Addressing Questions Concerning Equity in Longitudinal Studies of School Effectiveness and Accountability: Modeling Heterogeneity in Relationships Between Initial Status and Rates of Change

CSE Report 614

Kilchan Choi and Michael Seltzer
CRESST / University of California, Los Angeles

October 2003

Center for the Study of Evaluation
National Center for Research on Evaluation,
Standards, and Student Testing
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90095-1522
(310) 206-1532

Project 3.1 Validity of Measures of Progress
Kilchan Choi and Michael Seltzer, Project Co-Directors, CRESST/UCLA

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R305B60002, as administered by the Institute of Education Sciences, U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the positions or policies of the National Institute on Student Achievement, Curriculum, and Assessment, the Institute of Education Sciences, or the U.S. Department of Education.

ADDRESSING QUESTIONS CONCERNING EQUITY IN LONGITUDINAL STUDIES OF SCHOOL EFFECTIVENESS AND ACCOUNTABILITY: MODELING HETEROGENEITY IN RELATIONSHIPS BETWEEN INITIAL STATUS AND RATES OF CHANGE

Kilchan Choi
CRESST/University of California, Los Angeles

Michael Seltzer
CRESST/University of California, Los Angeles

Abstract

Attending to the relationship between where individuals start (e.g., their initial status) and how rapidly they progress (e.g., their rates of change) can help draw attention to possible concerns regarding the distribution of student achievement within schools in longitudinal studies of school effectiveness. Focusing on the relationship between initial status and rates of change, we address questions concerning equity: Why is it that student achievement is distributed in a more equitable fashion in some schools than in other schools? To what extent might this be due to school characteristics, school policies, or school practices? Do schools that start off with high mean achievement have a more positive relationship or negative relationship? To address questions of this kind, we present a latent variable regression modeling strategy that incorporates latent variable regression into a three-level hierarchical modeling framework (LVR-HM3). To illustrate key ideas and the distinctive features of the LVR-HM3, we fit a series of LVR-HM3s to the data from the Longitudinal Study of American Youth (LSAY) using the Gibbs sampler. We also present results from sensitivity analyses that involve employing *t*-distributional assumptions in LVR-HM3s, and we examine the convergence of the Gibbs sampler using different formulations and parameterization of the LVR-HM3. In a final section, we discuss some implications and possibilities that arise in longitudinal multi-site intervention studies using LVR-HM3s.

Introduction

This paper is motivated by the importance of examining the relationship between where students start (i.e., initial status) and how rapidly they progress (i.e., rates of change) in longitudinal studies (Blomqvist, 1977; Muthen & Curran, 1997; Rogosa & Willett, 1985; Seltzer, Choi, & Thum, 2001a, 2001b). In the area of educational indicators, for example, the relationship between initial status and rates of change might be considered an indicator of how equitably student achievement is distributed within schools (Seltzer, Choi, & Thum, 2001a). When the relationship is positive, for example, for students with relatively low initial levels of achievement, their growth in achievement tends to be slower than students with relatively high initial status. As a result, initial gaps in achievement among students become magnified over time, and student achievement is distributed in an inequitable fashion. In contrast, in the case of a negative relationship, as initial status increases, growth rates tend to decrease. Thus, initial gaps among students tend to diminish over time. Furthermore, we know that relationships of these kinds are likely to vary across schools. In some schools, students' initial levels of math achievement might be extremely consequential in terms of their rates of change (i.e., those students who start high might progress rapidly, while those who start low might progress very slowly). In other schools, however, students' initial status might be far less consequential in terms of their rates of change.

Along these lines, our attention is naturally drawn to questions of why these relationships are different across schools: How do the differences in various kinds of school characteristics relate to the differences in the within-school relationship between initial status and rates of change? In other words, why is it that student achievement is distributed in a more equitable fashion in some schools than in other schools? To what extent might this be due to school characteristics, school policies, or school practices? What are the school-level factors that are related to this underlying distribution of student achievement? Do schools that start off with high mean achievement have a more positive relationship or negative relationship?

Addressing the above questions implies the need to analyze data that have a three-level hierarchical structure. The repeated measures over time (level-1 units) are nested within students (level-2 units) who in turn are nested within different classrooms or schools (level-3 units). Furthermore, such questions involve combining two key modeling features: a) a latent variable regression in a within-school (level-2) model in which student rates of change are regressed on initial status, and b) treating the latent variable regression coefficients as varying across schools in a between-school (level-3) model.

