❧ 10 ❧

LATENT VARIABLE MODELING OF LONGITUDINAL AND MULTILEVEL DATA

Bengt Muthén*

An overview is given of modeling of longitudinal and multilevel data using a latent variable framework. Particular emphasis is placed on growth modeling. A latent variable model is presented for three-level data, where the modeling of the longitudinal part of the data imposes both a covariance and a mean structure. Examples are discussed where repeated observations are made on students sampled within classrooms and schools.

1. INTRODUCTION

The concept of a latent variable is a convenient way to represent statistical variation not only in conventional psychometric terms with respect to constructs measured with error, but also in the context of models with random coefficients and variance components. These features will be studied in this paper. The random coefficient feature is shown to present a useful way to study change and growth over time. The variance component feature is shown to reflect correctly common cluster sampling procedures.

*University of California, Los Angeles

This paper gives an overview of some aspects of latent variable modeling in the context of growth and clustered data. A new multilevel latent variable model is presented which not only has a covariance structure but also a mean structure, where the mean structure arises naturally from the growth perspective. Emphasis is placed on the benefits that can be gained from multilevel as opposed to conventional modeling, which ignores the multilevel data structure. Data from large-scale educational surveys are used to illustrate the points.

The paper is organized as follows. Sections 2–6 discuss theory and Sections 7 and 8 applications. Section 2 discusses aggregated versus disaggregated modeling and Section 3 intraclass correlations and design effects in the context of a two-level latent variable model. In Section 4, a two-level latent variable model and its estimation for continuous-normal data will be presented as a basis for analyses. Section 5 shows how a three-level model can be applied to growth modeling and how it can be reformulated as a two-level model. Section 6 shows how this modeling can be fit into the two-level latent variable framework and how the estimation can be carried out by conventional structural equation modeling software. The remaining sections present applications. Section 7 uses two-wave data on mathematics achievement for students sampled within classrooms. Section 7.1 discusses measurement error when data have both within- and between-group variation and gives an example of estimating reliability for multiple indicators of a latent variable. Section 7.2 uses the same example to discuss change over time in within- and between-group variation taking unreliability into account. Section 8 takes the discussion of change over time further using a four-wave data set on students sampled within schools. Here, a growth model is formulated for the relationships between socio-economic status, attitude toward math, and mathematics achievement. Issues related to the assessment of stability and cross-lagged effects are also discussed.

2. AGGREGATED VERSUS DISAGGREGATED MODELING

Consider the following two-level, hierarchical data structure. Let $u_{gi} = (u_{gi1} \ u_{gi2} \ldots u_{gip})'$ denote a p-dimensional vector for randomly sampled groups ($g = 1, 2, \ldots, G$) and randomly sampled individuals within each such group ($i = 1, 2, \ldots, N_g$). We may write the corresponding (total) covariance matrix as a sum of a between- and a within-group part,

$$\Sigma_T = \Sigma_B + \Sigma_W. \tag{1}$$

In a typical educational example, Σ_W refers to student-level variation and Σ_B refers to class-level or school-level variation. In line with Muthén and Satorra (1995; see also Skinner, Holt, and Smith 1989) we will use the term "aggregated modeling" when the usual sample covariance matrix S_T is analyzed with respect to parameters of Σ_T and "disaggregated modeling" when the analysis refers to parameters of Σ_W and Σ_B. In our terms, a multi-level model is a disaggregated model for multilevel data. Such data can, however, also be analyzed by an aggregated model—i.e., a model for the total covariance matrix Σ_T.

In terms of conventional maximum-likelihood covariance structure analysis (e.g., see Bollen 1989) for estimating Σ_T parameters and drawing inferences, multilevel data present complications of correlated observations due to cluster sampling. Special procedures are needed to properly compute standard errors of estimates and chi-square tests of model fit. Effects of ignoring the multilevel structure and using conventional procedures for simple random sampling are illustrated in the next section in the context of a latent variable model. The model is that of a conventional analysis in that the usual set of latent variable parameters is involved.

In a disaggregated (or multilevel) model, the parameters themselves change from those of the conventional analysis. A much richer model with both within and between parameters is used to describe both individual- and group-level phenomena.

It is of interest to compare Σ_T analysis and Σ_W analysis with respect to the magnitude of estimates. This comparison has a strong practical flavor because if the differences are small, the multilevel aspects of the data can be ignored apart from perhaps small corrections of standard errors and chi square. This is frequently the case. Even in such cases, however, there may be information in the data that can be described in interesting ways by parameters of Σ_B. In other words, a frequent shortcoming when ignoring the multilevel structure of the data is not what is misestimated but what is not learned.

3. DESIGN EFFECTS

Drawing on Muthén and Satorra (1995), this section gives a brief overview of effects of the cluster sampling in multilevel data on the standard errors and test of model fit used in conventional covariance structure analysis assuming simple random sampling.

Consider the well-known design effect (deff) formula for the variance estimate of a mean with cluster size c and intraclass correlation ρ,

$$V_C/V_{SRS} = 1 + (c - 1)\rho, \qquad (2)$$

where V_C is the (true) variance of the estimator under cluster sampling and V_{SRS} is the corresponding (incorrect) variance assuming simple random sampling (Cochran 1977). The intraclass correlation is defined as the amount of between-group variation divided by the total amount of variation (between plus within). This formula points out that the common underestimation of standard errors when incorrectly assuming SRS is due to the combined effects of group size (c) and intraclass correlations (ρ's). Given that educational data often have large group sizes in the range of 20–60, even a rather small intraclass correlation value of 0.10 can have huge effects. However, it is not clear how much guidance, if any, this formula gives in terms of multivariate analysis and the fitting of latent variable models (see also Skinner, Holt, and Smith 1989). Muthén and Satorra (1995) carried out a Monte Carlo study to shed some light on the magnitude of these effects.

In our experience with survey data, common values for the intraclass correlations range from 0.00 to 0.50 where the higher range values have been observed for educational achievement test scores and the lower range for attitudinal measurements and health-related measures. Both the way the groups are formed and the content of the variables have major effects on the intraclass correlations. Groups formed as geographical segments in alcohol use surveys indicated intraclass correlations in the range of 0.02 to 0.07 for amount of drinking, alcohol dependence, and alcohol abuse. Equally low values have been observed in educational surveys when it comes to attitudinal variables related to career interests of students sampled within schools. In contrast, mathematics achievement scores for U.S. eighth graders show proportions of variance due to class components of around 0.30–0.40 and due to school components of around 0.15–0.20.

Muthén and Satorra (1995) generated data according to a ten-variable multilevel latent variable model with a two-factor simple structure. This is a disaggregated model of the kind described above. In this case, the loading matrices are equal across the two levels, which means that the same covariance structure model holds on all three levels: within, between, and total. Conventional analysis of the total matrix can then be studied in a case where the model is correct, but standard errors and test of model fit are not. Data were generated as 200 randomly generated groups and group sizes

(total sample size) 7 (1400), 15 (3000), 30 (6000), and 60 (12000). These are common values in educational achievement surveys. One thousand replications were used.

