DOCUMENT RESUME

ED 379 322 TM 022 693

AUTHOR Muthen, Bengt

TITLE Latent Variable Modeling of Longitudinal and Multilevel Data. Project 2.4, Quantitative Models To Monitor the Status and Progress of Learning and Performance and Their Antecedents.

INSTITUTION National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.

SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.

PUB DATE Oct 94

CONTRACT R117G10027

NOTE 31p.

PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.

DESCRIPTORS Change; Error of Measurement; Learning; *Longitudinal Studies; Mathematics Tests; Models; *Psychometrics; Sampling

IDENTIFIERS *Latent Variables; *Multilevel Analysis; Variance (Statistical)

ABSTRACT

The modeling of longitudinal and multilevel data using a latent variable framework is reviewed. Particular emphasis is placed on growth modeling. Examples are discussed where repeated observations are made on students sampled within classrooms and schools. 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 are explored. The random coefficient feature is shown to be a useful way to study change and growth over time, while the variance component feature is shown to correctly reflect common cluster sampling procedures. Four tables and four figures are included. (Contains 19 references.) (Author/SLD)

National Center for Research on
Evaluation, Standards, and Student Testing

Final Deliverable – October 1994

Project 2.4 Quantitative Models to Monitor the
Status and Progress of Learning and Performance
and Their Antecedents

Latent Variable Modeling of Longitudinal and Multilevel Data

Bengt Muthén, Project Director
University of California, Los Angeles
Graduate School of Education & Information Studies

U.S. Department of Education
Office of Educational Research and Improvement
Grant No. R117G10027 CFDA Catalog No. 84.117G

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90024-1522
(310) 206-1532

The work reported herein was supported in part by a grant from the Educational Research and Development Center Program cooperative agreement R117G10027 and CFDA Catalog number 84.117G as administered by the Office of Educational Research and Improvement, U.S. Department of Education; in part by grant AA 08651-01 from NIAAA; and in part by grant 40859 from the National Institute for Mental Health.

The findings and opinions expressed in this report do not reflect the policies of the Office of Education Research and Improvement or the U.S. Department of Education or of the other aforementioned funding agencies.

LATENT VARIABLE MODELING OF LONGITUDINAL AND MULTILEVEL DATA [1,2]

Bengt Muthén

CRESST/University of California, Los Angeles
Graduate School of Education & Information Studies

Abstract

An overview is given of modeling of longitudinal and multilevel data using a latent variable framework. Particular emphasis is placed on growth modeling. 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 correctly reflect common cluster sampling procedures.

This paper gives an overview of some aspects of latent variable modeling in the context of growth and clustered data. Emphasis will be 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 will be used to illustrate the points.

[1] Invited paper for the annual meeting of the American Sociological Association, Section on Methodology, Showcase Session.
[2] I thank Ginger Nelson Goff for expert assistance.

The paper is organized as follows. Sections 2–6 will discuss theory and Sections 7 and 8 applications. To save space, the theory sections will by necessity be terse. Some results are given for easy reference and the reader is referred to previous papers for the modeling rationale and the derivations of estimators (see, e.g., Muthén, 1990, 1992, 1994a, 1994b). In Section 2, aggregated versus disaggregated modeling will be discussed. Section 3 discusses 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. In Section 5, it is shown how a three-level model can be applied to growth modeling and how it can be re-formulated as a two-level model. Section 6 shows how this modeling can be fit into the two-level latent variable framework. It is shown that 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 towards 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 data structure. Let y_{gi} denote a p-dimensional vector for randomly sampled groups and randomly sampled individuals within each such group and decompose the y_{gi} into between- and within-group variation,

$$y_{gi} = y_{B_g} + y_{W_{gi}} \tag{1}$$

and consider the decomposition of the corresponding (total) covariance matrix into a within- and a between-group part,

$$\Sigma_T = \Sigma_B + \Sigma_W \tag{2}$$

In a typical educational example, Σ_W refers to student-level variation and Σ_B refers to class-level or school-level variation. It is assumed that parameters of the covariance matrices capture the essential aspect of the data. In line with Muthén and Satorra (1993) (see also Skinner, Holt, & 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 multilevel model is a disaggregated model for multilevel data. Such data can, however, also be analyzed by an aggregated model, that is, a model for the total covariance matrix Σ_T. In terms of estimating Σ_T parameters and drawing inferences, multilevel data present the usual 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, however, that of a conventional analysis in that the usual set of latent variable parameters are involved. In a disaggregated (or multilevel) model the aspiration level is higher in that 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.