In this paper, we propose a latent variable modeling strategy that integrates latent variable regressions in a 3-level hierarchical modeling framework (LVR-HM3). Through analyses of the data from a sub-sample of the Longitudinal Study of American Youth (LSAY; Miller, Kimmel, Hoffer, & Nelson, 2000), we wish to show how LVR-HM3s provide a way of illuminating important differences across schools in the relationship between initial status and rates of progress, and how it can enrich the kinds of questions concerning equity that we are able to examine in studies of school effectiveness.

This paper consists of the following sections. First, we outline our approach in the following section. We then present a brief overview of other strategies for combining hierarchical modeling techniques with latent variable modeling strategies. In the illustrative example section, we fit a series of LVR-HM3s to the LSAY data and illustrate its use. Next, we conduct sensitivity analyses by employing t distributional assumptions at levels 1, 2, and 3. We then compare the convergence of the Gibbs sampler using different formulations and parameterizations of the LVR-HM3. In the final section of this paper, we will discuss possible implications and extensions of the LVR-HM3.

Latent Variable Regression in a Hierarchical Modeling Framework

Latent Variable Regression in a 2-Level Hierarchical Model (LVR-HM2)

For heuristic purpose, we now specify a simple 2-level hierarchical model (2-level HM) for longitudinal analysis. We then incorporate a latent variable regression (LVR) into this model. Building upon this latent variable regression 2-level hierarchical model, we specify a latent variable regression 3-level hierarchical model (LVR-HM3).

In growth curve analyses, the n_i repeated observations nested within person i are modeled as a function of time ($Time_{ti}$) in a level-1 or within-individual model. Modeling change as a linear function of time we have:

$$Y_{ti} = \pi_{0i} + \pi_{1i}Time_{ti} + \varepsilon_{ti} \qquad \varepsilon_{ti} \sim N(0, \sigma^2), \qquad (1)$$

where π_{0i} represents the status of person i at $Time_{ti} = 0$, and π_{1i} is the growth rate for person i. The ε_{ti} are residuals assumed normally distributed with mean 0 and variance σ^2.

The hallmark of the HM is that individual growth parameters are treated as varying across individuals. The variability is represented in a level-2 or between-individual model, which often contains predictors capturing information regarding students' background characteristics, their educational experiences, and the like. For illustrative purposes, we pose a simple level-2 model that does not contain any predictors:

$$\pi_{0i} = \beta_{00} + r_{0i} \qquad r_{0i} \sim N(0, \tau_{00})$$

$$\pi_{1i} = \beta_{10} + r_{1i} \qquad r_{1i} \sim N(0, \tau_{11}),\ Cov(r_{0i}, r_{1i}) = \tau_{01} = \tau_{10}. \qquad (2)$$

In the above equation, β_{00} represents the mean initial status and β_{10} the mean growth rate for the population of interest; the level-2 residuals (i.e., random effects) r_{0i} and r_{1i} capture the deviation of initial status for person i from β_{00}, and the deviation of the growth rate for person i from β_{10}, respectively. We further assume that the r_{0i} and r_{1i} are normally distributed with mean 0, variance τ_{00} and τ_{11}, respectively, and with covariance τ_{01}($Cov(r_{0i}, r_{1i}) = \tau_{01}$).

Specifying models to investigate how differences in initial status relate to differences in rates of change in essence implies modeling individual growth rate parameters (π_{1i}) as a function of individual initial status parameters (π_{0i}). In other words, we need to treat initial status not only as a dependent variable in the level 2 model but also as a predictor variable for the rate of change at level 2 (see Equation 3).

$$\pi_{0i} = \beta_{00} + r_{0i} \qquad r_{0i} \sim N(0, \tau_{00})$$

$$\pi_{1i} = \beta_{10} + b(\pi_{0i} - \beta_{00}) + r_{1i} \quad r_{1i} \sim N(0, \tau_{11}) \quad Cov(r_{0i}, r_{1i}) = 0. \qquad (3)$$

In the second equation, individual growth rates (π_{1i}) are modeled as a function of initial status (π_{0i}). The key parameter of interest is b, which captures the expected change in growth rate when initial status increases one unit. b is termed a latent variable coefficient. In the above equation, β_{10} represents the expected rates of growth when initial status is equal to the grand mean. The variance for the rate of change (τ_{11}) represents the amount of variation in growth rates that remains after taking into account initial status. Since we are conditioning on initial status in the level-2 model for growth rates ($\pi_{1i} \mid \pi_{0i}$), we assume that $Cov(r_{0i}, r_{1i}) = 0$.