Table 1 gives chi-square test statistics for a conventional analysis incorrectly assuming simple random sampling. The model has 34 degrees of freedom. Using the terms above, this is an analysis of an aggregated model using the usual sample covariance matrix S_T. The within and between parameters are not separately estimated; only the parameters of the total matrix are. It is seen that an inflation in chi-square values is obtained by increasing both group size and intraclass correlations, implying that models would be unnecessarily rejected. Only for small values of the intraclass correlations and the group size might the distortion be ignorable— for example, for the combinations (0.05, 7), (0.05, 15), and (0.10, 7). Judging from this table it seems that even for a rather small intraclass

TABLE 1
Chi-Square Testing with Cluster Data

Intraclass Correlation	Group Size			
	7	15	30	60
0.05				
Chi-square				
Mean	35	36	38	41
Var	68	72	80	96
5%[a]	5.6	7.6	10.6	20.4
1%	1.4	1.6	2.8	7.7
0.10				
Chi-square				
Mean	36	40	46	58
Var	75	89	117	189
5%	8.5	16.0	37.6	73.6
1%	1.0	5.2	17.6	52.1
0.20				
Chi-square				
Mean	42	52	73	114
Var	100	152	302	734
5%	23.5	57.7	93.1	99.9
1%	8.6	35.0	83.1	99.4

[a]Percentage of replications where model was rejected at 5-percent level.

correlation of 0.10, the distortions may be large if the group size exceeds 15. The standard errors of the estimates show an analogous pattern in terms of deflated values. Muthén and Satorra (1995) go on to show how standard errors and chi-square tests of fit can be corrected by taking the clustering into account. They also show that the ML estimator of the disaggregated, multilevel model performs well, but the estimator does have problems of convergence at small intraclass correlation values and small group sizes and is also sensitive to deviations from normality. In the normal case with intraclass correlations of 0.10 and groups sizes ranging from 7 to 60, the multilevel ML estimator also performs well when the number of groups is reduced from 200 to 50. In our experience, when the number of groups is much less than 50, this estimator does not give trustworthy results.

We conclude from these simulations that ignoring the multilevel nature of the data and carrying out a conventional covariance structure analysis may very well lead to serious distortions of conventional chi-square tests of model fit and standard errors of estimates.

4. A TWO-LEVEL (DISAGGREGATED) MODEL

This section briefly reviews the theory for two-level modeling and estimation. Specific latent variable models are not discussed here. The specific latent variable model used in growth modeling is given in the next section, where it is shown how it fits into the framework given in the present section.

In line with McDonald and Goldstein (1989) and Muthén (1989, 1990), suppose that there are G groups, of which group g has N_g members ($g = 1, 2, \ldots, G$). Let $\mathbf{z}_g$ be a vector of length r containing the values of group-level variables for group g and let $\mathbf{u}_{gi}$ be a vector of length p containing the values of individual-level variables for the ith individual in group g. Arrange the data vector for which independent observations are obtained as

$$\mathbf{d}'_g = (\mathbf{z}'_g, \mathbf{u}'_{g1}, \mathbf{u}'_{g2}, \ldots, \mathbf{u}'_{gN_g}), \tag{3}$$

where we note that the length of $\mathbf{d}_g$ varies across groups. The mean vector and covariance matrix are

$$\boldsymbol{\mu}'_{d_g} = [\boldsymbol{\mu}'_z, \mathbf{1}'_{N_g} \otimes \boldsymbol{\mu}'_u] \tag{4}$$

$$\Sigma_{d_g} = \begin{pmatrix} \Sigma_{zz} & \text{symmetric} \\ \mathbf{1}_{N_g} \otimes \Sigma_{uz} & \mathbf{I}_{N_g} \otimes \Sigma_W + \mathbf{1}_{N_g} \mathbf{1}'_{N_g} \otimes \Sigma_B \end{pmatrix}. \tag{5}$$

Here, the symbol $\otimes$ denotes a Kronecker product defined as follows. For an $m \times n$ matrix A and an $s \times t$ matrix B, $A \otimes B$ is the $ms \times nt$ matrix

$$A \otimes B = \begin{pmatrix} a_{11}B & \ldots & a_{1n}B \\ \vdots & \ldots & \vdots \\ a_{m1}B & \ldots & a_{mn}B \end{pmatrix}. \tag{6}$$

Σ_W and Σ_B are $p \times p$ within-group and between-group covariance matrices for the u variables. Muthén (1994a: 378–82) discusses the above covariance structure and contrasts it with that of conventional covariance structure analysis.

The elements of μ_z, μ_u, Σ_{zz}, Σ_{uz}, Σ_W, and Σ_B are functions of the parameters of the model. Assuming multivariate normality of $\mathbf{d}_g$, the ML estimator minimizes the function

$$F = \sum_{g=1}^{G} \{\log|\Sigma_{d_g}| + (\mathbf{d}_g - \mu_{d_g})' \Sigma_{d_g}^{-1} (\mathbf{d}_g - \mu_{d_g})\} \tag{7}$$

with respect to the parameters of the model. Here, the sizes of the arrays involving u variables are determined by the product $N_g \times p$, which is large if there are many individuals per group. A remarkable fact is that the likelihood can be expressed in a form that reduces the sizes of the arrays involving u variables to depend only on p (cf. McDonald and Goldstein 1989; Muthén 1989, 1990),

$$F = \sum_{d=1}^{D} G_d\{\ln|\Sigma_{B_d}| + tr[\Sigma_{B_d}^{-1}(S_{B_d} + N_d(t_d - \mu)(t_d - \mu)')]\}$$
$$+ (N - G)\{\ln|\Sigma_W| + tr[\Sigma_W^{-1} S_{PW}]\}, \tag{8}$$

where

$$\Sigma_{B_d} = \begin{pmatrix} N_d\Sigma_{zz} & \text{symmetric} \\ N_d\Sigma_{yz} & \Sigma_W + N_d\Sigma_B \end{pmatrix},$$

$$S_{B_d} = N_d G_d^{-1} \sum_{k=1}^{G_d} \begin{pmatrix} z_{dk} - \bar{z}_d \\ \bar{u}_{dk} - \bar{u}_d \end{pmatrix} [(z_{dk} - \bar{z}_d)'(\bar{u}_{dk} - \bar{u}_d)'],$$

$$t_d - \mu = \begin{pmatrix} \bar{z}_d - \mu_z \\ \bar{u}_d - \mu_u \end{pmatrix},$$

$$S_{PW} = (N - G)^{-1} \sum_{g=1}^{G} \sum_{i=1}^{N_g} (u_{gi} - \bar{u}_g)(u_{gi} - \bar{u}_g)'.$$

Here, D denotes the number of groups of a distinct size, d is an index denoting a distinct group size category with group size N_d, G_d denotes the number of groups of that size, S_{B_d} denotes a between-group sample covariance matrix, and S_{PW} is the usual pooled-within sample covariance matrix.

Muthén (1989,1990) pointed out that the minimization of the ML fitting function defined by equation (8) can be carried out by conventional structural equation modeling software, apart from a slight modification due to the possibility of singular sample covariance matrices for groups with small G_d values. A multiple-group analysis is carried out for $D + 1$ groups, the first D groups having sample size G_d and the last group having sample size $N - G$. Equality constraints are imposed across the groups for the elements of the parameter arrays μ, Σ_{zz}, Σ_{uz}, Σ_B, and Σ_W (see Muthén 1990 for details).

