25}$$

where the first column of $\boldsymbol{\Lambda}$ is a vector of 1s to define the intercept and the second column contains the specific values of time. The structural (or level-2) equation is given as

$$\boldsymbol{\eta}_i = \boldsymbol{\mu} + \boldsymbol{\zeta}_i, \tag{26}$$

in which there are again no level-2 predictors. Finally, the reduced form equation is

[1] Recall that there is an implicit column vector of 1s to define the intercept in Equation 21, but I do not include this here to explicate that we have no additional level-2 predictors.

(27) $$\mathbf{y}_i = \mathbf{\Lambda\mu} + \mathbf{\Lambda\zeta}_i + \boldsymbol{\varepsilon}_i.$$

Note that, analogous to Equation 22, $\mathbf{\Lambda}$ represents the design matrix for both the random and fixed model effects. Using traditional SEM notation, the covariance structure implied by Equation 27 is

(28) $$\mathbf{\Sigma}(\boldsymbol{\theta}) = \mathbf{\Lambda\Psi\Lambda}' + \mathbf{\Theta}_{\varepsilon},$$

and the mean structure is

(29) $$\boldsymbol{\mu}(\boldsymbol{\theta}) = \mathbf{\Lambda\mu}.$$

Note that Equations 28 and 29 are sub-parameterizations of the more general SEM shown in Equations 5 and 6. Also note the symmetry between Equations 23 and 28, and between Equations 24 and 29. It can be seen that for the two-level growth model described here, the multilevel and structural equation models are related such that

(30) $$\begin{aligned} \mathbf{X} &= \mathbf{\Lambda} \\ \mathbf{T} &= \mathbf{\Psi} \\ \mathbf{\Sigma}_r &= \mathbf{\Theta}_{\varepsilon} \\ \mathbf{\Gamma} &= \boldsymbol{\mu} \end{aligned}$$

Whereas in the MLM, time is entered as a predictor variable in the design matrix $\mathbf{X}$, in the SEM time is entered through the fixing of parameters in the factor loading matrix $\mathbf{\Lambda}$. The associated matrices defining the fixed and random effects are identical (Bauer, in press). Here lies the isomorphism.

As I noted earlier, various aspects of these relations have been explored by Bauer and Curran (2002), Neale et al. (1999), Newsom (2002), Rovine and Molenaar (1998, 2000, 2001), and most thoroughly by Bauer (in press). The question that intrigues me here is, *Can we use this strategy to estimate a broader class of multilevel models with the SEM framework?* The brief answer is yes, and the general approach we will use is to code the values of the level-1 predictors into the factor loading matrix within the SEM. However, an important initial question to ask is, *Why would we want to trouble ourselves to do this at all?* The MLM is unquestionably a powerful and well developed method that performs exceptionally well across a variety of applied research settings. Why work to replicate this model within the SEM framework? I suggest two reasons. First, I believe it is important to show these equivalencies to better understand specific aspects of each of these modeling strategies in isolation. Second, I believe that the SEM may provide advantages over the current MLM under certain empirical conditions. I will now present a detailed exploration of these models using an empirical data example.

Fitting Multilevel Models Using the SEM Framework: The Balanced Case

I will use empirical data drawn from the High School and Beyond (HSB) study to demonstrate the estimation of multilevel models within SEM. The HSB data has been used to demonstrate multilevel models elsewhere (Raudenbush & Bryk, 2002; Singer, 1998), and I use these same data here. The data consist of a total of N_T = 7185 students nested within J = 160 classrooms with class sizes ranging from N_j = 14 to N_j = 67 with a median class size of 47. The outcome of interest is a standardized math achievement test with a grand mean of $\bar{y}$ = 12.75, and grand standard deviation of sd = 6.88. I will consider two level-1 predictors of math achievement (child gender and child minority status) and one level-2 predictor (school socioeconomic status, or SES).

The Balanced Condition. As a starting point, I extracted a subset of the full HSB data to result in a balanced data structure (i.e., equal numbers of students within all classrooms). Although restrictive, I will fully expand these models to the unbalanced condition later. I randomly selected N_j = 8 students nested within J = 53 classrooms for a total sample of N_T = 424 children. I then created a data matrix **X** that was of dimensions J by N_j (i.e., 53 by 8) in which each row represented a given classroom with each row-wise cell representing the eight children within that classroom. The general form of this data matrix is presented in Table 1.

Table 1
The J = 53 by N_J = 8 Raw Data Matrix **X** which is the Unit of Analysis for the Unconditional Random Effects Regression Models Estimated in SEM

	child 1	child 2	child 3	child 4	child 5	child 6	child 7	child 8
class 1	$y_{1,1}$	$y_{1,2}$	$y_{1,3}$	$y_{1,4}$	$y_{1,5}$	$y_{1,6}$	$y_{1,7}$	$y_{1,8}$
class 2	$y_{2,1}$	$y_{2,2}$	$y_{2,3}$	$y_{2,4}$	$y_{2,5}$	$y_{2,6}$	$y_{2,7}$	$y_{2,8}$
class 3	$y_{3,1}$	$y_{3,2}$	$y_{3,3}$	$y_{3,4}$	$y_{3,5}$	$y_{3,6}$	$y_{3,7}$	$y_{3,8}$
•	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•
class 53	$y_{53,1}$	$y_{53,2}$	$y_{53,3}$	$y_{53,4}$	$y_{53,5}$	$y_{53,6}$	$y_{53,7}$	$y_{53,8}$

Although this is a data matrix of the math achievement scores for all $N_T = 424$ children, this is a very different matrix than those with which we typically work in SEM. Specifically, as in the standard SEM, each row is exchangeable with all other rows; however, in this case every cell within each row is exchangeable with all other cells within that same row. That is, the ordering of the cells is *completely arbitrary* within any given row. This is similar to the arbitrary denoting of "twin 1" and "twin 2" in studies of twin pairs (e.g., Neale & Cardon, 1992). Thus, in Table 1 the first row represents the eight children within the first classroom, but the order of these eight children within the first row is inconsequential. This data matrix will be the unit of analysis for our first multilevel model.

We will use the standard ML estimator in SEM to fit our first two-level model. To do so, we must calculate our usual covariance matrix and mean vector from our data matrix **X** for use in Equation 7. However, like the raw data matrix above, the covariance matrix and mean vector are of a particularly peculiar sort. Because we are considering a balanced design, $N_j = 8$ for all $J = 53$ classrooms. The covariance matrix **S** is thus of dimension N_j by N_j representing the number of children within each classroom. Similarly, the mean vector $\bar{\mathbf{y}}$ is of dimension N_j by 1, also representing the number of children within each classroom. The general structure of this matrix and vector is given in Equation 31.

$$(31)\qquad \mathbf{S}=\begin{pmatrix} s_{11} & & & & & & & \\ s_{21} & s_{22} & & & & & & \\ s_{31} & s_{32} & s_{33} & & & & & \\ s_{41} & s_{42} & s_{43} & s_{44} & & & & \\ s_{51} & s_{52} & s_{53} & s_{54} & s_{55} & & & \\ s_{61} & s_{62} & s_{63} & s_{64} & s_{65} & s_{66} & & \\ s_{71} & s_{72} & s_{73} & s_{74} & s_{75} & s_{76} & s_{77} & \\ s_{81} & s_{82} & s_{83} & s_{84} & s_{85} & s_{86} & s_{87} & s_{88} \end{pmatrix} \qquad \bar{\mathbf{y}}=\begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \bar{y}_3 \\ \bar{y}_4 \\ \bar{y}_5 \\ \bar{y}_6 \\ \bar{y}_7 \\ \bar{y}_8 \end{pmatrix}$$

Prior to fitting our first SEM, consider the elements of **S** and $\bar{\mathbf{y}}$. The first diagonal element of **S** (e.g., s_{11}) represents the variance of the math achievement scores for the students who are (arbitrarily) ordered first within each of the 53 classrooms, the second diagonal element represents those ordered second, and so on. Further, the first off-diagonal element of **S** (e.g., s_{21}) represent the covariance of math achievement scores for the students who are ordered second within each class with those who are ordered first within each class, and so on. Finally, the first element of the mean vector $\bar{\mathbf{y}}$ (e.g., $\bar{y}_1$) represents the mean math achievement of the students who are ordered first within each class, the second element for those ordered second in each class, and so on. As you can see, this is an extremely odd covariance matrix and mean vector from a traditional SEM perspective.

This becomes odder still when you consider that the order of children within classroom is arbitrary, meaning that there are an extremely large number of distinctly unique but equally valid covariance matrices and mean vectors that will all result in identical model results. Given there are N_j children within J classrooms, there are $(N_j!)^J$ possible combinations of children within each classroom. For $J = 53$ and $N_j = 8$ there are many millions of possible orderings of children within and across rows. This in turn results in many millions of unique covariance matrices and mean vectors, each of which contain sufficient information to equally fit our MLMs of interest. What we will see is that although the sample $\mathbf{S}$ and $\bar{\mathbf{y}}$ are seemingly arbitrary, we will impose the necessary restrictions needed to estimate the MLM through the parameterization of the measurement model via $\mathbf{\Lambda}$ in the SEM.

I will begin by estimating a standard random effects ANOVA model within the SEM. This is simply estimating a random mean at level-1, and allowing these means to vary randomly over classrooms (e.g., Raudenbush & Bryk, 2002, p. 23). Recall that the key to the equivalency between the SEM and MLM was that we incorporate any level-1 predictors into the factor loading matrix in the SEM. Following this strategy here, in order to define the random effects ANOVA model as an SEM, we require a column vector of 1's at level-1. To accomplish this in the SEM, we simply define a single latent factor upon which all eight indicators of math achievement load with a value fixed to 1. This is presented in Figure 1.