A theme in our discussion is the comparison of 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, the most 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 (1993), 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,

$$V_C / V_{SRS} = 1 + (c - 1)\,\rho \qquad (3)$$

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 (icc) 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 icc's (ρ's). Given that educational data often have large groups sizes in the range of 20--60, even a rather small icc 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, & Smith, 1989). Muthén and Satorra (1993) 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 icc's 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 icc's. Groups formed as geographical segments in alcohol use surveys indicated icc's 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-Satorra 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, $\Sigma_B = \Sigma_W$, 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 (1,400), 15 (3,000), 30 (6,000), and 60 (12,000). 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 sampe covariance matrix S_T. The within and between parameters are not separately estimated, but only the parameters of the total matrix. It is seen that an inflation in chi-square values is obtained both by increasing group size and increasing icc's, implying that models would be unnecessarily rejected. Only for small values of the icc's and the group size might the distortion be ignorable, such as for the combinations (0.005, 7), (0.05, 15), and (0.10, 7). Judging from this table it seems that even for a rather small

Table 1

Chi-Square Testing With Cluster Data

	Group size			
Intraclass correlation	7	15	30	60
0.05				
Chi-square				
Mean	35	36	38	41
Var	68	72	80	96
5%	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

icc 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-Satorra 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 icc values and small group sizes and is also sensitive to deviations from normality. In the normal case with icc's 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, reducing the number of groups much below 50 does not give trustworthy results by this estimator.

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 gives a brief review of 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), assume $g = 1, 2, \ldots, G$ independently observed groups with $i = 1, 2, \ldots, N_g$ individual observations within group g. Let z and y represent group- and individual-level variables, respectively. Arrange the data vector for which independent observations are obtained as

$$\mathbf{d}_g' = (\mathbf{z}_g', \mathbf{y}_{g1}', \mathbf{y}_{g2}', \ldots, \mathbf{y}_{gN_g}') \tag{4}$$

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

$$\mu_{\mathbf{d}_g}' = [\mu_{\mathbf{z}}', \mathbf{1}_{N_g}' \otimes \mu_{\mathbf{y}}'] \tag{5}$$

$$\Sigma_{\mathbf{d}_g} = \begin{pmatrix} \Sigma_{\mathbf{zz}} & \text{symmetric} \\ \mathbf{1}_{N_g} \otimes \Sigma_{\mathbf{yz}} & \mathbf{I}_{N_g} \otimes \Sigma_W + \mathbf{1}_{N_g} \mathbf{1}'_{N_g} \otimes \Sigma_B \end{pmatrix} \tag{6}$$

Muthén (1994a, pp. 378-382) discusses the above covariance structure and contrasts it with that of conventional covariance structure analysis.

Assuming multivariate normality of $\mathbf{d}_g$, the ML estimator minimizes the function

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

Here, the parameter arrays are potentially of large size if there are many individuals per group. A remarkable simplification which makes the sizes not depend on group size is given as (cf. McDonald & Goldstein, 1989; Muthén, 1989, 1990)

$$F = \Sigma_d^D G_d \{ \ln |\Sigma_{B_d}| + tr\, [\Sigma_{B_d}^{-1} (S_{B_d} + N_d (\bar{v}_d - \mu)(\bar{v}_d - \mu)')] \}$$

$$+ (N - G) \{ \ln |\Sigma_W| + tr\, [\Sigma_W^{-1} S_{PW}] \} \tag{8}$$

where

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

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

$$\bar{v}_d - \mu = \begin{pmatrix} \bar{z}_d - \mu_z \\ \bar{y}_d - \mu_y \end{pmatrix}$$

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

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_{yz}, Σ_B, Σ_W. To obtain the correct chi-square test of model fit, a separate H_1 analysis needs to be done (see Muthén, 1990 for details).

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

$$F' = G\{\ln|\Sigma_{B_c}| + tr\,[\Sigma_{B_c}^{-1}(S_B + c\,(\bar{v} - \mu)\,(\bar{v} - \mu)')]\} \\ + (N - G)\{\ln|\Sigma_W| + tr\,[\Sigma_W^{-1} S_{PW}]\} \quad (9)$$

where the definition of the terms simplifies relative to equation 14 due to ignoring the variation in group size, dropping the d subcript, 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, that is, the group size is constant for all groups, this gives the ML estimator. 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 structure 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 (see, e.g., Laird & Ware, 1982), or multilevel modeling (see, e.g., Bock, 1989), describes individual differences in growth. In this way, it goes beyond conventional structural equation modeling of longitudinal data and its focus on auto-regressive models (see, e.g. Jöreskog & Sorbom, 1977; Wheaton, Muthén, Alwin, & Summers, 1977). Random-coefficient modeling for three-level data is described, for example, in Goldstein (1987), Bock (1989), and Bryk and Raudenbush (1992).