Muthén (1989, 1990) also suggested an ad hoc estimator that considered only two groups,

$$F' = G\{\ln|\Sigma_{B_c}| + tr[\Sigma_{B_c}^{-1}(S_B + c(t - \mu)(t - \mu)')]\}$$

$$+ (N - G)\{\ln|\Sigma_W| + tr[\Sigma_W^{-1} S_{PW}]\}, \tag{9}$$

where the definition of the terms simplifies relative to Equation (8) due to ignoring the variation in group size, dropping the d subscript, and using $D = 1$, $G_d = G$, and $N_d = c$, where c is the average group size (see Muthén 1990 for details). When data are balanced—i.e., when the group size is constant for all groups—the ML estimator will be obtained. Experience with the ad hoc estimator for covariance structure models with unbalanced data indicates that the estimates, and also the standard errors and chi-square test of model fit, are quite close to those obtained by the true ML estimator. This observation has also been made for growth models where a mean structure is added to the covariance structure, see Muthén (1994b).

In Section 6 we will return to the specifics of how the mean and covariance structures of equations (8) and (9) can be represented in conventional structural equation modeling software for the case of growth modeling. The growth model will be presented next.

5. A THREE-LEVEL HIERARCHICAL MODEL

Random coefficient growth modeling (e.g., see Laird and Ware 1982), or multilevel modeling (e.g., see Bock 1989), describes individual differ-

ences in growth. In this way, it goes beyond conventional structural equation modeling of longitudinal data and its focus on autoregressive models (e.g., see Jöreskog and Sörbom 1977; Wheaton et al. 1977). Random-coefficient modeling for three-level data has been described (e.g., see Goldstein [1987]; Bock [1989]; Bryk and Raudenbush [1992]) as follows.

Consider the three-level data

Group   : $g = 1, 2, \ldots, G$
(school, class)
Individual   : $i = 1, 2, \ldots, n$
Time   : $t = 1, 2, \ldots, T$
y_{git} : individual-level, outcome variable
x_{it} : individual-level, time-related variable (age, grade)
w_{git} : individual-level, time-varying covariate
w_{gi} : individual-level, time-invariant covariate
z_g : group-level variable

and the growth equation,

$$y_{git} = \alpha_{gi} + \beta_{gi} x_{it} + \gamma_{git} w_{git} + \zeta_{git}. \tag{10}$$

An important special case that will be the focus of this paper is where the time-related variable $x_{it} = x_t$. This means that for the tth occasion, all individuals have the same x_t value. An example of this is educational achievement studies where t corresponds to grade. The x_t values are, for example, $0, 1, 2, \ldots, T - 1$ for linear growth. We will also restrict attention to the case of $\gamma_{git} = \gamma_t$. Both restrictions are necessary in order to fit the model into currently available software and estimation techniques for the latent variable framework to be discussed in the next section.

The three levels of the growth model are then

$$y_{git} = \alpha_{gi} + x_t \beta_{gi} + \gamma_t w_{git} + \zeta_{git}, \tag{11}$$

$$\begin{cases} \alpha_{gi} = \alpha_g + \pi_\alpha w_{gi} + \delta_{\alpha_{gi}} \\ \beta_{gi} = \beta_g + \pi_\beta w_{gi} + \delta_{\beta_{gi}}, \end{cases} \tag{12}$$

$$\begin{cases} \alpha_g = \alpha + \kappa_\alpha z_g + \delta_{\alpha_g} \\ \beta_g = \beta + \kappa_\beta z_g + \delta_{\beta_g}. \end{cases} \tag{13}$$

In the case of growth modeling using a simple random sample of individuals, it is possible to translate the growth model from a two-level

model to a one-level model by considering a $T \times 1$ vector of outcome variables y for each individual. Analogously, we may reduce the three-level model to two levels. For example, in a model without covariates and group-level variables,

$$\mathbf{y}_{gi} = \begin{pmatrix} y_{gi1} \\ \vdots \\ y_{giT} \end{pmatrix} = [1\mathbf{x}] \begin{pmatrix} \alpha_{gi} \\ \beta_{gi} \end{pmatrix} + \zeta_{gi}, \tag{14}$$

which may be expressed as the sum of a between- and a within-group component,

$$\mathbf{y}_{gi} = y_g^* + y_{gi}^*, \tag{15}$$

where

$$y_g^* = [1\mathbf{x}] \begin{pmatrix} \alpha_g \\ \beta_g \end{pmatrix} + \zeta_g^* \tag{16}$$

and

$$y_{gi}^* = [1\mathbf{x}] \begin{pmatrix} \delta_{\alpha gi} \\ \delta_{\beta gi} \end{pmatrix} + \zeta_{gi}^*, \tag{17}$$

where ζ_{gi} is the sum of the two uncorrelated components ζ_g^* and ζ_{gi}^*. Equations (16) and (17) will be further discussed from a latent variable perspective in the next section.

6. LATENT VARIABLE FORMULATION

For the case of simple random sampling of individuals, Meredith and Tisak (1984, 1990) have shown that the random coefficient model of the previous section can be formulated as a latent variable model (for applications in psychology, see McArdle and Epstein [1987]; for applications in education, see Muthén [1993] and Willett and Sayer [1993]; for applications in mental health, see Muthén [1983, 1991]). The latent variable formulation can be directly extended to the three-level data case. The basic idea can be simply described as follows. In the example of equation (16), α_g and β_g are latent variables varying across individuals and the coefficient matrix $[1\mathbf{x}]$ corresponds to a factor analysis loading matrix relating the response variables to the latent variables. In equation (17) the corre-

sponding latent variables are $\delta_{\alpha_{gi}}$ and $\delta_{\beta_{gi}}$, whereas the loading matrix is the same. We find that

$$E(\mathbf{y}_{gi}) = E(y_g^*) + E(y_{gi}^*) = [1\mathbf{x}]\begin{pmatrix} \alpha \\ \beta \end{pmatrix} + 0, \tag{18}$$

$$V(\mathbf{y}_{gi}) = V(y_g^*) + V(y_{gi}^*) = \Sigma_B + \Sigma_W, \tag{19}$$

$$\Sigma_B = [1\mathbf{x}]V\begin{pmatrix} \delta_{\alpha g} \\ \delta_{\beta g} \end{pmatrix}[1\mathbf{x}]' + V(\zeta_g^*), \tag{20}$$

$$\Sigma_W = [1\mathbf{x}]V\begin{pmatrix} \delta_{\alpha gi} \\ \delta_{\beta gi} \end{pmatrix}[1\mathbf{x}]' + V(\zeta_{gi}^*), \tag{21}$$

so that the parameters of this latent variable model are α, β, and the elements of the covariance matrices in equations (20) and (21). This example fits into the framework of Section 4 by noting that here

$$\mathbf{u} = \mathbf{y}, \tag{22}$$

where $\mathbf{y}$ is the $T \times 1$ vector of equation (14).