There are three parameters to be estimated in this SEM. There is a single residual that is equated across all eight items (where *item* represents *child*); this is equated given the arbitrary ordering of children within classroom, and this represents the level-1 random effect (e.g., $\hat{\sigma}^2$). There is a single variance of the latent factor, and this represents the level-2 random effect (e.g., $\hat{\tau}_{00}$). Finally, there is a mean of the latent factor (denoted by the triangle), and this represents the single fixed effect (i.e., $\hat{\gamma}_{00}$).

I estimated the model presented in Figure 1 using the standard maximum likelihood estimator based on the single covariance matrix and mean vector summarizing the data arranged as in Table 1. The log likelihood for this model was $LL = 2834.05$. The mean (and standard error) of the latent factor was 11.167 (.412), the factor variance was 3.494 (1.794), and the residual variance was 44.025 (3.232) resulting in an intraclass correlation (ICC) of 3.494/(44.024 + 3.494) = .074 indicating an appreciable degree of nesting in the data. For comparison, I replicated this same random effects ANOVA model under the standard MLM framework using restricted ML. All of the results from the MLM

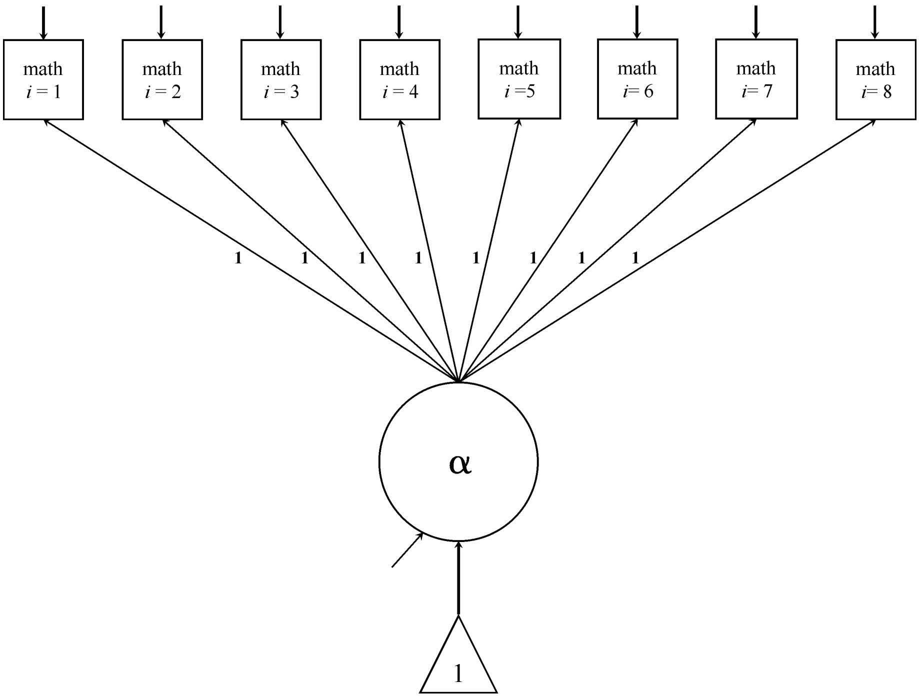

Figure 1
Path Diagram for the SEM Estimation of a Random Effects ANOVA

matched those from the SEM to the third decimal. This empirically supports the analytical results showing that the standard SEM based on a single covariance matrix and mean vector results in a true disaggregation of level-1 and level-2 random effects when parameterizing the level-2 effects as latent factors. Although I used *Mplus* (Muthén & Muthén, 1998) for the SEM analyses and SAS PROC MIXED (SAS Inc., 2000) for the MLM analyses, all of the results I present here would be replicated using any comparable software package.

Let us now turn to the inclusion of a single level-1 predictor variable, the minority status of the child. This was coded 0 if the child did not self-report as a member of a minority group and was coded 1 if they did report being a member of a minority group. The question of interest is whether child minority status is related to math achievement, and if the magnitude of this effect varied over classroom. To estimate this model within the SEM, we must not only define a latent factor for the random intercept, but we must also define a latent factor for the random slope associated with minority status. To do this, we must modify the data matrix so that order becomes more important. The raw data matrix is thus the same as that presented in Table

1 with the additional restriction that the first four children listed in a given row are minority status 0, and the second four are minority status 1. I intentionally selected the subset of the HSB data so that minority status was balanced within classroom, but I will relax this restriction later.

It is important to note that we have no information about minority status *within* the data matrix **X**; that is, there is no measured variable that defines minority status for each child. Instead, minority status is solely represented in terms of the *ordering* of the observations within each row, and we will capitalize on this ordering to define a latent factor to represent this effect in the SEM. Also note that in the random effects ANOVA model the order of children within a row was arbitrary and thus children could be considered fully exchangeable. However, here we have *conditional* exchangeability such that order is arbitrary *within* minority status, but not across minority status. So the first four children are ordered arbitrarily, but these must remain in the first four elements of the first row. Outside of this ordering of cases, no new information is present in this data matrix compared to the earlier one. This is because we will incorporate the values of the level-1 predictors into the SEM via the factor loading matrix. This multilevel model is presented in Figure 2.

Figure 2 highlights several key additions to the previous model. First, it is important to realize that although we are now including minority status as a level-1 predictor, we are analyzing *exactly the same* covariance matrix and mean vector as we used before. It is not required that we use this same covariance matrix, and we could equivalently use any of a large number of potential matrices within the restrictions of conditional exchangeability. However, we can use the same matrix as before given that the values of minority status are not embedded in the data set, but are instead entered into the model via the factor loading matrix that defines a second latent factor. More specifically, we have included a new latent factor representing the fixed and random effects associated with minority status as a level-1 predictor. The factor loadings are set equal to 0 for children with a minority status of 0, and these are set equal to 1 for children with a minority status of 1.

It is now clear why it was important to order children so that the non-minority children were listed first followed by the minority children. There are two fixed effects (a mean of each latent factor), one level-1 random effect (the item-level residual variance equated over items), two level-2 random effects (the variance of each latent factor), and the covariance between the level-2 random effects (the covariance between the two latent factors). This parameterization represents a random regression model with one level-1 predictor (minority status) and no level-2 predictors. The SEM and MLM results are presented in Table 2.

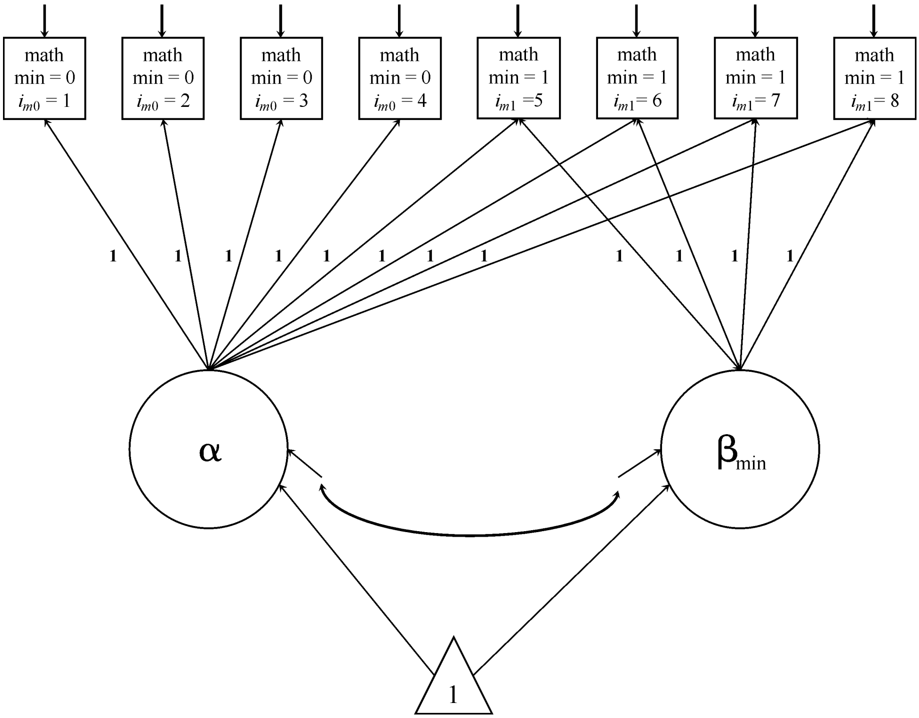

Figure 2
Path Diagram for the SEM Estimation of a Random Effects Regression with Minority Status as the Sole Level-1 Predictor

Table 2
SEM and MLM Results for Random Regression with Child Minority Status as the Sole Level-1 Predictor

	SEM	MLM
$\hat{\mu}_{\alpha}$	13.530 (.481)	13.530 (.481)
$\hat{\mu}_{\beta_{min}}$	–4.725 (.654)	–4.725 (.654)
$\hat{\psi}_{\alpha}$	3.197 (2.493)	3.197 (2.493)
$\hat{\psi}_{\beta_{min}}$	4.516 (4.638)	4.516 (4.638)
$\hat{\psi}_{\alpha,\beta}$	.127 (2.701)	.127 (2.701)
$\hat{\theta}_{\varepsilon}$	36.355 (2.883)	36.355 (2.883)

As expected, the SEM and MLM results are identical. Of key interest is the significant fixed effect of minority status indicating that, on average, minority children report standardized math tests scores that are 4.7 units lower than non-minority children. This is reflected in the fixed-effect regression in the multilevel model and the mean of the minority latent factor in the SEM. Further, there is not a significant random effect associated with minority status indicating that the magnitude of the effect does not vary significantly over classroom. It is important to note that this random effect is represented as the latent factor variance in the SEM, and the covariance between the random intercept and slope is represented in the corresponding factor covariance.