Consider the three-level data

Group (School, class)	:	$g = 1,2,\ldots,G$
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)
v_{git}	:	individual-level, time-varying covariate
w_{gi}	:	individual-level, time-invariant covariate
z_g	:	group-level variable

and the growth equation,

$$y_{git} = \alpha_{gi} + \beta_{gi}\, x_{it} + \gamma_{git}\, v_{git} + \zeta_{git}$$

An important special case that will be the focus of this paper is where the time-related variable $x_{it} = x_t$. An example of this is educational achievement studies where x_t corresponds to grade. The x_t values are for example 0, 1, 2, ..., T–1 for linear growth. We will also restrict attention to the case of $\gamma_{git} = \gamma_{gt}$. The three levels of the growth model are then

$$y_{git} = \alpha_{gi} + x_t\, \beta_{gi} + \gamma_{gt}\, v_{git} + \zeta_{git} \tag{11}$$

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

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

where the variation in the individual-level, time-invariant covariate w_{gi} is decomposed into between- and within-group parts

$$w_{gi} = w_{B_g} + w_{W_{gi}} \tag{14}$$

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 x 1 vector of outcome variables y for each individual. Analogously, we may reduce the three-level model to two levels as follows.

$$\mathbf{y}_{gi} = \begin{pmatrix} y_{gi1} \\ \vdots \\ y_{giT} \end{pmatrix} = [1\ \mathrm{x}] \begin{pmatrix} \alpha_{gi} \\ \beta_{gi} \end{pmatrix} + \zeta_{gi} \tag{15}$$

which may be expressed in five terms

$$\mathbf{y}_{gi} = [1\ \mathrm{x}] \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + [1\ \mathrm{x}] \begin{pmatrix} \delta_{\alpha g} \\ \delta_{\beta g} \end{pmatrix} + \zeta_g^* + [1\ \mathrm{x}] \begin{pmatrix} \delta_{\alpha gi} \\ \delta_{\beta gi} \end{pmatrix} + \zeta_{gi}^* \tag{16}$$

The first term represents the mean as a function of the mean of the initial status and the mean of the growth rate. The second and third terms correspond to between-group (school) variation. The fourth and fifth terms correspond to within-group variation.

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 & Epstein, 1987; for applications in education, see

Muthén, 1993 and Willett & Sayer, 1993; for applications in mental health, see Muthén, 1983, 1991). The basic idea can be simply described as follows. In equation 1, α_i is a latent variable varying across individuals. Assuming the special case of $x_{it} = x_t$, the x variable becomes a constant which multiplies a second latent variable β_i.

The latent variable formulation can be directly extended to the three-level data case. In line with Muthén (1989, 1990), Figure 1 shows a path diagram which is useful in implementing the multilevel estimation using the multilevel fitting function F or F'. The figure corresponds to the case of no covariates v, w, or z. It shows how the covariance structure

$$\Sigma_W + N_d \Sigma_B \tag{17}$$

can be represented by latent variables, introducing a latent between-level variable for each outcome variable y. On the within side, we note that the α factor influences the y's with coefficients 1 at all time points. The constants of x_t are the coefficients for the influence of the β factor on the y variables. This makes it clear that non-linear growth can be accommodated by estimating the x_t coefficients, for example, holding the first two values fixed at 0 and 1, respectively, for identification purposes. The within-level α and β factors correspond to the $\delta_{\alpha_{ig}}$ and $\delta_{\beta_{ig}}$ residuals of equation 12. The between-level α and β factors correspond to the δ_{α_g} and δ_{β_g} residuals of equation 13. From equation 16 it is clear that the influence from these two factors is the same on the between side as it is on the within side. Corresponding to this, in Figure 1 the Σ_B structure is identical to the Σ_W structure. A strength of the latent variable approach is that this equality assumption can easily be relaxed. For example, it may not be necessary include between-group variation in the growth rate. These latent between-level variables may also be related to observed between-level variables z_g as in Section 4.

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 five observed variable means being expressed as functions of the means of the α and β factors, here applied on the between side, see equation 16. Equation 8 indicates that the means need to be included on the between side of Figure 1 given that the mean term of F is scaled by N_d,