Muthén (1989, 1990) showed how the multilevel fitting functions F and F' of equations (8) and (9) could be implemented in existing structural equation modeling software using multiple-group analysis. Equation (8) shows that there are D such groups that involve between-level parameters and one group that involves within-level parameters. We will focus on how to implement the D mean and covariance structures of equation (8) for the example we are considering, where the mean structure appears in the expression

$$N_d(t_d - \mu)(t_d - \mu)' = N_d(\bar{u}_d - \mu_u)(\bar{u}_d - \mu_u)', \tag{23}$$

and the covariance structure in the expression

$$\Sigma_{B_d} = \Sigma_W + N_d\Sigma_B. \tag{24}$$

Figure 1 shows a path diagram that is useful in describing this mean and covariance structure. The figure again corresponds to the case of no covariates w, or group-level variables z. Here, $T = 3$. Figure 1 shows the implementation of the model structure in the example of equations (18)–(21). The top part of the figure introduces the mean structure and the between-level covariance structure of the model by using the latent variables y_g^* of equation (16) premultiplied by the square root of N_d to match
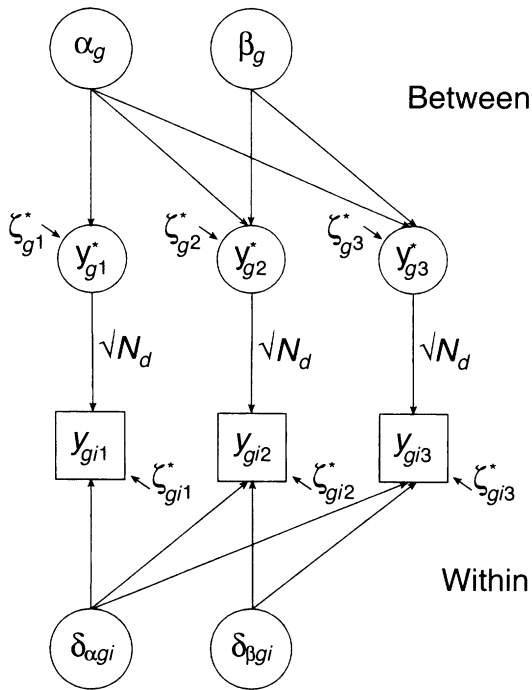
FIGURE 1. Latent variable growth model formulation for two-level, three-wave data.

equations (23) and (24). The bottom part of the figure introduces the within-level covariance structure. On the within side, we note that in our example, the $\delta_{\alpha_{gi}}$ factor influences the y's with coefficients 1 at all time points. The influence of the $\delta_{\beta_{gi}}$ factor on the y variables is captured by the constants of x_t. This makes it clear that nonlinear growth can be accommodated by estimating the x_t coefficients—e.g., holding the first two values fixed at 0 and 1, respectively, for identification purposes. The between-level α_g and β_g factors influence the between-level y_g^* variables with the same coefficients as on the within side; the factor loading matrices are the same and only the factor and residual covariance matrices differ. A strength of the latent variable approach is that this loading matrix equality assumption can easily be relaxed. For example, it may not be necessary to include between-group variation in the growth rate so that the between-level variation is represented by only one factor.

A special feature of the growth model is the mean structure imposed on μ in the ML fitting function of equation (8), where μ represents the means of group- and individual-level variables. In the specific growth model shown in Figure 1, the mean structure arises from the three observed y variable means being expressed as functions of the means of the α_g and β_g factors. Equation (8) indicates that the means need to be included on the between side of Figure 1, while the means on the within side are fixed at zero. This implies that dummy zero means are entered for the within group. The number of degrees of freedom for the chi-square test of model fit obtained in conventional software then needs to be reduced by the number of individual-level variables.

The latent between-level variables may also be related to observed between-level variables z, as in Section 4 and Section 5. Furthermore, it is straightforward to add individual-level covariates such as the w variables in equations (10)–(13), defining

$$\mathbf{u}' = (\mathbf{y}' \quad \mathbf{w}'), \tag{25}$$

where $\mathbf{w}$ is a vector of all time-invariant and time-varying covariates. While z only contributes between-group variability, $\mathbf{u}$ contributes both between-group and within-group variability, as seen in equation (6).

The model in Figure 1 can also be generalized to applications with multiple indicators of latent variable constructs instead of single outcome measurements y at each time point. The covariates may also be latent variables with multiple indicators. Furthermore, estimates may be obtained for the individual growth curves by estimating the individual values of the intercept and slope factors α and β. This relates to empirical Bayes estimation in the conventional growth literature (e.g., see Bock 1989).

The determination of model identifiability can draw on regular latent variable modeling rules by observing in equation (8) that it is sufficient that the individual-level parameters can be identified from the within-group covariance matrix and that the group-level parameters (and the means) can be identified from the between-group covariance matrix (and the means).

Further details and references on latent variable modeling with two-level data are given in Muthén (1994a), where suggestions for analysis strategies are also given. Software for calculating the necessary sample statistics, including intraclass correlations, is available from Statlib at *http: //lib.stat.cmu.edu/general/latent.2level* or by sending the E-mail message "send latent.2level from general" to *statlib@stat.cmu.edu*.

7. ANALYSIS OF TWO-WAVE ACHIEVEMENT DATA

We will first consider data from the Second International Mathematics Study (SIMS) drawing on analyses presented in Muthén (1991b, 1992). Here, a national probability sample of school districts was selected proportional to size; a probability sample of schools was selected proportional to size within school district, and two classes were randomly drawn within each school. The data consist of 3724 students observed in 197 classes from 113 schools; the class sizes varied from 2 to 38, with a typical value of around 20. Eight variables are considered corresponding to various areas of eighth-grade mathematics. The same set of items was administered as a pretest in the fall of eighth grade and again as a posttest in the spring.

Muthén (1991b: 341) poses the following questions:

> The substantive questions of interest in this article are the variance decomposition of the subscores with respect to within-class student variation and between-class variation and the change of this decomposition from pretest to posttest. In the SIMS . . . such variance decomposition relates to the effects of tracking and differential curricula in eighth-grade math. On the one hand, one may hypothesize that effects of selection and instruction tend to increase between-class variation relative to within-class variation, assuming that the classes are homogeneous, have different performance levels to begin with, and show faster growth for higher initial performance level. On the other hand, one may hypothesize that eighth-grade exposure to new topics will increase individual differences among students within each class so that posttest within-class variation will be sizable relative to posttest between-class variation.

7.1. *Measurement Error and Reliability of Multiple Indicators*

Analyses addressing the above questions can be done for overall math performance, but it is also of interest to study if the differences vary from more basic to more advanced math topics. For example, one may ask if the differences are more marked for more advanced topics. When focusing on

specific subsets of math topics, the resulting variables consist of a sum of rather few items and therefore contain large amounts of measurement error. At grade eight, the math knowledge is not extensively differentiated and a unidimensional latent variable model may be formulated to estimate the reliabilities for a set of such variables. Muthén (1991b) formulated a multilevel factor analysis model for the two-wave data. Given that the amount of across-school variation was small relative to the across-classroom variation, the school distinction was ignored and the data analyzed as a two-level structure. At each time point, unidimensionality was specified for both within- and between-class variation, letting factors and measurement errors correlate across time on each level. Table 2 presents estimates from both the multilevel factor analysis (MFA) model (see the within and between columns) and a conventional analysis (see the total columns). Reliability is estimated from the factor model as the proportion of variance in the indicator accounted for by the factor. As is seen from Table 2, the estimated student-level (within) reliabilities are considerably lower than reliabilities obtained from a total analysis.