Finally, I expanded this model to include the combined level-1 effects of child minority status and child gender (where gender equal to 0 reflects girls and gender equal to 1 reflects boys). As before, we need not include additional information into the data matrix given that we will enter the values of gender into the factor loading matrix. However, we must impose one more degree of ordering in our raw data matrix. Specifically, within minority status equal to 0, the first two children must be girls, and the second two boys; and within minority equal to 1, the first two children must be girls and the second two must be boys. There remains exchangeability of children within row, but now only within gender ordered within minority. For example, the two girls that are minority of 0 can be exchanged with one another, but these two cannot appear anywhere else in that particular row. We thus continue to work with the same raw data matrix presented in Table 1, but with the realization that within a given row the first two cells represent math achievement scores for female non-minorities, the next two for male non-minorities, the next two for female minorities, and the final two for male minorities. The is no variable measuring gender or minority; we will define this via $\mathbf{\Lambda}$.

We will again analyze the same covariance matrix and mean vector as before, but we are now going to consider the inclusion of two level-1 variables, each of which has a fixed and a random effect. The corresponding SEM needed to estimate this model is presented in Figure 3. Note that the factor loadings for the intercept term are set to 1 for all indicators, the loadings are set to 1 for indicators representing males, and the loadings are set to 1 for indicators representing minorities. Each latent factor mean and variance represents the fixed and random effects, respectively. The results of this model are presented in Table 3. Again, all results are identical between the SEM and MLM estimation of this model. The fixed effects reflect that the mean of math achievement is 12.994 for non-minority female children, and that there is a 1.07 increment for males and a 4.725 decrement for minorities. Finally, there is no significant random variability in any of these effects ($p < .05$). To reiterate,

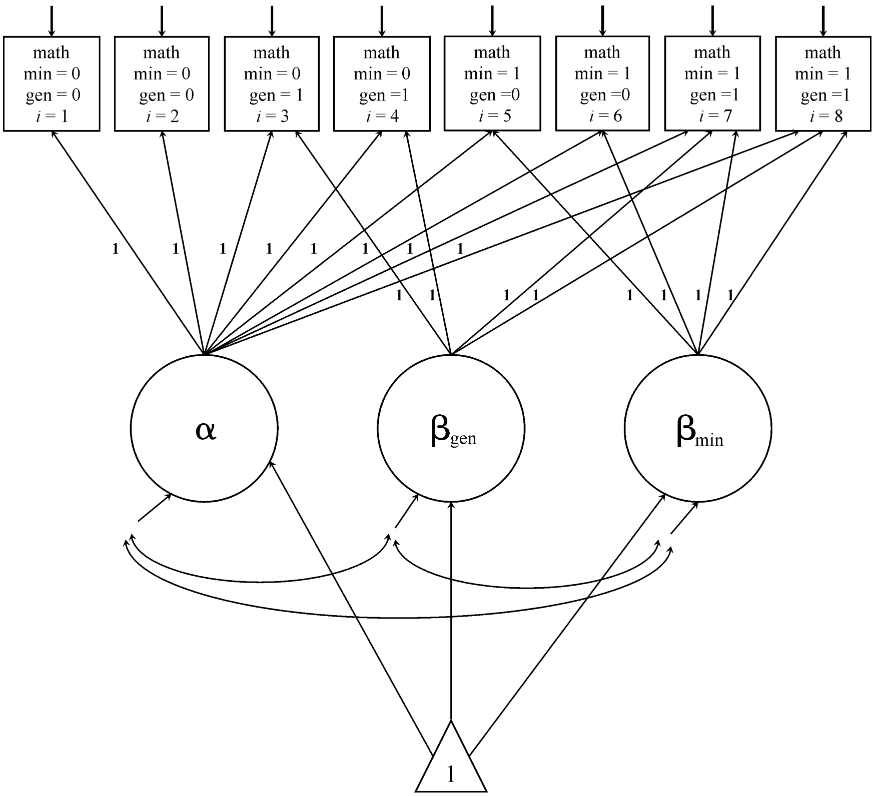

Figure 3
Path Diagram for the SEM Estimation of a Random Effects Regression with Minority Status and Gender as the Two Level-1 Predictors

Table 3
SEM and MLM Results for Random Regression with Child Minority Status and Child Gender as the Two Level-1 Predictors

	SEM		MLM	
$\hat{\mu}_{\alpha}$	12.994	(.604)	12.994	(.604)
$\hat{\mu}_{\beta_{gen}}$	1.07	(.641)	1.07	(.641)
$\hat{\mu}_{\beta_{min}}$	-4.725	(.654)	-4.725	(.654)
$\hat{\psi}_{\alpha}$	6.387	(3.916)	6.387	(3.916)
$\hat{\psi}_{\beta_{gen}}$	4.531	(4.485)	4.531	(4.485)
$\hat{\psi}_{\beta_{min}}$	5.463	(4.656)	5.463	(4.656)
$\hat{\theta}_{\varepsilon}$	34.463	(2.994)	34.463	(2.994)

this is a fully random regression model estimated as a standard SEM based upon a single covariance matrix and a single mean vector.

Continuous Level-1 Predictors. Up to this point I have considered only level-1 predictors that are dichotomous, thus making the integration of these effects quite easy via the factor loading matrix. There are of course many instances in which we would like to consider continuous level-1 predictor variables as well. This can be fully accomplished in the SEM, although several challenges are quickly encountered. Recall that the general rule is that all possible values of the level-1 predictor must be defined in the factor loading matrix. With a dichotomous measure this is easily accomplished because there are only two potential values on which the variable was observed. However, for a continuous measure this can become exceedingly tedious and introduces an interesting problem with missing data (e.g., it is not likely that all values of the level-1 continuous covariate may be observed across all classrooms). However, through the use of *definition variables* (as is currently available in Mx, Neale et al., 1999; see also Mehta & West, 2000), these problems can be surmounted. The key to this approach is the use of individual factor loading matrices, where here "individual" is serving as "classroom". More specifically, a factor loading matrix would be defined that is unique to each classroom j; the elements of this matrix would contain each observed value of the continuous level-1 covariate. Finally, the discrepancy function would be fitted to the model by pooling over all classroom-specific matrices. Given space constraints, I do not demonstrate the inclusion of continuous level-1 covariates here, but see Bauer (in press) and Neale et al. (1999, p. 133) for empirical examples of this approach.

Level-2 Predictors. The models I have described above have not included any level-2 predictor variables. That is, I have considered two characteristics associated with the individual children, and we have allowed the magnitude of these effects to vary over level-2 units, but we have not yet included predictors at the level of the classroom. Given that we have defined latent factors to represent the fixed and random effects of the level-1 predictors, we can simply regress these latent factors on our level-2 measures of interest. To accomplish this, we must now augment our data matrix $\mathbf{X}$ with an additional column to incorporate the level-2 covariates of interest.[2]

To demonstrate this model, I will consider a single classroom level predictor representing the mean socioeconomic status (SES) of the school.

[2] Given that we will simply regress the latent factors on level-2 predictors, these predictors may be categorical or continuous; the challenges encountered when incorporating continuous level-1 predictors in the SEM do not apply to level-2 predictors.

Because there is only one classroom per school, school SES is a classroom-level measure. The structure of the new data matrix **X** is presented in Table 4. As can be seen, the data matrix is identical to that used before with the exception that there is an added column representing school SES. Given that each row represents one classroom, the first eight cells continue to represent the math scores for the eight children nested within that classroom, but the ninth cell represents the school measure of SES for that classroom.

Based on the this data matrix **X**, we again compute our standard covariance matrix and mean vector, but this continues to exhibit several curious properties. For example, the portion of **S** related to the students within classroom is the same as before (e.g., Equation 31); however, the newly added diagonal element for school SES (denoted s_{ww}) represents the sample variance of SES across the J = 53 schools. Further, the off-diagonal elements associated with w (e.g., $s_{j,w}$) represents the covariance of the N_j child in classroom j with the level-2 covariate w. This is now a partitioned matrix containing information about the covariance structure of our measures within level-1, within level-2, and between level-1 and level-2; we are thus disaggregating the multilevel covariance structure for estimation within the SEM framework. As before, this matrix has little intrinsic meaning until we parameterize the level-1 measures via the factor loading matrix in the SEM.