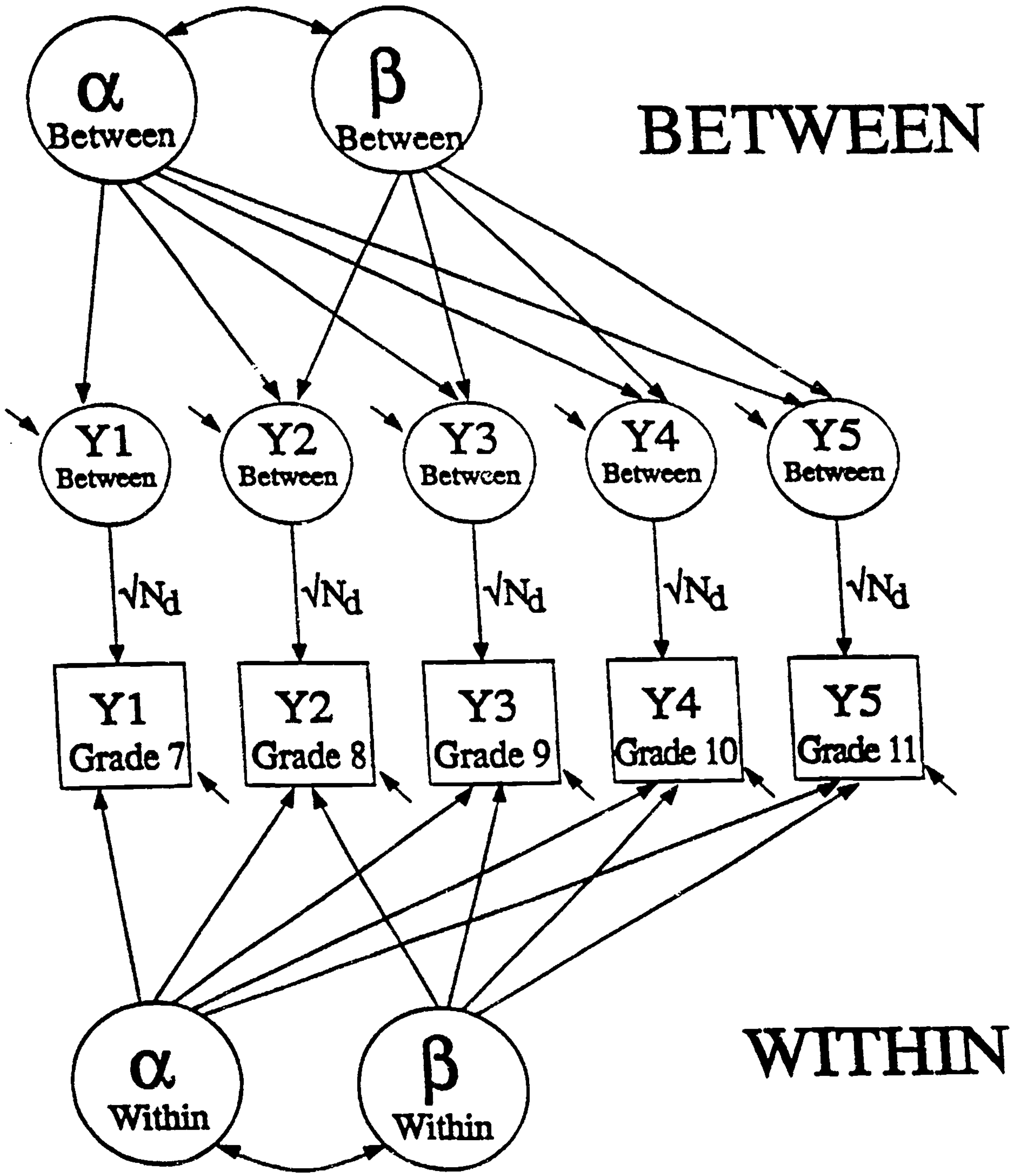


Figure 1. Latent variable growth model formulation for two-level, five-wave data.

while the means on the within side are fixed at zero. This implies that dummy zero means are entered for the within group. The degrees of freedom for the chi-square test of model fit obtained in conventional software then needs to be reduced by the number of y variables.

Further details and references on latent variable modeling with two-level data are given in Muthén (1994b), also giving suggestions for analysis strategies. Software is available from the author for calculating the necessary sample statistics, including intraclass correlations.

It is clear that the Figure 1 model can be easily 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. Estimates may also 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 (see, e.g., Bock, 1989).

7. Analysis of Two-Wave Achievement Data

We will first consider data from the Second International Mathematics Study (SIMS; Crosswhite et al., 1985) drawing on analyses presented in Muthén (1991, 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 3,724 students observed in 197 classes from 113 schools with class sizes varying 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 were administered as a pretest in the fall of eighth grade and again as a posttest in the spring.

Muthén (1991) 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 tc 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 (1991) 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 & Novick, 1968). Here, however, it seems important to make the distinction shown in Figure 2.