In psychometrics it is well-known that reliabilities are lower in more homogeneous groups (Lord and Novick 1968). Here, however, it seems important to make the distinction shown in Figure 2.

Figure 2 (a) corresponds directly to the Lord and Novick case. The three line segments may be seen as representing three different classrooms

TABLE 2
The Second International Mathematics Study:
Reliabilities of Math Achievement at Two Time Points

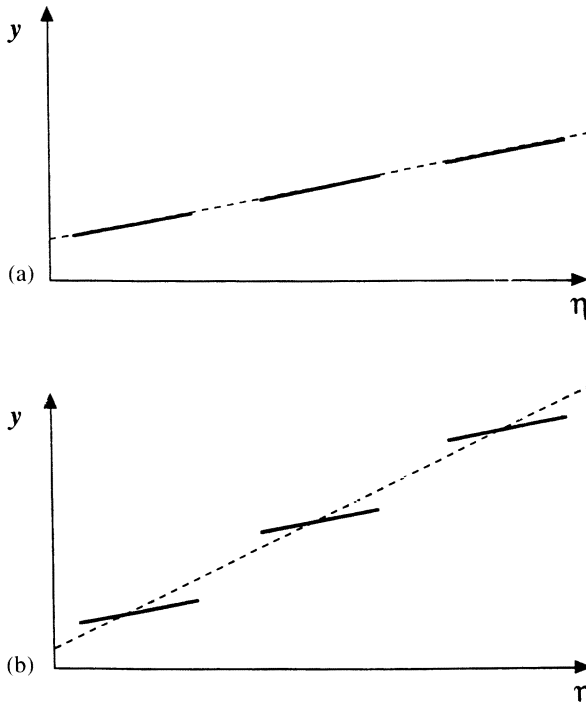| Variables	Number of Items	Pretest			Posttest		
		Total	MFA		Total	MFA	
			Within	Between		Within	Between
RPP	8	.61	.44	.96	.68	.52	.97
FRACT	8	.60	.38	.97	.68	.49	.98
EQ EXP	6	.36	.18	.83	.55	.32	.92
INTNUM	2	.34	.18	.81	.43	.25	.88
STESTI	5	.44	.25	.86	.52	.34	.89
AREAVOL	2	.29	.18	.82	.38	.23	.84
COORVIS	3	.34	.18	.92	.42	.26	.80
PFIGURE	5	.32	.17	.78	.46	.31	.77

FIGURE 2. Regression of an indicator on its latent variable.

with different student factor values η and student test score values y. The regression line for all classrooms is given as a broken line. All classrooms have the same intercept and slope. For any given classroom, the range of the factor is restricted and due to this restriction in range the reliability is attenuated relative to that of all classrooms.

Figure 2 (b) probably corresponds more closely to the situation at hand. Here, the three classrooms have the same slopes but different intercepts. The regression for the total analysis is marked as a broken line. It gives a steeper slope and a higher reliability than for any of the classrooms. One can argue, however, that the higher reliability is incorrectly obtained by analyzing a set of heterogeneous subpopulations as if they were one single population (cf. Muthén 1989). In contrast, the multilevel model captures the varying intercepts feature and reveals the lower within-classroom reliability that holds for each classroom.

The Table 2 between-classroom reliabilities are considerably higher than the within-classroom values. These between coefficients concern reliable variation across classrooms and therefore have another interpretation than the student-level reliabilities. The results indicate that what distinguishes classrooms with respect to math performance is largely explained by a single dimension—i.e., a total score—and that on the whole the topics measure this dimension rather similarly.

7.2. Attenuation of Intraclass Correlations by Measurement Error

We will consider the size of the intraclass correlations as indicators of school heterogeneity. This can be seen as a function of social stratification giving across-school differences in student "intake," as well as differences in the teaching and what schools do with a varied student intake. The U.S. math curriculum in grades 7–10 is very varied with large differences in emphasis on more basic topics such as arithmetic and more advanced topics such as geometry and algebra. Ability groupings ("tracking") is often used. In some other countries, however, a more egalitarian teaching approach is taken, the curriculum is more homogeneous, and the social stratification less strong. In international studies the relative sizes of variance components for student, class, and school are used to describe such differences (e.g., see Schmidt, Wolfe, and Kifer 1993).

Table 3 gives conventional variance component results from nested, random-effects ANOVA in the form of the proportion of variance between classrooms relative to the total variance. This is the same as the intraclass correlation measure. It is seen that the intraclass correlations increase from pretest to posttest. The problem with these values is, however, that they are likely to be attenuated by the influence of measurement error. This is because student-level measurement error adds to the within-part of the total variance—i.e., the denominator of the intraclass correlation. The distortion is made worse by the fact that the student-level measurement error is likely to decrease from pretest to posttest due to more familiarity with the topics tested.

The MFA columns of Table 3 give the multilevel factor analysis assessment of intraclass correlations using the one-factor model in the previous subsection. Here, the intraclass correlations are computed using the between and within variances for the factor variable, not including measurement error variance. It is seen that these intraclass correlations are considerably higher and indicate a slight decrease over time. This is a

TABLE 3
The Second International Mathematics Study:
Intraclass Correlations (proportion between classroom
variance) of Math Achievement at Two Time Points

Variables	Number of Items	ANOVA		MFA	
		Pretest	Posttest	Pretest	Posttest
RPP	8	.34	.38	.54	.52
FRACT	8	.38	.41	.60	.58
EQ EXP	6	.27	.39	.65	.64
INTNUM	2	.29	.31	.63	.61
STESTI	5	.33	.34	.58	.56
AREAVOL	2	.17	.24	.54	.52
COORVIS	3	.21	.32	.57	.55
PFIGURE	5	.23	.33	.60	.54

change in the opposite direction from the ANOVA results. Results from
ANOVA would therefore give misleading evidence for answering the ques-
tions posed in Muthén (1991b).

8. ANALYSIS OF FOUR-WAVE DATA BY GROWTH MODELING

The Longitudinal Study of American Youth (LSAY) is a national study of
performance in and attitudes toward science and mathematics. It is con-
ducted as a longitudinal survey of two cohorts spanning grades 7 to 12.
LSAY uses a national probability sample of about 50 public schools, test-
ing an average of about 50 students per school every fall starting in 1987.
Data from four time points, grades 7–10, and one cohort will be used to
illustrate the methodology for analysis of individual differences in growth.
         In this analysis, mathematics achievement and attitudes toward math
will be related to each other and to the socioeconomic status of the family.
The data to be analyzed consists of a total sample of 1869 students in 50
schools with complete data on all variables in the analysis. Mathematics
achievement is quantified as a latent variable (theta) score obtained by IRT
techniques using multiple test forms and a large number of items including
arithmetic, geometry, and algebra. The intraclass correlations for the math
achievement variable for the four grades are estimated as 0.18, 0.13, 0.15,

0.14, indicating a noteworthy degree of across-school variation in achievement. Attitude toward math was measured by a summed score using items having to do with how hard the student finds math, whether math makes the student anxious, whether the student finds math important, etc. As expected, the intraclass correlations for the attitude variable are considerably lower than for achievement. They are estimated as 0.05, 0.06, 0.04, and 0.02. The Pearson product-moment correlations between achievement and attitude are estimated as 0.4–0.6 for each of the four time points. The measure of socioeconomic status pertains to parents' educational levels, occupational status, and the report of some resources in the home. It has an intraclass correlation of 0.17.