Latent variable regression in a 3-level hierarchical model. We now extend the above two-level model and specify a simple LVR-HM3 in which we do not include any student- and school-level observed predictors. At level 1 (see Equation 4), the outcome of interest, Y_{tij}, for person *i*, in school *j*, at time *t*, is modeled as a function of $Time_{tij}$. In a level-2 (within-school/between-individual) model, rate of change for student *i* in school *j* (π_{1ij}) is modeled as a function of a student's initial status.

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}Time_{tij} + \varepsilon_{tij} \qquad \varepsilon_{tij} \sim N\,(0, \sigma^2) \qquad (4)$$

$$\pi_{0ij} = \beta_{00j} + r_{0ij} \qquad r_{0ij} \sim N\,(0, \tau_{\pi 0j})$$

$$\pi_{1ij} = \beta_{10j} + \mathrm{Bw}_j\,(\pi_{0ij} - \beta_{00j}) + r_{1ij} \qquad r_{1ij} \sim N\,(0, \tau_{\pi 1j}). \qquad (5)$$

In this model, the latent variable regression coefficient, Bw_j, captures the relationship between initial status and rate of change for students in school *j*. In other words, compared to the latent variable regression coefficient, b, in Equation 3, each of *J* schools (*j*=1, . . . , *J*) has a different latent variable regression coefficient. Thus, we refer to the within-school latent variable regression coefficients, Bw_j, as within-school initial status/rate of change slopes (Seltzer, Choi & Thum, 2001a). Furthermore, the level-2 random effects for school *j* (i.e., r_{0ij}, and r_{0ij}) are assumed independently and normally distributed with mean 0 and variances $\tau_{\pi 0j}$ and $\tau_{\pi 1j}$, respectively. Importantly note that we allow the variances of r_{0ij} and r_{0ij} to differ across schools. Additionally, we can also include student characteristics both in the above two equations.

At level 3, level-2 parameters (i.e., school mean initial status (β_{00j}), the expected rate of change for school *j* (β_{10j}), and the within-school initial status/rate of change slope for school *j* (Bw_j) are viewed as varying across schools. That is, level-2 parameters are treated as outcomes in a level-3 (between-school) model. Interestingly, we can pose a model in the second equation below where school mean rate of change (β_{10j}) is modeled as a function of school mean initial status (β_{00j}). In addition, the within-school initial status/rate of change slope (Bw_j) can be also modeled as a function of school mean initial status.

$$\beta_{00j} = \gamma_{000} + u_{00j} \qquad u_{00j} \sim N\,(0, \tau_{\beta 00})$$

$$\beta_{10j} = \gamma_{100} + \mathrm{Bb}(\beta_{00j} - \gamma_{000}) + u_{10j} \qquad u_{10j} \sim N\,(0, \tau_{\beta 10})$$

$$\mathrm{Bw}_j = \mathrm{Bw}_0 + \mathrm{Bw}_1(\beta_{00j} - \gamma_{000}) + u_{\mathrm{Bw}j} \qquad u_{\mathrm{Bw}j} \sim N\,(0, \tau_{\mathrm{Bw}}) \qquad (6)$$

In the first equation, γ_{000} represents grand mean initial status. In the second equation, a latent variable regression coefficient, Bb, captures the between-school relationship between mean initial status and mean rates of change. Thus, in contrast to the within-school initial status/rate of change slope(Bw_j), we term this latent variable regression coefficient the between-school mean initial status/mean rate of change slope. In the last equation, Bw_0 is the expected within-school initial status/rate of change slope when school mean initial status is equal to the grand mean. Bw_1 is also a latent variable regression coefficient which captures the effect of school mean initial status on the within-school initial status/rate of change slope.

In the above level-3 model, three random effects, u_{00j}, u_{10j} and u_{Bwj} are assumed normally distributed with mean 0 and variances, $\tau_{\beta 00}$, $\tau_{\beta 10}$, τ_{Bw}, respectively. Regarding covariance components in the above model, we assume that $Cov\,(\beta_{00j}, \beta_{10j}) = 0$ and $Cov\,(\beta_{00j}, Bw_j) = 0$, because both β_{10j} and Bw_j are conditioned on β_{00j} in the second and third equations above. However, the covariance between β_{10j} and Bw_j is defined to be $\tau_{\beta 10,Bw}$ ($Cov\,(\beta_{10j}, Bw_j) = \tau_{\beta 10,Bw}$).