To demonstrate this, I estimated a SEM to define a random regression with two level-1 predictors (child minority status and gender) in which all effects were random, and I regressed these random effects on a single level-

Table 4
The Raw Data Matrix **X** Ordered First by Minority and Within Minority by Gender and is Augmented with a Column for the Level-2 Predictor School SES

	child 1 min = 0 gen = 0	child 2 min = 0 gen = 0	child 3 min = 0 gen = 1	child 4 min = 0 gen = 1	child 5 min = 1 gen = 0	child 6 min = 1 gen = 0	child 7 min = 1 gen = 1	child 8 min = 1 gen = 1	School SES (w_1)
class 1	$y_{1,1}$	$y_{1,2}$	$y_{1,3}$	$y_{1,4}$	$y_{1,5}$	$y_{1,6}$	$y_{1,7}$	$y_{1,8}$	$w_{1,1}$
class 2	$y_{2,1}$	$y_{2,2}$	$y_{2,3}$	$y_{2,4}$	$y_{2,5}$	$y_{2,6}$	$y_{2,7}$	$y_{2,8}$	$w_{2,1}$
class 3	$y_{3,1}$	$y_{3,2}$	$y_{3,3}$	$y_{3,4}$	$y_{3,5}$	$y_{3,6}$	$y_{3,7}$	$y_{3,8}$	$w_{3,1}$
•	•	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•	•
class 53	$y_{53,1}$	$y_{53,2}$	$y_{53,3}$	$y_{53,4}$	$y_{53,5}$	$y_{53,6}$	$y_{53,7}$	$y_{53,8}$	$w_{53,1}$

2 measure (school SES). A path diagram of this model is presented in Figure 4, and the results of the SEM and MLM results are presented in Table 5. As expected, the results are identical for the SEM and MLM strategies. There is a significant prediction of the random intercept of math achievement as a function of school SES such that higher SES is associated with higher mean math scores for female, non-minority students (resulting from the coding used for the level-1 effects). Further, there is no significant prediction of either the gender or minority effects, but this too is expected given no evidence was found for variability in these effects in the earlier model that contained no level-2 predictors. These results demonstrate that we can use the standard SEM with a single covariance matrix and mean vector to perfectly replicate a multilevel model containing both level-1 and level-2 predictors and random effects at both levels for the balanced case.

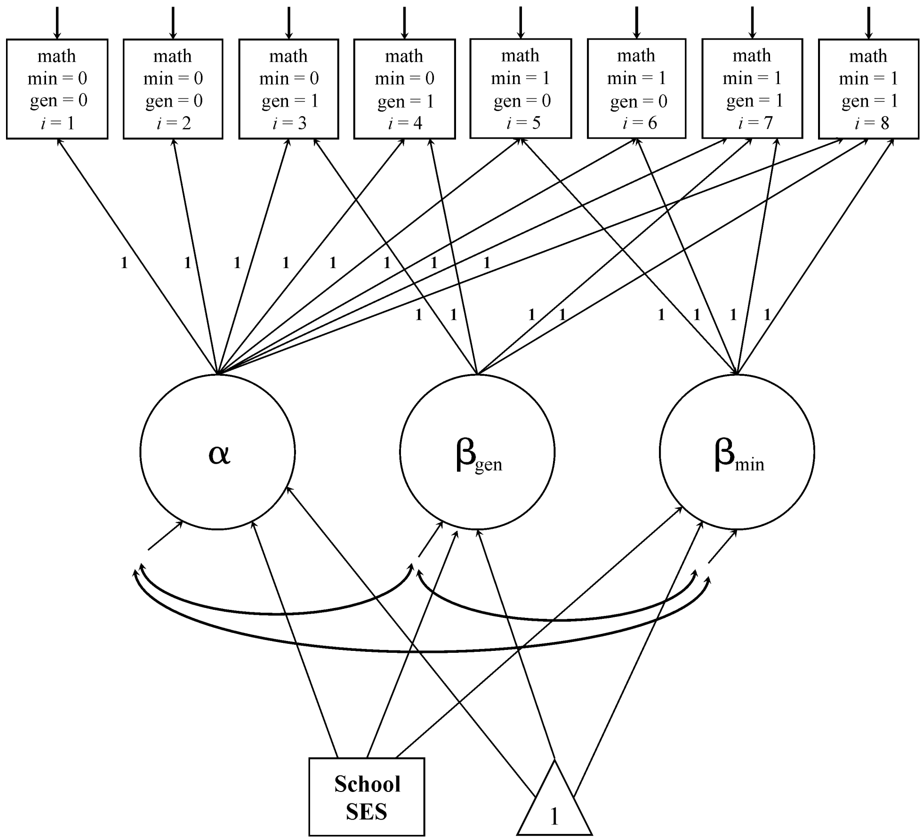

Figure 4
Path Diagram for the SEM Estimation of a Fully Multilevel Model with Two Level-1 Predictors and One Level-2 Predictor

Table 5
SEM and MLM Results for the Gender, Minority Status, and School SES

	SEM		MLM	
$\hat{\mu}_{\alpha}$	13.118	(.567)	13.118	(.567)
$\hat{\mu}_{\beta_{gen}}$	1.085	(.642)	1.085	(.642)
$\hat{\mu}_{\beta_{min}}$	–4.703	(.655)	–4.703	(.655)
$\hat{\gamma}_{\alpha}$	4.069	(1.494)	4.069	(1.494)
$\hat{\gamma}_{\beta_{gen}}$	.476	(1.692)	.476	(1.692)
$\hat{\gamma}_{\beta_{min}}$	.743	(1.726)	.743	(1.726)
$\hat{\theta}_{\varepsilon}$	34.463	(2.994)	34.463	(2.994)

Fitting Multilevel Models Using the SEM Framework: The Unbalanced Case

Up to this point I have considered a highly contrived balanced situation in which the within-class sample size is equal over all classrooms, and there were exactly half of the children within each class who were female and exactly half who were self-identified minority. I did this to orient to the general problem and to demonstrate that under these conditions the multilevel model could be estimated as a SEM using the standard ML estimator based on a single aggregate covariance matrix and mean vector. Of course these contrived conditions would rarely be encountered in practice. However, with some basic extensions we can continue to use the SEM framework to estimate more complex (and more realistic) multilevel models.

To expand the SEM to the unbalanced case, we must briefly turn back to the ML fit function defined earlier in Equation 7. As is explicated in this equation, the fit function is based upon a single sample covariance matrix and a single sample mean vector. In the fully balanced case, these are sufficient statistics for estimation of the models discussed thus far. However, in the unbalanced case we can no longer use a single covariance matrix and mean vector given that the level-1 sample size varies across level-2 units; these aggregate covariance matrices and mean vectors no longer represent sufficient statistics for model estimation. However, we can address this issue through the use of direct maximum likelihood (direct ML).

The direct ML estimator has been used extensively in standard SEM analyses to properly model data in which some portion is missing (see Allison, 2001, and Schafer & Graham, 2002, for details). We can make use of the direct ML estimator here because we can conceptualize the unbalanced nested design as a form of missing data. For example, if one classroom were to have N_j = 15 children and a second classroom were to have N_j = 10 children, we can consider the difference of five children to be missing in the second classroom. Of course it is not missing in the literal sense of the word, but it is in the eyes of the estimator.[3] This approach will allow us to estimate the multilevel model as a SEM for unbalanced designs.

The direct ML fit function used in the general SEM can be expressed as

$$F_{DML} = \sum_{i=1}^{N} \log |\Sigma_{i,mm}| + \sum_{i=1}^{N} (y_{i,m} - \mu_{i,m})' \Sigma_{i,mm}^{-1} (y_{i,m} - \mu_{i,m}) \tag{32}$$

where m denotes the measured (or observed) data taken on individual i within a sample of size N (Wothke, 2000, Equation 1; see also Arbuckle, 1996). However, for our use here, each row of the data matrix refers to a given *classroom*, and each element within a given row refers to a specific *child* nested within that classroom. Thus, in the situation in which we estimate a multilevel model as a SEM, the direct ML fit function defined in Equation 32 allows for unbalanced designs when we consider $i = j$ and $N = J$ indicating that the level of the individual is in actuality the classroom, and the pattern of "missingness" is in actuality the unbalanced design in which sample size varies over classroom. I will use this approach to estimate a random effects regression within the SEM based on the full HSB data set.

Random Effects ANOVA Model. To demonstrate this approach, I will re-estimate several of the earlier models, but will consider the entire HSB data set and not just the highly structured subsample. Unfortunately, here we begin to pay the reaper in that the data management aspect of this endeavor is becoming increasingly tedious. In order to estimate the unbalanced multilevel model using the direct ML estimator in SEM, we must first identify the largest level-1 sample size over level-2 units because this will define the total number of columns in the data matrix. For the HSB data, we find that the maximum level-1 sample size is N_j = 67.

Thus, for our initial models we will compute a J = 160 by N_j = 67 data matrix **X** in which the data for each of J = 160 classrooms is contained in a given row, and the varying sample sizes across classrooms is expressed as "missing" for any number of observations fewer than N_j = 67. For example, if classroom j = 5 were to have N_5 = 10 children, the first 10 elements of row 5 in data matrix **X** would contain the individual children's math achievement

[3] Briefly allowing the rather spooky notion of estimators having eyes.

scores, and the remaining 67 – 10 = 57 elements would be denoted as missing. This general data structure is presented in Table 6. Because order of classroom is arbitrary, I place the largest class as the first row simply to demonstrate that one classroom has N_j = 67 children. However, all other classrooms will have only as many columns as children, and all remaining columns are missing (denoted in the table with cross-hatching).[4] We will now use this data matrix with the direct ML estimator to replicate our earlier models but using the full HSB data set.

The key difference from the earlier models is that we now must define our latent factors with 67 individual indicators reflecting that this is the largest level-1 sample size. Thus, if we were to replicate the random effects ANOVA model using the full HSB sample, we will define a single latent factor with 67 indicators, all of which were related to the factor with loadings set equal to 1. I do not present the model results here but, as predicted, the results from this SEM using the direct ML estimator precisely equal those of the standard multilevel model.