For simplicity in the analyses to be presented, the two-group ad hoc estimator discussed in Section 4 will be used, not the full-information maximum-likelihood estimator. This means that the standard errors and chi-square tests of model fit are not exact but are approximations; given our experience, they are presumably quite reasonable ones. Nevertheless, statements about significance and model fit should not be interpreted in exact terms.

8.1. Two-Level Modeling

In this section, two-level modeling of the LSAY data will be outlined, both in terms of the growth model and, as a contrast, in terms of a conventional autoregressive model. The primary analysis considers a growth model that extends the single-variable, two-level growth model of Figure 1 to a simultaneous model of the growth process for both achievement and attitude. SES will be used as a student-level, time-invariant covariate, explaining part of the variation in these two growth processes. No observed variables on the school level will be used in this case, but school-level variables can be easily incorporated in the general model. The model is described graphically in Figure 3.

Let the top row of observed variables (squares) represent achievement at each of the four time points and the bottom row the corresponding attitudes. The SES covariate is the observed variable to the left in the figure.

Consider first the student- (within-)level part of Figure 3. The latent variable (circle) to the right of the observed variable of SES is hypothesized to influence four latent variables, the intercept (initial status) factor and slope (growth rate) factor for achievement (the top two latent vari-

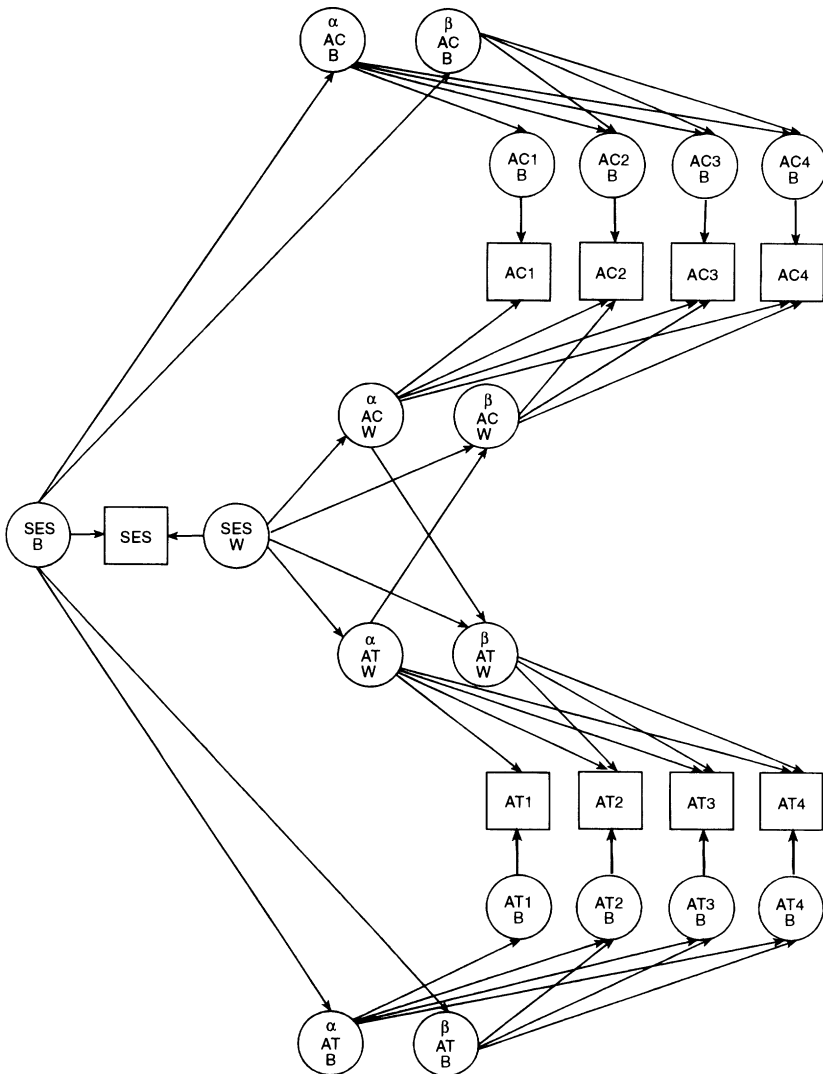**FIGURE 3.** Two-level, four-wave growth model for achievement and attitude related to socioeconomic status.

ables) and the intercept and slope factors for attitude (the bottom two latent variables). The intercept for each growth process is hypothesized to have a positive influence on the slope of the other growth process. In order not to clutter the picture, residuals and their correlations are not drawn in the

figure, but a residual correlation is included for the intercepts as well as the slopes. For each growth process, the model is as discussed in connection with Figure 1. Preliminary analyses suggest that nonlinear growth for achievement should be allowed for by estimating the growth steps from grade 8 to 9 and from grade 9 to 10, while for attitude a linear process is sufficient. In fact, for attitude, a slight decline is observed over time. The reason for this is not clear, but does perhaps reflect that among a sizable part of the student population there is an initial positive attitude about math that wears off over the grades either because math gets harder or because they stop taking math. For each growth process, correlations are allowed for among residuals at adjacent time points. Residual correlations are also allowed for across processes at each time point. Cross-lagged effects between the outcome variables are allowed for but not shown in the figure. It should be noted that even without cross-lagged effects the model postulates that achievement and attitude do influence each other via their growth intercepts and slopes. For example, if the initial status factor for attitude has a positive influence on the growth rate factor for achievement, initial attitude has a positive influence on later achievement scores.

The hierarchical nature of the data is taken into account by inclusion of the between- (school-)level part of the model. The between-level part of Figure 3 is similar to the within-level part. Starting with the SES variable to the left in the figure, it is seen that the variation in this variable is decomposed into two latent variables, one for the within variation and one for the between variation (the between factor is to the left of the SES square). At the top and the bottom of the figure are given the between-level intercept and slope factors for achievement and attitude, respectively. As in Figure 1, the influence of these factors on achievement/attitude is specified to have the same structure and parameter values as for the within part of the model. A minor difference here is that the intercept for one process is not specified to influence the slope of the other process, but all four intercept and slope factor residuals are instead allowed to be freely correlated. Also, on the between side, correlations among adjacent residuals over time are not included in the model.