Using the above level-3 model as a baseline model, measures of various school-level characteristics can be entered as predictors in the above equations. By doing this, in studies on school effectiveness, we can identify factors that appear to eventuate in high mean rates of progress, and in relatively weak relationships, or even negative relationships, between initial status and rates of change.

From an estimation standpoint, we employ a fully Bayesian approach using Markov Chain Monte Carlo (MCMC) method (e.g., the Gibbs sampler) (see, e.g., Carlin & Louis, 1996; Gelfand & Smith, 1990; Gelman, Carlin, Stern, & Rubin, 1995; Gilks, Richardson, & Spiegelhalter, 1996; Tanner, 1996). That is, we use the Gibbs sampler to simulate the marginal posterior distributions of parameters of interest (e.g., Bw_1, Bb). The resulting posteriors provide the bases of point estimates and intervals. Thus, while LVR-HM3s are extremely complex from an estimation standpoint, it is the use of MCMC that makes fitting models of this kind very feasible. Furthermore, to study the sensitivity of inferences to outlying time-series data points, outlying individuals, and outlying schools, we can readily use MCMC to fit LVR-HM3s under *t* distributional assumptions at each level. To implement this approach, we use the software program WinBUGS (Spiegelhalter, Thomas, & Best, 2000), which provides a fairly easy means of implementing the Gibbs sampler in a wide range of modeling settings. Note that BUGS is a near-acronym for *Bayesian analysis using Gibbs sampling*.

Review of Various Extensions of Hierarchical Modeling and Structural Equation Modeling for Analyzing Multilevel Longitudinal Data

Growth modeling techniques have been widely used in many disciplines for studying individual change. One tradition generally found in psychometrics is referred to as latent curve analysis or latent variable Structural Equation Modeling (SEM; Muthen & Curran, 1997). Ever since Meredith and Tisak (1984, 1990) demonstrated how covariance structure analysis could be applied to longitudinal data, many extensions of this work can now be found in the latent variable literature (Chou, Bentler & Pentz, 1998; McArdle & Epstein, 1987; Muthen, 1991; Muthen & Curran). This tradition has focused on variance-covariance structure analysis using latent variables, and regressions among latent variables constitute a key feature of this approach. Thus, viewing initial status and rates of change as latent variables, and specifying a variance-covariance matrix for these latent variables, we can readily model rates of change as a function of initial status within the SEM framework. However, this approach, as currently implemented in a number of software programs, is problematic when data are not time-structured. (See, however, Muthen, Kaplan, & Hollis, 1987; Muthen & Muthen, 2002.)

A second tradition is referred to as random coefficient modeling (Gibbons, Hedeker, Waternaux, & Davis, 1988; Laird & Ware, 1982; Liang & Zeger, 1986), Hierarchical Linear Modeling (HLM; Bryk & Raudenbush, 1992), and multilevel modeling (Goldstein, 1987, 1995; Longford, 1993). This approach is more flexible in terms of data structures in that each individual is allowed to have a different number of observations and different spacings between observations (see, e.g., Bryk & Raudenbush, 1987; Muthen & Curran, 1997; Raudenbush, 1998; Willett & Sayer, 1994). A key feature of this approach is that individual initial status and rates of change are treated as varying across individuals (i.e., random coefficients; for comparisons of these two traditions, see Muthen & Curran and Willett & Sayer).

Recently, several pioneering researchers have developed strategies for incorporating latent variable regressions in hierarchical modeling settings (Chou et al., 2000; Muthen, 1997; Raudenbush & Sampson, 1999). First, Raudenbush and Sampson's approach can be easily applied to settings where we wish to regress rates of change on initial status, and it can be implemented in the latest release of the

HLM software program (Raudenbush, Bryk, Cheong, & Congdon, 2000). In this strategy, the regression coefficients for latent predictors (e.g., initial status) are estimated by a two-stage procedure. At the first stage, we estimate fixed effects and a variance-covariance matrix by fitting a more standard model (i.e., one that does not contain latent variable regressions). At the second stage, we use the elements of the estimated variance-covariance matrix to obtain estimates of latent variable regression coefficients; we then use these LVR estimates and the fixed effects estimates from the standard model to obtain various fixed effects estimates in the latent variable regression model (e.g., estimates of direct effects; see Raudenbush and Bryk [2001, Chapter 11] for an example). Thus, for example, to estimate b in Equation 3, first we fit the 2-level model to the data defined by Equations 1 and 2. Then, as a second stage in this process, we estimate b by dividing the estimate of τ_{01} by the estimate of τ_{00} : $\hat{b} = \dfrac{\hat{\tau}_{01}}{\hat{\tau}_{00}}$.