Table 6
The $J = 160$ by $N_j = 67$ Raw Data Matrix **X** Where Classroom Size is not Balanced

	child 1	child 2	child 3	child 4	•	•	child 66	child 67
class 1	$y_{1,1}$	$y_{1,2}$	$y_{1,3}$	$y_{1,4}$	•	•	$y_{1,66}$	$y_{1,67}$
class 2	$y_{2,1}$	$y_{2,2}$	$y_{2,3}$	$y_{2,4}$	•	•	$y_{2,66}$	
class 3	$y_{3,1}$	$y_{3,2}$	$y_{3,3}$					
•	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•
class 160	$y_{160,1}$	$y_{160,2}$						

[4] Because the ordering of cells within a given row is completely arbitrary, these can be written out in any of a large number of ways. For this example I listed the data starting from left to right for half of the rows and starting from right to left for the other half (e.g., for a classroom with 10 students, I would list the math scores in columns 57 through 67). This not only highlights the arbitrariness of the ordering, but also allowed for more efficient estimation of the models given less spareseness of row and column coverage within the raw data matrix.

A Single Level-1 Predictor. It becomes more complicated to incorporate level-1 predictors for the unbalanced condition, but only modestly so. Whereas for the random effects ANOVA model we only had to identify the maximum sample size across all classrooms, we now have to consider the maximum sample size within each value of the level-1 predictor of interest. For example, to include minority status as a level-1 predictor using the full HSB data set, we must again order the data so that all of the children of minority status equal to 0 appear first followed by all of the children of minority status equal to 1. However, because we are working with an unbalanced design, the specific number of minority children in each classroom varies over classrooms. To account for this, we must identify the largest number of minority status equal to 0 plus the largest number of minority status equal to 1, and this combined value identifies the total number of columns of our data matrix **X**.

In the full HSB data set we find that the largest number of children of minority status equal to 0 in a given classroom is $N_{m=0} = 66$, and the largest number of children of minority status equal to 1 in a given classroom is $N_{m=1} = 64$. Thus, our data matrix contains $J = 160$ rows and 66 + 64 = 130 columns. This data structure is shown in Table 7. We can now fit a two-factor SEM to this data matrix **X** in which there are 130 indicators, all of which load on the random intercept factor with factor loadings set equal to 1, and indicators 67 through 130 load on the minority status factor with factor loadings set equal to 1. I do not present these results here but, again as predicted, all parameter estimates and standard errors from the SEM with direct ML estimation are precisely equal to those obtained from a standard multilevel analysis of the same model.

Table 7
The J = 160 by N_J =130 Raw Data Matrix **X** for the Full HSB Data Set

	$n_0 = 66$				$n_1 = 64$			
	child 1 min = 0	child 2 min = 0	•••	child max min = 0	child 1 min = 1	child 2 min = 1	•••	child max min = 1
class 1	$y_{1,1(0)}$	$y_{1,2(0)}$	•••	$y_{1,\max(0)}$	$y_{1,1(1)}$	$y_{1,2(1)}$	•••	$y_{1,\max(1)}$
class 2	$y_{2,1(0)}$	$y_{2,2(0)}$	•••	$y_{2,\max(0)}$	$y_{2,1(1)}$	$y_{2,2(1)}$	•••	$y_{2,\max(1)}$
class 3	$y_{3,1(0)}$	$y_{3,2(0)}$	•••	$y_{3,\max(0)}$	$y_{3,1(1)}$	$y_{3,2(1)}$	•••	$y_{3,\max(1)}$
•	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•
class 160	$y_{160,1(0)}$	$y_{160,2(0)}$	•••	$y_{160,\max(1)}$	$y_{160,1(1)}$	$y_{160,2(1)}$	•••	$y_{160,\max(1)}$

A Full Multilevel Model with Both Level-1 and Level-2 Predictors. Finally, we can include one or more level-2 predictors just as we did before. We simply augment the data matrix **X** in Table 7 with a single column (here, column number 131) which denotes the value of school SES for each of the $J = 160$ individual schools. We then define the random intercept and random minority status factors as before, and regress these two factors on our level-2 measure of school SES. This model is presented in Figure 5. Again, I do not present the full results here. As analytically predicted, the parameter estimates and standard errors from the SEM estimation of the unbalanced nested model with both level-1 and level-2 predictors are precisely equal to those obtained using standard MLM analysis of the same model. This model could be extended to include additional level-1 or level-2 predictors following the strategies outlined above.

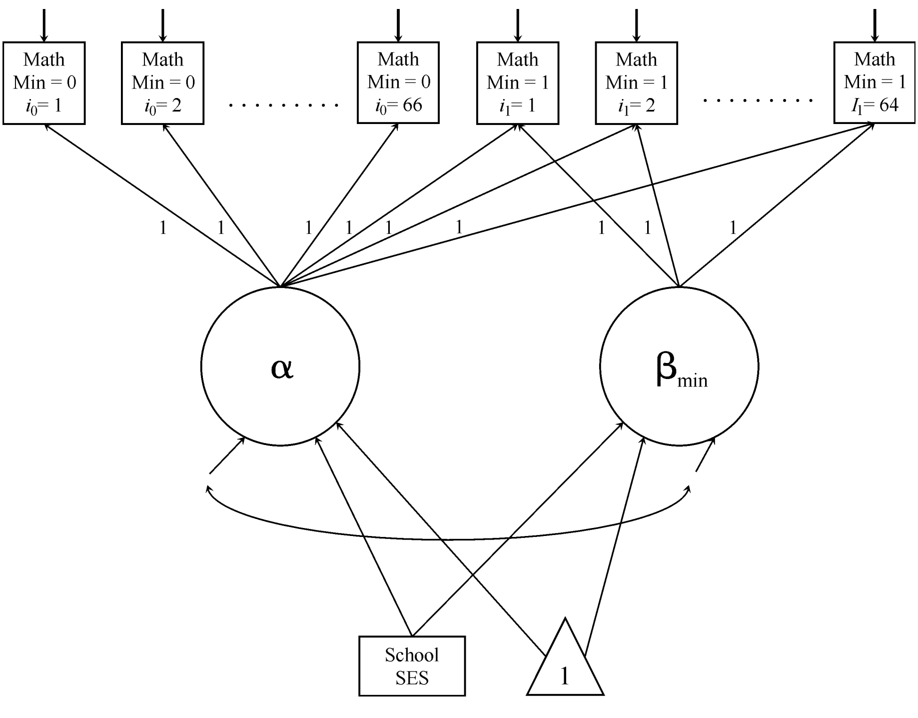

Figure 5
Path Diagram for the SEM Estimation of the Full HSB Data Set with 130 Indicators to Define the Intercept and Minority Latent Factors

Why Bother?

I have demonstrated both analytically and empirically that a two-level multilevel model with both level-1 and level-2 predictors and fixed and random effects can be equivalently estimated within the standard SEM framework. If the design is balanced, the multilevel model can be estimated using ML based on a single covariance matrix and mean vector. If the design is unbalanced, the multilevel model can be estimated using full information ML allowing for unequal level-1 sample size across level-2 units. Finally, discrete level-1 predictors can be included through specific definitions of the factor loading matrix, continuous level-1 predictors can be included through the use of definition variable methodology, and any form of level-2 predictors can be included as predictors of the latent factors.

Given these analytical and empirical results, I make the following proposal: *Any two-level linear multilevel model can be estimated as a structural equation model given that this is essentially a data management problem*; thus the intentionally lighthearted title of my article. However, in the spirit of never getting something for nothing, for any multilevel model of any reasonable complexity, estimating this as a SEM becomes a remarkably complex, tedious, and error-prone task. A quite reasonable question then is, *Why bother*? Given that I can fully estimate these multilevel models in any number of elegant software packages dedicated to such an endeavor, why should I consider these complex SEMs that accomplish the same thing? I see two possible answers to this question.

First, I believe that understanding how a multilevel model might be estimated within the structural equation modeling framework helps us to better understand both approaches to model estimation. I am in no way advocating the widespread estimation of multilevel models using SEM software. I do hope, however, that the delineation of how these two modeling strategies are related to one another helps us to better consider each approach in isolation. I believe this alone is worth the effort.

However, I believe there is a second advantage to such a consideration as well. If we can replicate a multilevel model using the standard SEM framework, might we then potentially capitalize upon other strengths of the general SEM to improve upon our multilevel analyses? I believe that we can indeed accomplish this, and this may in turn improve our methods for analyzing data from both the SEM and MLM perspectives. I will conclude with a description of several of these potential applications that I find particularly intriguing.

Potential Expansions of the Multilevel Model via SEM

Multiple Indicator Level-2 Predictors. Recall that, just as in the standard ordinary least squares regression model, the MLM makes the assumption that all predictors are error free. Given the use of manifest variables in the SEMs presented above, this same assumption is being made here as well. Violation of this assumption undermines the asymptotic properties of the ML estimator and the resulting parameter estimates are no longer unbiased (Bollen, 1989). However, one of the key strengths of the SEM is the ability to explicitly model measurement error through the use of multiple indicator latent factors. Given that the level-1 random effects are defined as latent factors in the SEM, we can simply now regress these "level-1" latent factors on one or more "level-2" latent factors. I put these terms in quotes because these latent factors are analytically equivalent in the SEM; only we know that the factors have fundamentally different meanings given our specific parameterization of the factor loading matrices.

A generic path diagram of a multilevel model with two correlated multiple indicator latent factors at level-2 is presented in Figure 6. The figure presents a situation in which there is a random level-1 intercept and slope, and each of these is regressed on two level-2 predictors which are modeled as multiple indicator latent factors. Just as in the comparison of OLS regression and SEM, the latent factors allow for the estimation and removal of measurement error from the exogenous predictors, and thus provide unbiased estimates of the regression coefficients relating the level-2 predictors to the level-1 random effects.