As a comparison to the above growth model, a more conventional autoregressive, cross-lagged model will also be analyzed. This is shown in Figure 4 in its two-level form. On the within level, the figure shows a lag-one autoregressive process for both achievement and attitude with lag-one cross-lagged effects, where SES is allowed to influence the outcomes at each time point. The between-level part of the model is here not given a
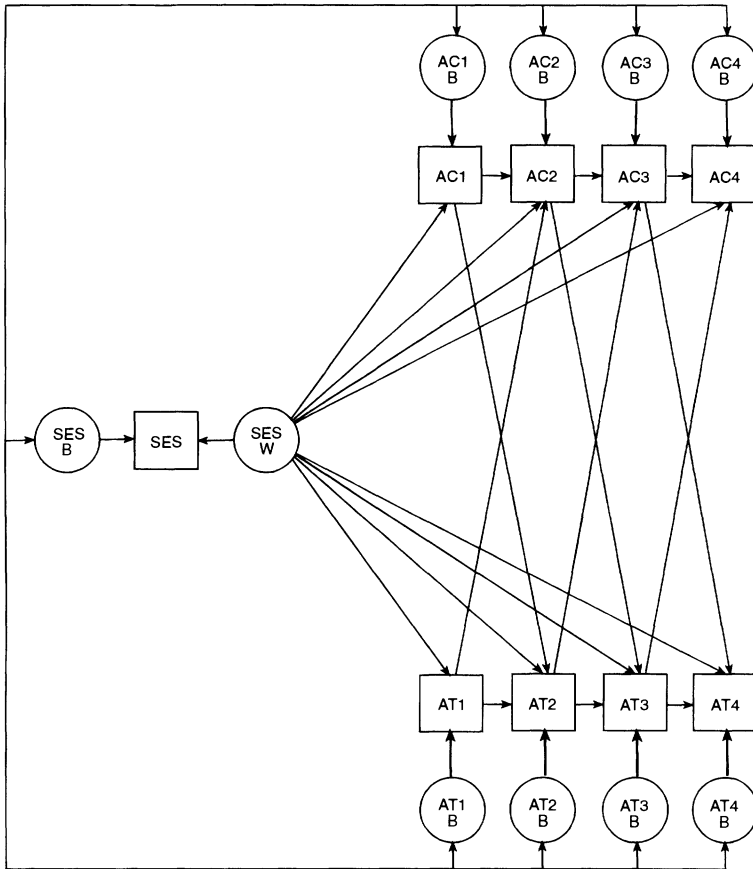
FIGURE 4. Two-level, four-wave autoregressive model for achievement and attitude related to socioeconomic status.

specific structure but the between-level covariance matrix is made unrestricted by allowing all between-level factors to freely correlate.

8.2. Analyses

The results of the model testing are summarized in Table 4. This table also includes BIC values for evaluating model fit, comparing the model in question against the unrestricted model (see Raftery 1993, 1995). It is of interest to first ignore the hierarchical nature of the data and to give the incorrect tests of fit for the single-level analogs of the autoregressive and

TABLE 4
Model Testing for LSAY Data ($n = 1,869$)

Model	Chi-square	Degrees of Freedom	BIC
Autoregressive Model			
Ignoring the two-level structure (chi-square values incorrect)			
Model 1: Lag-one	534.7	12	418
Model 2: Lag-three	22.3	6	−36
Two-level analysis			
Model 3: Lag-one	518.8	12	402
Model 4: Lag-three	28.1	6	−30
Growth Model			
Ignoring the two-level structure (chi-square value incorrect)			
Model 5: Growth model	44.0	8	−34
Two-level analysis			
Model 6: Growth model	68.4	30	−152

growth models. To this aim, the conventional maximum-likelihood fitting function is used. The lag-one autoregressive model did not fit well at all. To improve fit it was necessary to include a lag-three model for the auto-regressive part. The correct two-level tests of fit using the lag-one model of Figure 4 resulted in a clear rejection of the model, while a two-level, lag-three model gave a reasonable fit. The BIC values agree with this con-clusion. The degrees of freedom are the same for the single-level and two-level models because the two-level model doubles the number of parameters as well as the number of sample variances and covariances that are ana-lyzed (a mean structure is not involved in this model). The two-level, lag-three model shows positive and significant student-level cross-lagged effects of achievement and attitude on each other. The lag-three autoregres-sive structure of the model, however, makes it a rather complex and inele-gant representation of the data.

Turning to the growth model, Table 4 gives test results for the single-level model, which ignores the hierarchical nature of the data, and for the two-level model. As indicated by its BIC value, the two-level growth model is clearly preferred over the unrestricted model as well as the other models in Table 4. The estimates of this model are shown in Table 5. What is

TABLE 5
Results from the Two-Level Random Coefficient Growth Model

$n = 1869$
$\chi^2(30) = 68.38$

Within

	Parameter Estimates	t-Values
Cross-lags		
Achievement → Attitude		
Grade 7 → Grade 8	-0.001	-0.08
Grade 8 → Grade 9	-0.01	-0.79
Grade 9 → Grade 10	-0.01	-0.45
Attitude → Achievement		
Grade 7 → Grade 8	0.04	0.54
Grade 8 → Grade 9	-0.15	-1.74
Grade 9 → Grade 10	-0.15	-0.85
Growth Model		
Achievement initial status		
→ Attitude growth rate	0.003	0.39
Attitude initial status		
→ Achievement growth rate	0.23	1.29
Effects of SES on		
Achievement		
Initial status	2.93	10.21
Growth rate	0.16	1.84
Attitude		
Initial status	0.29	3.54
Growth rate	-0.08	-2.31

Between

	Parameter Estimates	t-Values
Effects of SES on		
Achievement		
Initial status	7.96	5.24
Growth rate	0.91	3.31
Attitude		
Initial status	0.31	0.91
Growth rate	0.12	0.93

Factor Residual (Co)Variances

Achievement

Initial status	57.84	14.50
Growth rate	1.16	2.17
Initial status, growth rate	1.57	1.16

Attitude

Initial status	4.24	1.33
Growth rate	0.71	0.67
Initial status, growth rate	-0.80	-0.50

Achievement, attitude

Initial status	4.38	6.71
Growth rate	0.28	1.06

Initial Status Intercept

Achievement	52.47	117.38
Attitude	11.36	117.85

Growth Curve

Achievement

Grade 7	0*
Grade 8	1*
Grade 9	2.60	12.81
Grade 10	3.85	11.86

Attitude

Grade 7	0*
Grade 8	1*
Grade 9	2*
Grade 10	3*

Growth Rate Intercept

Achievement	2.37	9.82
Attitude	-0.32	-9.10

Factor Residual (Co)Variances

Achievement

Initial status	6.11	3.38
Growth rate	0.08	1.26
Initial status, growth rate	0.15	0.64

Attitude

Initial status	0.19	1.18
Growth rate	0.02	0.66
Initial status, growth rate	-0.03	-0.48

Achievement, Attitude

Initial status	0.65	2.09
Growth rate	-0.02	-1.20
Initial status, growth rate	-0.06	-0.58
Growth rate, initial status	-0.08	-1.50

*Parameter is fixed in this model.

particularly interesting about the two-level growth model is that in con-
trast to the autoregressive model, none of the student-level cross-lagged
effects are significantly different from zero. This makes for a very parsi-
monious model where the achievement and attitude processes are instead
correlated via the correlations among their intercept and slope factors. The
correlation between the intercept factors (not shown in the table) is posi-
tive (0.27) while the slope factor correlation is ignorable (0.08). The in-
fluences from the intercepts to the slopes turn out to be not significant.