Even though Raudenbush and Sampson's (1999) strategy enables us to specify latent variable regressions in the hierarchical modeling framework, it has the following limitations. In the current implementation of the HLM program, when we apply this strategy to three-level hierarchical data sets, we are able to regress school mean rate of change on school mean initial status at level 3. However, we are not able to regress students' rates of change on their initial status at level 2. This kind of limitation precludes the possibility of modeling differences in the within-school relationship between initial status and rate of change as a function of various school or site characteristics.

A second approach can be founded in the work of Muthen (1994) who presents multilevel structural equation models (see also Muthen, 1997; Muthen & Satorra, 1995). In this approach, the sample data are decomposed into a pooled within-group matrix and a scaled between-group covariance matrix. Based on these matrices, the structural relations among within- and between-group variables can be estimated simultaneously using a multiple-group modeling technique.

His approach provides us with very useful and important latent variable regression modeling techniques for analyzing multilevel data. From an estimation standpoint, however, it is very computationally cumbersome and challenging when data are unbalanced (i.e., when cluster sizes [n_j] substantially vary across clusters). In such cases, full information maximum likelihood (FIML) estimation implies specifying a separate between-group model for each distinct group size so that large numbers of distinct group sizes make FIML extremely complex. As an alternative to FIML, Muthen proposes using a single between-group sampling covariance matrix based on a simpler ML-based estimate (MUML; Muthen, 1994). This approximation of the between-group sampling covariance reduces the computational complexity a great deal and provides fairly acceptable and accurate results for unbalanced nested data with large numbers of both level-1 and level-2 units (McDonald, 1994; Muthen,1994; Hox, 1993). However, a large simulation study conducted by Hox and Mass (2000) shows that small numbers of level-2 units and low intraclass correlations yield underestimates of residual variances, and standard errors for parameters in the between-group model that are too small.

We now consider limitations of the multilevel SEM approach from a modeling standpoint. Essentially, Muthen's approach summarizes unbalanced nested data structures by means of within- and between-group variance-covariance matrices, and his approach may be viewed as a random intercept model (Muthen, 2002). Thus, within this kind of modeling framework, it is not possible to include random slopes (Raudenbush, 2001; however, see Muthen & Muthen (2002) regarding the possibility of incorporating random slopes in 2-level latent variable models). In three-level hierarchical modeling settings where the numbers of level-2 and level-3 units are fairly large, the multilevel SEM approach allows us to estimate the average within-school relationship between initial status and rates of change and the between-school relationship between mean initial status and mean rate of change using the pooled within-school covariance matrix and the between-school covariance matrix, respectively. However, as in the case of Raudenbush and Sampson's approach, we cannot treat latent variable regression coefficients as varying across schools.

Finally, Chou et al. (2000) present a two-stage approach to multilevel SEM for multilevel data. Their two-stage approach is somewhat analogous to Burstein's (1980) "slopes-as-outcomes" approach to analyzing hierarchically structured data. At the first stage, they pose a very general mean and / or covariance structure model at the first level for each nesting unit. The estimated parameters for each of the nesting units provide a new data matrix for the second stage model. They note that these estimates provide a model-based summary of the level-1 data: estimates of the model-based mean for each nesting unit and the structural relations among level-1 variables. For three-level hierarchical data, for example, in the first stage of the procedure, we fit the LVR-HM2 depicted in Equations 1 and 3 to each school's data.

As a result, we obtain three sets of parameter estimates: estimates of the mean initial status (β_{00j}), the expected rate of change (β_{10j}), and the latent variable regression coefficient (Bw_j) for each school. In the second stage of the procedure, using these estimated parameters as a data set, we fit the level-3 model as presented in Equation 6.

As they point out, their less sophisticated two-stage estimation procedure may be problematic in certain situations. They argue that their two-stage estimation strategy performs efficiently for situations where the number of observations at the second (e.g., the number of students within a school) and the highest levels (e.g., the number of schools) are sufficiently large to rely on standard large-sample theory. The results based on fitting the level-3 model, however, do not take into account estimation error in the estimates of the level-2 parameters (i.e., β_{00j}, β_{10j}, Bw_j). In particular, due to estimation error in school mean initial status, which is used as a predictor for school mean rate and for the within-school initial status / rate of change slope, estimates of the latent variable regression coefficients in Equation 6 may tend to be attenuated. This may be especially problematic when level-2 sample sizes are small or moderate.