Multiple Indicator Dependent Variables. Just as we are able to include multiple indicator latent factors for our level-2 covariates, we can use this strategy for our dependent measure as well. That is, instead of modeling a manifest variable loading on the level-1 random effect factors, given proper data characteristics we can model this as a latent factor itself. Although measurement models have been proposed in the standard MLM approach, these are currently limited in that factor loadings must be fixed to unity and all residual variances are equated for all items (e.g., Raudenbush, Rowan, & Kang, 1991).[5] There is no such limitation in the SEM, and a hypothetical SEM with two correlated latent factors at level-1 is presented in Figure 7. An excellent example of this type of model was given in Bauer (in press) who

[5] Note that exciting recent work has focused on the integration of multilevel modeling and item response theory via MCMC estimation to model measurement error in MLMs (e.g., Fox & Glas, 2001, 2003). However, given the recency of these developments, these methods have not yet been incorporated in applied research settings.

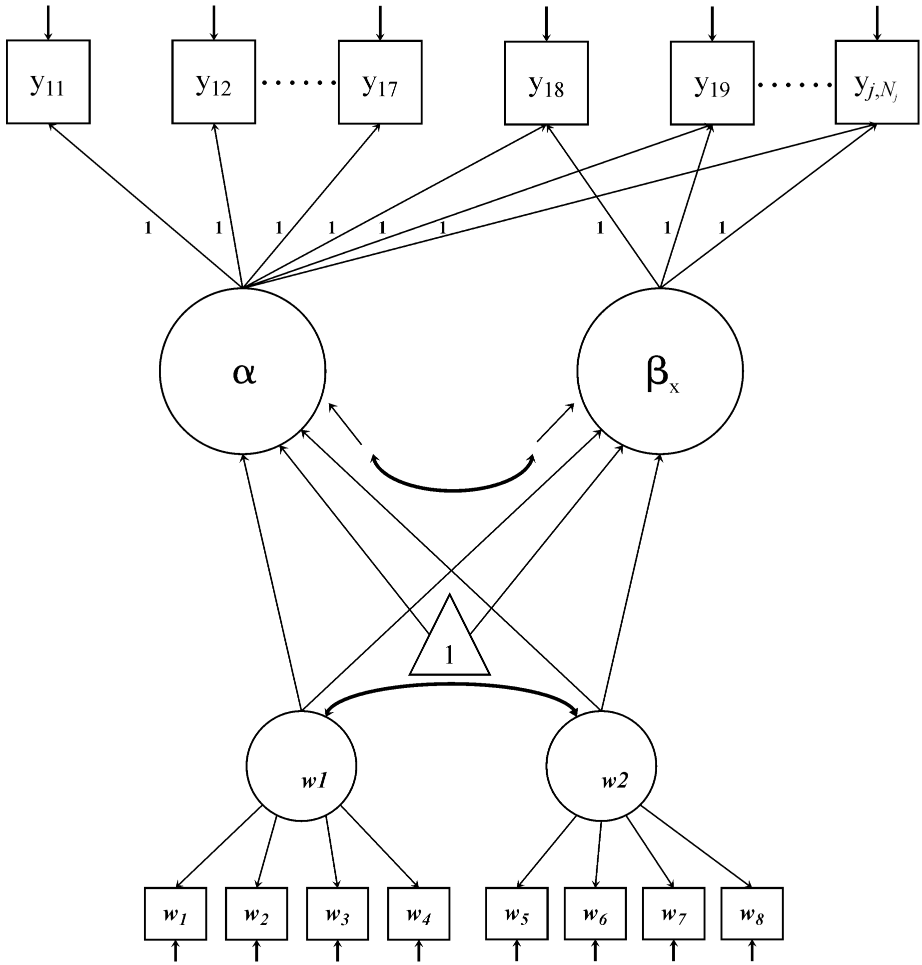

Figure 6
Hypothetical Model with Multiple Indicator Latent Factors Defined for the Level-2 Predictors

drew item level data from the HSB teacher survey. He used four individual items to assess teacher perceptions of control over school policy and five items assessing perceptions of control over the classroom. This represented a multilevel measurement model with two level-1 dependent variables modeled as multiple indicator latent factors, and these factors were then regressed on several level-2 predictors. Although current software makes estimating such a model quite tedious, it can be seen that there are many potentially interesting applications of models such as these.

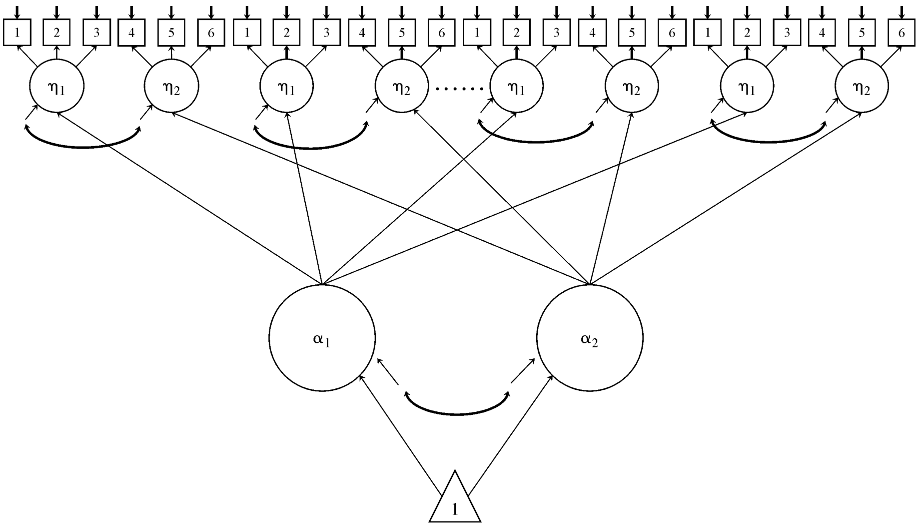

Figure 7
Hypothetical 3-level SEM with a Measurement Model for Two Correlated Constructs at Level-1

Three-Level Growth Models. One of the key advantages of the MLM growth model is that it can easily be expanded to higher levels of nesting (e.g., time nested within child nested within classroom). It has been quite challenging to estimate models such as these within the SEM framework. However, using the methods described above, the SEM can also be used to estimate a 3-level growth model, and a hypothetical model is presented in Figure 8. This is similar in form to the measurement model depicted in Figure 7, but here the two correlated latent factors at level-1 represent randomly varying intercepts and slopes. However, the second order latent factors (representing level-2) capture fixed and random effects associated with the children nested within, say, classrooms. Finally, these level-2 factors are then regressed on a level-3 covariate, here shown as a multiple-indicator latent factor. Thus the SEM can be applied to estimate a 3-level growth model with time nested within child nested within classroom.

Testing Within-Level and Across-Level Mediation. Baron and Kenny (1986) describe methods for testing mediation within the standard regression model. Krull and MacKinnon (1999, 2001) extended these methods to the multilevel model and, although promising, these methods are limited to testing a single mediator at a time. Given the estimation of MLMs within the SEM framework, we can augment these tests with those available within SEM.

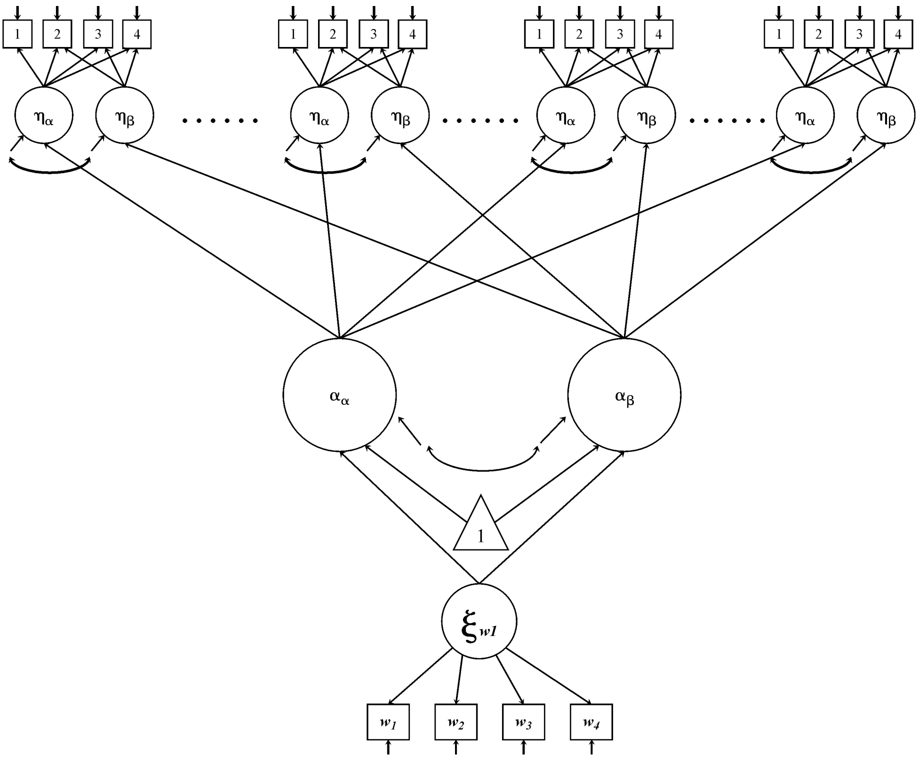

Figure 8
Hypothetical 3-level SEM with a Growth Model at Level-1, Fixed and Random Effects at Level-2, and a Multiple Indicator Latent Factor at Level-3

Specifically, one of the key strengths of the SEM framework is the ability to decompose total effects into direct, indirect and specific indirect effects, and to calculate standard errors for all of these (e.g., Bollen, 1987). In principal, this would allow for tests of mediation within level-1, within level-2, or simultaneously across both levels. Bauer (in press) extended his multivariate multilevel measurement model to formally test whether the size of the school mediated the effect of private versus public school in the prediction of the level-1 perceptions of control factors. More work is needed to better understand the advantages and disadvantages of these tests, but the SEM does offer interesting alternatives to testing mediation within the MLM.