    The student-level influence from SES is significantly positive for
both the achievement and attitude intercepts. It is insignificant for the
achievement slope and significantly negative for the attitude slope. It is not
clear what the negative effect represents, but this effect would be seen if
students from high SES homes have a strong initial positive attitude that
later becomes less positive. SES explains 12 percent of the student varia-
tion in the achievement intercept while it explains only 1 percent of the
student variation in the attitude intercept. In terms of the achievement
growth, the estimates indicate that relative to the positive growth from
grade 7 to 8, the growth is accelerated in later grades. For attitude, linear
growth is maintained.

    In the school-level part of the model, the correlation between achieve-
ment and attitude intercepts (not shown in the table) obtains a rather high
value, 0.61 (the student-level value is 0.27). On the school level it is seen
that SES does not have a significant influence on the attitude intercept or
slope factors. The influence on the achievement intercept and slope is, how-
ever, significantly positive. This reflects across-school heterogeneity in
neighborhood resources so that schools with higher SES families have both
higher initial achievement and stronger growth over grades. It is interesting
to note that significant *student*-level influence of SES on the student-level
achievement growth rate was not seen, while strongly significant *school*-
level influence of SES is seen on the school-level achievement growth rate.
This is an example of a difference in between-school and within-school
model structure. The latent variable approach readily accommodates such
model features.

9. CONCLUSIONS

This paper has presented a general model for latent variable growth anal-
ysis that takes into account cluster sampling. The model is of interest from
two methodological perspectives. First, it represents a multilevel latent vari-
able model that has not only a covariance structure but also a mean struc-

ture. Second, it represents a new latent growth model that has two levels. The latent variable formulation combines these two perspectives and makes for a very general modeling framework. The model has the advantage that it can be analyzed with existing structural equation modeling software.

It is clear from the real data analyses that much can be gained in terms of understanding the data if the multilevel analysis is used not only to compute correct standard errors and chi-square tests of model fit with nonindependent observations but also to interpret the parameters that cap- ture the nonindependence of the observations. The examples showed that different interpretations were obtained for parameters on the between- group level than on the within-group level.

The complexity of the models calls for sound modeling strategies, but these are difficult to formulate in general. Muthén (1994a) discusses latent variable modeling steps that are relevant for two-level data. The examples in the present paper illustrate the wide applicability of the new methodology. It is clear, however, that many other special types of multi- level growth models can be formulated and it will be interesting to have researchers from many fields explore these new possibilities. Further meth- odological research is also needed, such as regarding restricted maximum- likelihood estimation (REML) and robustness of inference in situations with small sample sizes and data that deviate from normality.

REFERENCES

Bock, R. Darrell. 1989. *Multilevel Analysis of Educational Data*. San Diego: Academic Press.

Bollen, Ken A. 1989. *Structural Equations with Latent Variables*. New York: Wiley.

Bryk, Anthony S., and Stephen W. Raudenbush. 1992. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage.

Cochran, William G. 1977. *Sampling Techniques*, 3rd ed. New York: Wiley.

Goldstein, Harvey I. 1987. *Multilevel Models in Educational and Social Research*. London: Oxford University Press.

Jöreskog, Karl G., and Dag Sörbom. 1977. "Statistical Models and Methods for Anal- ysis of Longitudinal Data." In *Latent Variables in Socio-Economic Models*, edited by D. J. Aigner and A. S. Goldberger, 285–325. Amsterdam: North-Holland.

Laird, Nan M., and James H. Ware. 1982. "Random-Effects Models for Longitudinal Data." *Biometrics* 38:963–74.

Lord, Frederic M., and Melvin R. Novick. 1968. *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.

McArdle, Jack J., and David Epstein. 1987. "Latent Growth Curves Within Develop- mental Structural Equation Models." *Child Development* 58:110–33.

McDonald, Rod, and Harvey Goldstein. 1989. "Balanced Versus Unbalanced Designs

for Linear Structural Relations in Two-Level Data." *British Journal of Mathematical and Statistical Psychology* 42:215–32.

Meredith, William, and John Tisak. 1984. "Tuckerizing Curves." Paper presented at the Psychometric Society annual meetings, Santa Barbara, CA.

———. 1989. "Latent Variable Modeling in Heterogeneous Populations." *Psychometrika* 54:557–85.

Muthén, Bengt. 1983. "Latent Variable Structural Equation Modeling with Categorical Data." *Journal of Econometrics* 22:43–65.

———. 1990. "Latent Curve Analysis." *Psychometrika* 55:107–122.

———. 1990. "Mean and Covariance Structure Analysis of Hierarchical Data." Paper presented at the Psychometric Society meeting in Princeton, New Jersey. UCLA Statistics Series #62, August 1990.

———. 1991a. "Analysis of Longitudinal Data Using Latent Variable Models with Varying Parameters." In *Best Methods for the Analysis of Change. Recent Advances, Unanswered Questions, Future Directions*, edited by L. Collins and J. Horn, 1–17. Washington: American Psychological Association.

———. 1991b. "Multilevel Factor Analysis of Class and Student Achievement Components." *Journal of Educational Measurement* 28:338–54.

———. 1993. "Latent Variable Modeling of Growth with Missing Data and Multilevel Data." In *Multivariate Analysis: Future Directions 2*, edited by C. R. Rao and C. M. Cuadras, 199–210. Amsterdam: North-Holland.

———. 1994a. "Multilevel Covariance Structure Analysis." *Sociological Methods and Research* 22:376–98.

———. 1994b. "Latent Variable Growth Modeling with Multilevel Data." Paper presented at the UCLA conference Latent Variable Modeling with Applications to Causality. March, 1994. To appear in *Latent Variable Modeling with Applications to Causality*, edited by Maia Berkane. Cambridge, MA: Springer-Verlag.

Muthén, Bengt, and Albert Satorra. 1995. "Complex Sample Data in Structural Equation Modeling." In *Sociological Methodology* 1995, edited by Peter V. Marsden, 267–316. Cambridge, MA: Blackwell.

Raftery, Adrian E. 1993. "Bayesian Model Selection in Structural Equation Models." In *Testing Structural Equation Models*, edited by Kenneth Bollen and J. Scott Long, 163–80. Newbury Park, CA: Sage.

———. 1995. "Bayesian Model Selection in Social Research." In *Sociological Methodology* 1995, edited by Peter V. Marsden, 111–63. Cambridge, MA: Blackwell.

Schmidt, W., R. G. Wolfe, and E. Kifer. (in press). "The Identification and Description of Student Growth in Mathematics Achievement." In L. Burstein (Ed.), *The IEA Study of Mathematics III: Student Growth and Classroom Processes*. Edited by L. Burstein. London: Pergamon.

Skinner, C. J., D. Holt, and T. M. F. Smith. 1989. *Analysis of Complex Surveys*. Chichester, England: Wiley.

Wheaton, Blair, Bengt Muthén, Duane Alwin, and Gene Summers. 1977. "Assessing Reliability and Stability in Panel Models." In *Sociological Methodology 1977*, edited by D. R. Heise, 84–136. San Francisco: Jossey-Bass.

Willett, John B., and Alice G. Sayer. 1993. "Using Covariance Structure Analysis to Detect Correlates and Predictors of Individual Change over Time." *Psychological Bulletin* 116:368–81.