The top panel of Figure 2 corresponds directly to the Lord and Novick case. The three line segments may be seen as representing three different classrooms 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

Table 2

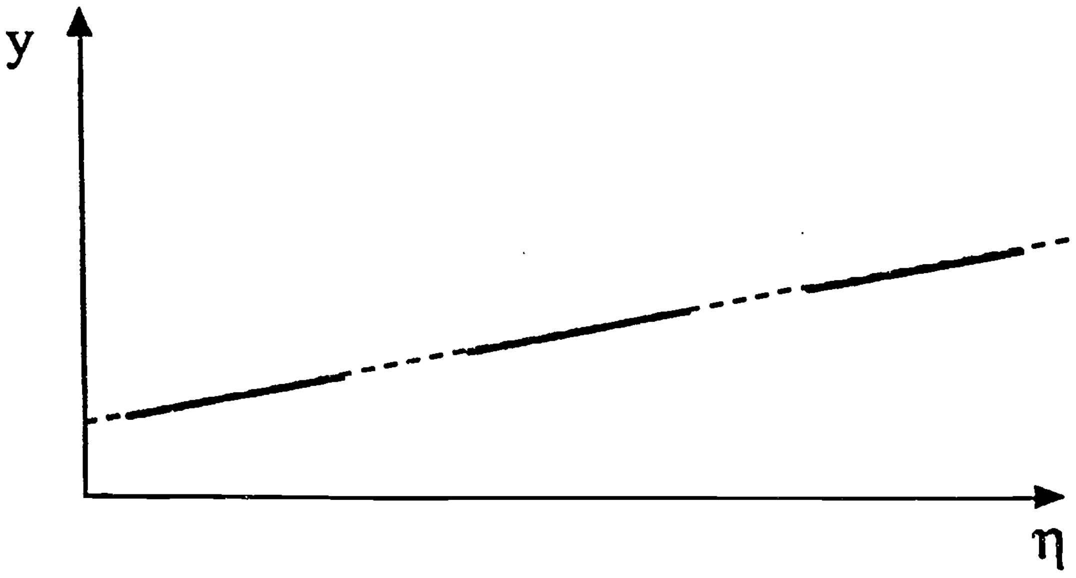
The Second International Mathematics Study: Analysis of Math Achievement From Two Time Points

		Reliabilities					
		Pretest			Posttest		
Variables	# Items	Total	MFA Within	MFA Between	Total	MFA Within	MFA 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

restricted and due to this restriction in range the reliability is attenuated relative to that of all classrooms.

The bottom panel of Figure 2 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 reliability which holds for each classroom.

The Table 2 between reliabilities are considerably higher than the within 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, that is, a total score, and that on the whole the topics measure this dimension rather similarly.



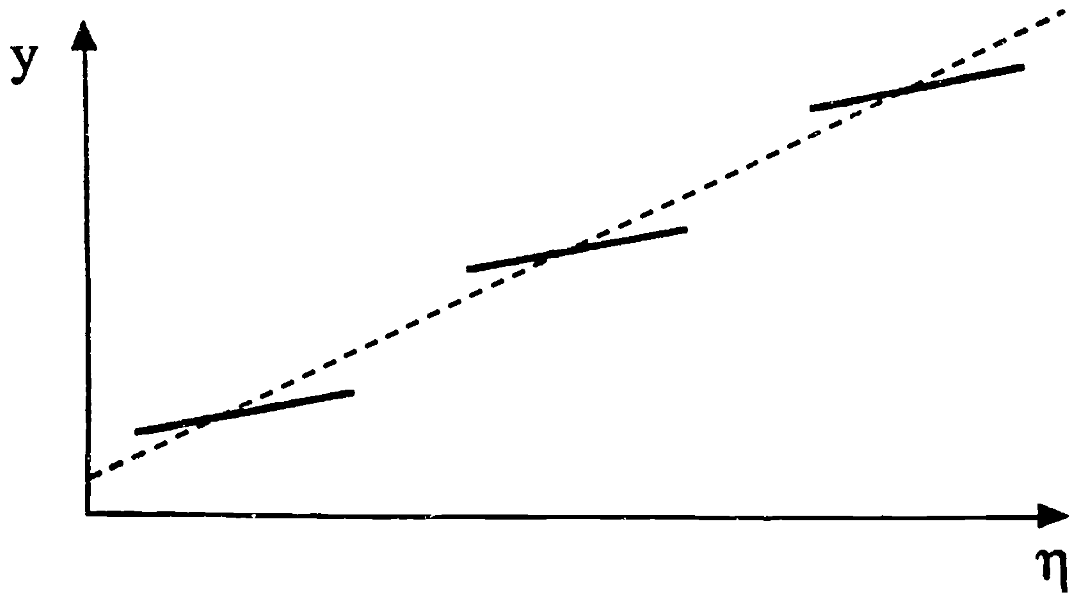


Figure 2. Regressions of an indicator on its latent variable.