Testing Within-Level and Across-Level Moderation. Because the MLM is an extension of the standard OLS regression, interactions are tested in precisely the same way. Namely, the unique contribution of the product between two main effects is assessed in the presence of the main effects, and a significant unique effect is indicative of a higher order interaction (e.g., Cohen, 1978). A complexity that arises in the MLM is that interactions may occur within just level-1, just level-2, or across levels 1 and 2 (the cross-level interaction). We have recently proposed methods for probing these interactions in the general MLM (Bauer & Curran, 2003), in the multilevel growth model (Curran, Bauer & Willoughby, in press-a) and in the SEM growth model (Curran, Bauer, & Willoughby, in press- b). Interestingly, the SEM approach to the multilevel model cannot only test all three types of interactions that arise in the MLM (with the cross-level interaction arising as indirect effects in the SEM; see Curran et al., in press-b), but the SEM can extend some of these tests in a powerful way. Specifically, one of the strengths of the general SEM is the ability to test a variety of interesting questions regarding similarities and differences in model parameters across discrete group membership. This is accomplished through the multiple group SEM in which models are defined within each of two or more discrete groups, and model parameters can be equated within or across groups to empirically evaluate a number of hypotheses. These hypotheses might relate to potential group differences in the structure of level-1 and level-2 random effects (e.g., thus allowing for heterogeneous covariance structures of random effects across group membership). Or these might consider differences in measurement structure or regression structure over groups. For example, in Curran and Muthén (1999) and Muthén and Curran (1997) we applied such models to evaluate the efficacy of treatment interventions within an experimental design. Regardless of specific question, the SEM provides a method for evaluating these questions of moderation in ways that are not currently easily accessible via the standard MLM.

Alternative Methods of Estimation. A distinct advantage of the multilevel model is that there are well developed methods for the incorporation of alternative link functions to allow for explicit modeling of dependent measures that are scaled as dichotomous, ordinal or count variables (e.g., Davidian & Giltinan, 1995; Vonesh & Chinchilli, 1997). However, the estimation of the standard multilevel model under ML currently assumes that any continuous dependent variable is normally distributed. In contrast, whereas the SEM does not currently allow for the use of link functions, there have been a variety of recent developments in alternative methods of estimation in the SEM that help overcome several limitations associated with the ML estimator. For example, Browne (1984) proposed the ADF estimator that is asymptotically distribution free and works well at large sample sizes and less complex models (Curran, West, & Finch, 1996; Muthén & Kaplan, 1985, 1992). Recent developments draw on elements of the ADF estimator which increases applicability in many research settings including diagonally weighted least squares (Jöreskog, 1994) and robust weighted least squares (Muthén, du Toit, & Spisic, 1997). Further, Satorra and Bentler (1990) proposed adjustments to the ML to correct standard errors and test statistics under nonnormality. Finally, Bollen (1996) developed a two-staged least squares estimator for use in SEMs that is asymptotically distribution free and less influenced by specification error. All of these alternative methods could be used in the estimation of SEMs with continuously but nonnormally distributed data that are not currently available in MLM.

Incorporation of Categorical Dependent Variables via WLS Estimation. As I noted above, the SEM cannot currently explicitly incorporate alternative link functions as is done in the nonlinear MLM. However, the SEM can incorporate categorical dependent measures through the analysis of polychoric correlations (e.g., Olsson, 1979) using weighted least squares (WLS) estimation (e.g., Muthén, 1983, 1984). This approach is based on the premise that the observed categories arise from the eclipsing of some threshold value associated with an unobserved underlying continuous variable (or latent response function). The structure among these unobserved continuous measures are estimated via polychoric correlations, and the SEM is fitted to this correlation structure using WLS estimation. Thus, a variety of potentially interesting multilevel SEMs could be fitted to categorical dependent variables allowing for the continued capitalization of other advantages offered by the SEM.

Multilevel Finite Mixture Modeling. There has been a recent flurry of work focused on the extension of the classic finite mixture model to several types of SEMs (Arminger & Stein, 1997; Jedidi, Jagpal, & DeSarbo, 1997; see Bauer & Curran, in press-a, for a review). Briefly, the general goal of mixture modeling is to identify two or more latent classes that represent subpopulations that are hypothesized to exist but could not be identified as such based strictly on the available observed measures. Instead, latent class membership must be probabilistically inferred as a function of the distributional structure of the data. Although clear evidence of latent classes may be attributable to specific characteristics of the data that do not relate to “true” latent class structure (e.g., Bauer & Curran, in press-a, in press-b, in press-c), the identification of multiple classes may also reflect true heterogeneity in the population. In principal, these mixture models could be applied using the methods described above to provide some empirically-based insight into the potential of population heterogeneity in the sample data. Although great care must be taken in the application and interpretation of these techniques in general, this is an intriguing area of future research for the multilevel model.

Omnibus Measures of Model Fit. One current challenge that is encountered with the MLM is the difficulty of calculating an omnibus measure of model fit. The reason is that there is no logical saturated model with which to compare a particular fitted model. Of course there are a variety of powerful methods for comparing nested models and using residual plots and information criterion measures to evaluate model fit (e.g., Raudenbush & Bryk, 2002), but there remains no single inferential test of the goodness-of-fit of a specific hypothesized MLM. In contrast, the SEM does allow for a natural saturated model to which any fitted model can be compared. This is reflected in the well known likelihood ratio test (LRT) which, under certain assumptions, follows an asymptotic central chi-square distribution to allow for formal inferential tests of omnibus model fit as well as the calculation of a variety of incremental fit indices. This LRT can be applied to evaluating multilevel models using the SEM framework. It is important to note, however, that if the SEM is fitted as I described above, the default test statistic provided by the software package is *not* correct. Instead, an alternative saturated model must be estimated that incorporates the unique data structure being analyzed (see Bauer, in press, for details). However, once defined, the LRT can be used in the usual way. Further work is needed to better understand the potential advantages and limitations of these measures in practice, but these provide a promising way of evaluating model fit in practice.

Limitations in Using SEM to Estimate Multilevel Models

As I noted at the opening of this article, my focus here has not been on the extension of the general SEM to nested data structures, and I leave this to the creative contributions of McDonald and Goldstein (1989), Bentler and Liang (2003), du Toit and du Toit (in press), and others. Instead, I have explored methods that allow for the estimation of the multilevel model using the existing SEM analytical framework. Despite the potential advantages, there are of course a number of distinct limitations. Of most importance, I believe that if no other elements of the SEM are incorporated in a multilevel model then the SEM approach has nothing unique to offer over the standard multilevel model.

I presented the random regressions earlier to empirically demonstrate the analytically derived predictions, but in every case the complex SEM simply replicated the standard MLM. My recommendation is that if a particular research hypothesis can be fully evaluated using a standard multilevel model, by all means use the MLM approach. However, if unique elements of the SEM are to be used that are not currently available in MLM, then the SEM approach should be considered, but with several caveats.

First, the SEM approach that I have described above is a data management nightmare. Many steps are necessary to properly structure the data and the SEM code quickly becomes extensive. These challenges could of course be overcome through the creation of new software that is dedicated to this task, but this is currently a tedious and error prone process. Second, the interpretation of the multilevel models estimated as SEMs is non-standard, and care is needed in the identification of the proper parameter estimates and the proper interpretations (e.g., latent factor means in the SEM represent regression coefficients in the MLM, and indirect effects in the SEM represent cross-level interactions in the MLM). Finally, it is important to realize that simply because a particular multilevel model can be parameterized using SEM, it does not mean that the model is even remotely meaningful. For example, it is trivial to re-specify the covariances among the latent factors representing the level-2 random effects in the SEM so that one factor is regressed upon another within level-1. Although numerically estimable, this may be nonsense from an analytic or substantive standpoint. It is thus critical that these models be parameterized in a way that the results are both valid and meaningful.

Although I chose to focus my entire article on a comparison of the SEM and MLM, an intriguing final question is whether there is a compelling reason to even embark on such an endeavor in the first place. Indeed, the boundaries between these two modeling strategies are becoming increasingly porous as is evidenced in that fully random regressions can be estimated in the SEM and latent variable measurement models can be estimated in the MLM. We seem to be approaching a point in which the terms SEM and MLM better distinguish historical roots and commercial software rather than the underlying statistical models. This is best evidenced in the recent developments of the generalized linear latent and multilevel model (or GLLAMM) by Rabe-Hesketh and colleagues (Rabe-Hesketh, Pickles, & Skrondal, 2001; Rabe-Hesketh, Skrondal, & Pickles, in press; Skrondal & Rabe-Hesketh, in press). GLLAMM is a general modeling framework in which there is no conceptualization of what is a "MLM" or what is a "SEM"; instead, the specific set of parameter matrices are selected that are needed to optimally test a given question of interest. The orientation of this approach is theoretically appealing and bodes well for the ongoing development of maximally general models that are focused less on historical lineage and more on the parameterization of the statistical model that optimally corresponds to the given theoretical model of interest.

References

Allison, P. D. (2001). *Missing data*. Thousand Oaks: Sage.

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling* (pp. 243-277). Mahwah, NJ: Erlbaum.