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") are 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 (see, e.g., Schmidt, Wolfe, & 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 are, however, that they are likely to be attenuated by the influence of measurement error. This is because

Table 3

The Second International Mathematics Study: Analysis of Math Achievement From Two Time Points

		Intraclass Correlations (proportion between classroom variance)			
		ANOVA		MFA	
Variables	# Items	Pre	Post	Pre	Post
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

student-level measurement error adds to the within-part of the total variance, that is, 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 change in the opposite direction from the ANOVA results. Results from ANOVA would therefore give misleading evidence for answering the questions posed in Muthén (1991).

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 towards science and mathematics. It is conducted as a longitudinal survey of two cohorts spanning Grades 7–12. LSAY uses a national probability sample of about 50 public schools, testing 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 socio-economic status (SES) of the family. The data to be analyzed consist of a total sample of 1,869 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, 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 socio-economic 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.

The analysis considers a growth model extending 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. 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 variables) 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 sizeable part of the student population there is an initial positive attitude about math which 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 are allowed for as indicated in the figure. It should be noted, however, 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

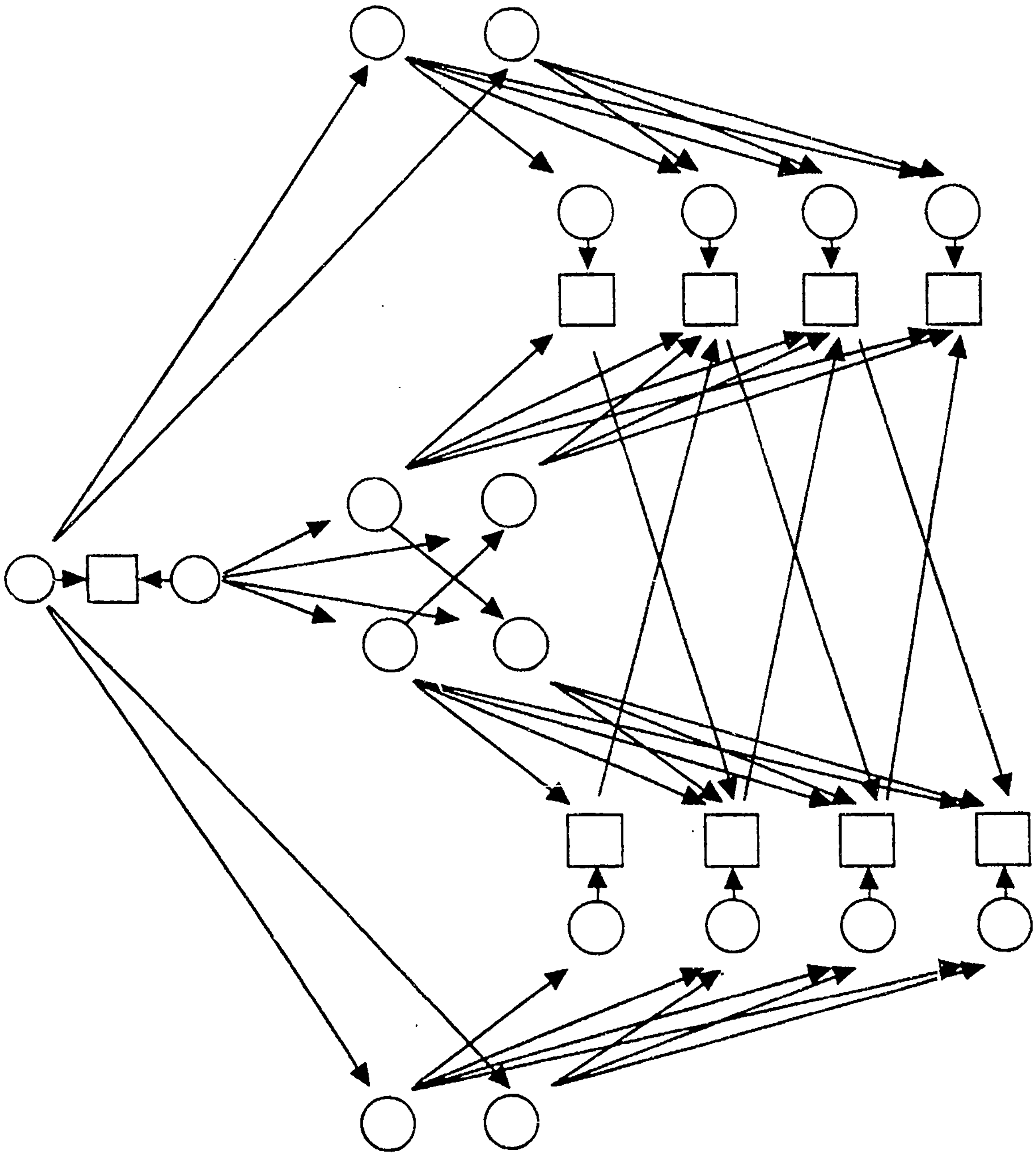


Figure 3. Two-level, four-wave growth model for achievement and attitude related to socio-economic status.

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

For simplicity in the analyses to be presented, the two-group ad hoc estimator discussed in Section 4 will be used and 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. Consequently, statements about significance and model fit should not be interpreted in exact terms.

It is of interest to first ignore the hierarchical nature of the data and give the incorrect tests of fit for the single-level analogs of the auto-regressive and

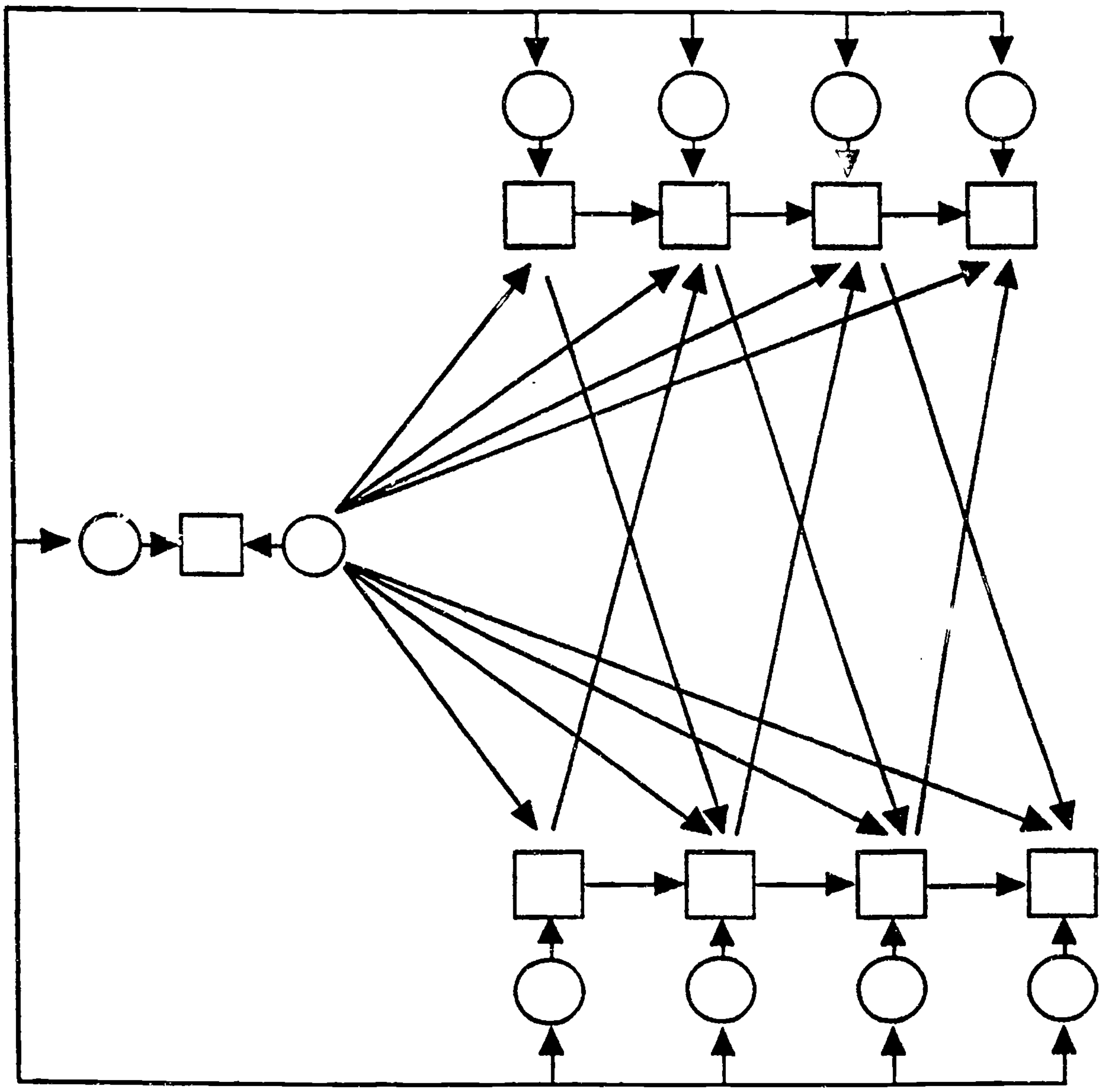


Figure 4. Two-level, four-wave auto-regressive model for achievement and attitude related to socio-economic status.