Arminger, G. & Stein, P. (1997). Finite mixtures of covariance structure models with regressors. *Sociological Methods and Research, 26*, 148-182.

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

Bauer, D. J. (in press). Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics*.

Bauer, D. J. & Curran, P. J. (2002, June). *Estimating multilevel linear models as structural equation models*. Paper presented at the annual meeting of the Psychometric Society, Chapel Hill, NC.

Bauer, D. J. & Curran, P. J. (2003). *Testing and probing interactions in fixed and multilevel regression models*. Manuscript submitted for publication.

Bauer, D. J., & Curran, P. J. (in press-a). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*.

Bauer, D. J. & Curran, P. J. (in press-b). Over-extracting latent trajectory classes: Much ado about nothing? *Psychological Methods*.

Bauer, D. J. & Curran, P. J. (in press-c). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods*.

Bentler, P. M. & Liang, J. (2003). Two-level mean and covariance structures: Maximum likelihood via an EM algorithm. In S. P. Reise & N. Duan (Eds.), *Multilevel modeling: Methodological advances, issues, and applications* (pp. 53-70). Mahwah, NJ: Erlbaum.

Bollen, K. A. (1987). Total, direct, and indirect effects in structural equation models. In C. C. Clogg (Ed.), *Sociological methodology* (pp. 37-69). Washington, DC: American Sociological Association.

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

Bollen, K. A. (1995). Structural equation models that are nonlinear in latent variables: A least squares estimator. In P. M. Marsden (Ed.) *Sociological methodology* (pp. 223-251). Washington, DC: American Sociological Association.

Bollen, K. A. (1996). An alternative 2SLS estimator for latent variable models. *Psychometrika, 61*, 109-21.

Browne, M. W. (1984). Asymptotic distribution free methods in analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37*, 62-83.

Bryk, A. S. & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, *101*, 147-158.

Bryk, A. S. & Raudenbush, S. W. (1992). *Hierarchical linear models. Applications and data analysis methods*. Thousand Oaks: Sage.

Burstein, L. (1980). The analysis of multilevel data in educational research and evaluation. *Review of Research in Education, 8*, 158-233.

Cohen, J. (1978). Partialled products are interactions; partialled powers are curve components. *Psychological Bulletin, 85*, 858-866.

Curran, P. J., Bauer, D. J, & Willoughby, M. T. (in press-a). Testing and probing interactions in hierarchical linear growth models. In C. S. Bergeman & S. M. Boker (Eds.), *The Notre Dame series on quantitative methodology, Volume 1: Methodological issues in aging research*. Mahwah, NJ: Lawrence Erlbaum Associates.

Curran, P. J., Bauer, D. J., & Willoughby, M. T. (in press-b). Testing and probing main effects and interactions in latent trajectory models. *Psychological Methods*.

Curran, P. J. & Muthén, B. O. (1999). The application of latent curve analysis to testing developmental theories in intervention research. *American Journal of Community Psychology, 27*, 567-595.

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

Davidian, M. & Giltinan, D.M. (1995). *Nonlinear models for repeated measurement data*. New York: Chapman & Hall.

du Toit, S. H. C. & du Toit, M. (in press). Multilevel structural equation modeling. In J. de Leeuw & I. G. G. Kreft (Eds.), *Handbook of quantitative multilevel analysis*. Boston: Kluwer Academic Publishers.

Fox, J.-P. & Glas, C. A. W. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika, 66*, 269-286.

Fox, J.-P. & Glas, C. A. W. (2003). Bayesian modeling of measurement error in predictor variables using item response theory. *Psychometrika, 68*, 169-191.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika, 73*, 43-56.

Goldstein, H. I. & McDonald, R. P. (1988). A general model for the analysis of multilevel data. *Psychometrika, 53*, 455-467.

Hox, J. J. (1998). Multilevel modeling: When and why. In I. Balderjahn, R. Mathar & M. Schader (Eds.), *Classification, data analysis, and data highways* (pp.147-154). New York: Springer Verlag.

Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997). Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science, 16*, 39-59.

Jöreskog, K. G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika, 59*, 381-389.

Krull, J. L. & MacKinnon, D. P. (1999). Multilevel mediation modeling in group-based intervention studies. *Evaluation Review, 23*, 418-444.

Krull, J. L. & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research, 36*, 249-277.

Laird, N. M. & Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics, 38*, 963-974.

Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika, 74*, 817-827.

MacCallum, R. C., Kim, C., Malarkey, W., & Kielcolt-Glaser, J. (1997). Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioral Research, 32*, 215-253.

Mason, W. M., Wong, G. Y., & Entwisle, B. (1983). Contextual analysis through the multilevel linear model. *Sociological Methodology*, 72-103.

McArdle, J. J. & Hamagami, F. (1996). Multilevel models from a multiple group structural equation perspective. In G. Marcoulides and R. Schumacker (Eds.), *Advanced structural equation modeling techniques* (pp. 89-124). Hillsdale, NJ: Lawrence Erlbaum Associates.

McDonald, R. P. (1993). A general model for two-level data with responses missing at random. *Psychometrika, 58*, 575-585.

McDonald, R. P. (1994). The bilevel reticular action model for path analysis with latent variables. *Sociological Methods and Research, 22*, 399-413.

McDonald, R. P. & Goldstein, H. (1989). Balanced versus unbalanced designs for linear structural relations in two-level data. *British Journal of Mathematical and Statistical Psychology, 42*, 215-232.

Mehta, P. D. & West, S. G. (2000). Putting the individual back into individual growth curves. *Psychological Methods, 5*, 23-43.

Meredith, W. & Tisak, J. (1984, July). *"Tuckerizing" growth curves*. Paper presented at the annual meeting of the Psychometric Society, Santa Barbara, CA.

Meredith, W. & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55*, 107-122.

Muthén, B. O. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics, 22*, 48-65.

Muthén, B. O. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49*, 115-132.

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika, 42*, 215-232.

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research, 22*, 376-398.

Muthén, B. O. & Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods, 2*, 371-402.

Muthén, B. O., du Toit, S. H. C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes*. Unpublished manuscript.

Muthén, B. O. & Kaplan, D. (1985). A comparison of methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology, 38*, 171-189.

Muthén, B. O. & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology, 45*, 19-30.

Muthén, L. & Muthén, B .O. (1998). *Mplus user's guide*. Los Angeles: Authors.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (1999). *Mx: Statistical modeling* (Computer Manual, 5th Edition). Richmond, VA: Virginia Commonwealth University, Department of Psychiatry.

Neale, M. C. & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*. Dordrecht, NL: Kluwer Academic Publishers.

Newsom, J. T. (2002). A multilevel structural equation model for dyadic data. *Structural Equation Modeling, 9*, 431-447.

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika, 44*, 443-460.

Rabe-Hesketh, S., Pickles, A., & Skrondal, A. (2001). GLLAMM: A class of models and a Stata program. *Multilevel Modelling Newsletter, 13*, 17-23.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (in press). Generalized multilevel structural equation modelling. *Psychometrika*.

Rao, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrika, 51*, 83-90.

Raudenbush, S. W. (2001). Toward a coherent framework for comparing trajectories of individual change. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 35-64). Washington, DC: American Psychological Association.

Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models. Applications and data analysis methods* (2nd ed.). Thousand Oaks: Sage.

Raudenbush, S. W., Rowan, B. & Kang, S. J. (1991). A multilevel, multivariate model for studying school climate with estimation via the EM algorithm and application to U.S. high-school data. *Journal of Educational Statistics, 16*, 295-330.

Rovine, M. J. & Molenaar, P. C. M. (1998). A nonstandard method for estimating a linear growth model in LISREL. *International Journal of Behavioral Development, 22*, 453-473.

Rovine, M. J. & Molenaar, P. C. M. (2000). A structural modeling approach to a multilevel random coefficients model. *Multivariate Behavioral Research, 35*, 51-88.

Rovine, M. J. & Molenaar, P. C. M. (2001). A structural equations modeling approach to the general linear mixed model. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 65-96). Washington, DC: American Psychological Association.

SAS (2000). *SAS/STAT user's guide: Version 8, Volume 1, 2, & 3*. Cary: Author.

Satorra, A. & Bentler, P. M. (1990). Model conditions for Asymptotic robustness in the analysis of linear relations. *Computational Statistics & Data Analysis, 10*, 235-249.

Schafer, J. L. & Graham, J. W. (2002) Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147-177.

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth curve models. *Journal of Educational and Behavioral Statistics, 24*, 323-355.

Skrondal, A. & Rabe-Hesketh, S. (in press). *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models*. Chapman & Hall/CRC.

Snijders, T. & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.

Tucker, L. R. (1958). Determination of parameters of a functional relation by factor analysis. *Psychometrika, 23*, 19-23.

Vonesh, E. F. & Chinchilli, V. M. (1997), *Linear and nonlinear models for the analysis of repeated measurements*. New York: Marcel Dekker.

Willett, J. B. & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin, 116*, 363-381.

Wothke, W. (2000). Longitudinal and multi-group modeling with missing data. In T. D. Little, K. U. Schnabel & J. Baumert (Eds.) *Modeling longitudinal and multiple group data: Practical issues, applied approaches, and specific examples* (pp. 219-240). Hillsdale, NJ: Erlbaum.

Yuan, K.-H. & Bentler, P. M. (2001). A unified approach to multigroup structural equation modeling with nonstandard samples. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: New developments and techniques* (pp. 35-56). Mahwah, NJ: Lawrence Erlbaum Associates.