growth models. To this aim, the conventional maximum-likelihood fitting function is used. The lag-one auto-regressive model obtained a chi-square value of 534.7 with 12 degrees of freedom. To improve fit it was necessary to include a lag-three model for the auto-regressive part and this gave a chi-square value of 22.3 with 6 df. The correct two-level tests of fit using the lag-one model of Figure

4 resulted in a chi-square value of 518.8 with 12 df, while a two-level, lag-three model gave a chi-square value of 28.1 with 6 df. 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 analyzed (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 auto-regressive structure of the model, however, makes it a rather complex and unelegant representation of the data.

Turning to the growth model, the single-level model which ignores the hierarchical nature of the data obtained the incorrect chi-square value of 44.0 with 8 df ($p < 0.001$). The two-level model obtained the chi-square value of 68.4 with 39 df ($p = 0.003$). This may perhaps be regarded as a reasonable fit at n = 1,869. The estimates of this model are given in Table 4.

What is particularly interesting about the two-level growth model is that in contrast to the auto-regressive model, none of the student-level cross-lagged effects are significantly different from zero. This makes for a very parsimonious 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 positive (0.27) while the slope factor correlation is ignorable (0.08). The influences 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 which later becomes less positive. SES explains 12% of the student variation in the achievement intercept while it explains only 1% 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.

Table 4

Results From Two-Level Random Coefficient Growth Model (n = 1,869)

	χ^2 (39) = 68.38 Parameter estimates	t-values
Within		
Cross-lags		
Timepoints		
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
Achievement Initial Status → Attitude 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
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

Table 4 (continued)

	χ^2 (39) = 68.38 Parameter estimates	t-values
Growth Curve		
Achievement		
7th Grade	0*	
8th Grade	1*	
9th Grade	2.60	12.81
10th Grade	3.85	11.86
Attitude		
7th Grade	0*	
8th Grade	1*	
9th Grade	2*	
10th Grade	3*	
Growth Rate Intercept		
Achievement	2.37	9.82
Attitude	-0.32	-9.10
Between		
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	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, Gro.vth 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.

In the school-level part of the model, the correlation between achievement 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, however, 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.

References

Bock, R. D. (1989). *Multilevel analysis of educational data.* San Diego, CA: Academic Press.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage Publications.

Goldstein, H. I. (1987). *Multilevel models in educational and social research.* London: Oxford University Press.

Jöreskog, K. G., & Sorbom, D. (1977). Statistical models and methods for analysis of longitudinal data. In D. J. Aigner & A. S. Goldberger (Eds.), *Latent variables in socio-economic models* (pp. 285-325). Amsterdam: North-Holland

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

McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, *58*, 110-133.

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.

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

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

Muthén, B. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics*, *22*, 43-65.

Muthén, B. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, *54*, 557-585.

Muthén, B. (1990). *Mean and covariance structure analysis of hierarchical data.* Paper presented at the Psychometric Society meeting in Princeton, NJ. (UCLA Statistics Series #62, August 1990.)

Muthén, B. (1991). Analysis of longitudinal data using latent variable models with varying parameters. In L. Collins & J. Horn (Eds.), *Best methods for the analysis of change. Recent advances, unanswered questions, future directions* (pp. 1-17). Washington, DC: American Psychological Association.

Muthén, B. (1993). Latent variable modeling of growth with missing data and multilevel data. In C. R. Rao & C. M. Cuadras (Eds.), *Multivariate analysis: Future directions 2* (pp. 199-210). Amsterdam: North-Holland.

Muthén, B. (1994a). Multilevel covariance structure analysis. In J. J Hox & I. G. Kreft (Eds.), *Multilevel analysis methods* (special issue) *Sociological Methods & Research* (Vol. 22, pp. 376-391). Thousand Oakes, CA: Sage Publications.

Muthén, B. (1994b). *Latent variable growth modeling with multilevel data.* Presented at the UCLA conference Latent Variable Modeling with Applications to Causality, March, Los Angeles.

Muthén, B., & Satorra, A. (1993). *Complex sample data in structural equation modeling.* Under review.

Wheaton, B., Muthén, B., Alwin, D., & Summers. G. (1977). Assessing reliability and stability in panel models. In D. R. Heise (Ed.), *Sociological methodology* (pp. 84-136). San Francisco: Jossey-Bass.

Willett, J. B., & Sayer, A. G. (1993). Using covariance structure analysis to detect correlates and predictors of individual change over time. Harvard University. Forthcoming in *Psychological Bulletin*.