An Illustrative Example Using Data From LSAY

To help illustrate various key points and ideas of the LVR-HM, we fit a series of LVR-HM3 to analyze the time-series data for students in 45 schools in the Longitudinal Study of American Youth (LSAY; Miller et al., 2000). In this paper we focus on mathematics achievement scores collected at the start of Grades 7, 8, 9, and 10 in 45 different schools. The total number of students in the sample is 2,628 and the average number of students per school is 58. The sample included approximately 31 teachers per school.

This section presents a series of LVR-HM3s. In the first model (Model 1), we specify a latent variable regression in the within-school (level-2) model. As a result, we can estimate within-school initial status / rate of change slopes for each of the 45 schools. In the between-school (level-3) model, correlation coefficients are estimated among school mean initial status, school mean rate of change, and the within-school initial status / rate of change slopes.

In Model 2, instead of estimating correlation coefficients at level-3, we include latent variable regressions at level-3: school mean rate of change is regressed on school mean initial status, and the within-school initial status / rate of change slope is also regressed on school mean initial status.

Finally, viewing Model 2 as a baseline model, in Model 3 we include student-level predictors for initial status and rates of change in the level-2 model. In addition, school-level characteristics are included as predictors for outcomes in the level-3 model.

Unconditional LVR-HM3 (Model 1)

We first begin by specifying a simple LVR-HM3 (Model 1) in which no student-level predictors or school-level predictors are included in the model. However, we include initial status as a predictor for rates of change in the within-school (level-2) model. The following within-student (level-1) model is posed for the time-series data for each of the students in each of the 45 schools.

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}(Grade_{tij} - 7) + \varepsilon_{tij} \qquad \varepsilon_{tij} \sim N(0, \sigma^2) \qquad (7)$$

In the above level-1 (within-student) model, a student's math scores from grade 7 through grade 10 (Y_{tij}) are modeled as a function of a time variable ($Grade_{tij}$). $Grade_{tij}$ represents the grade for student *i* in school *j* at measurement occasion *t*, and it takes on a value 7 for grade 7, 8 for grade 8, 9 for grade 9 and 10 for grade 10. Since we centered $Grade_{tjj}$ around a value of 7, π_{0ij} represents the expected math achievement score for student *i* in school *j* at the start of grade 7 (i.e., initial status for student *i* in school *j*). π_{1ij} represents the expected rate of change for student *i* in school *j*.

Next, we pose a between-student or within-school (level-2) model with a latent variable regression as follows:

$$\begin{aligned} &\pi_{0ij} = \beta_{00j} + r_{0ij} && r_{0ij} \sim N(0, \tau_{\pi 0j}) \\ &\pi_{1ij} = \beta_{10j} + Bw_j(\pi_{0ij} - \beta_{00j}) + r_{1ij} && r_{1ij} \sim N(0, \tau_{\pi 1j}) \qquad (8) \\ & && Cov(r_{0ij}, r_{1ij}) = 0 \end{aligned}$$

A key parameter of interest in this model is Bw_j, which is a latent variable regression coefficient. Bw_j is a slope parameter relating initial status and rates of change for students in school j. By virtue of centering π_{0ij} around school mean initial status (i.e., β_{00j}), β_{10j} represents expected rate of change for school j when student initial status is equal to the school mean initial status.

Since we include initial status as a predictor for rates of change in the within-school model, $\tau_{\pi 1j}$ represents the residual variance in rates of change after we take into account differences in initial status. Furthermore, including π_{0ij} as a predictor for π_{1ij}, we now assume that the covariance between the residuals, r_{0ij} and r_{1ij}, is equal to 0.

We treat β_{00j}, β_{10j}, and Bw_j as outcomes in a level-3 (between-school) model:

$$\beta_{00j} = \gamma_{000} + u_{00j} \qquad u_{00j} \sim N(0, \tau_{\beta 00})$$

$$\beta_{10j} = \gamma_{100} + u_{10j} \qquad u_{10j} \sim N(0, \tau_{\beta 10})$$

$$\mathrm{Bw}_j = \mathrm{Bw}_0 + u_{\mathrm{Bw}j} \qquad u_{\mathrm{Bw}j} \sim N(0, \tau_{\mathrm{Bw}}) \tag{9}$$

$$Cov(u_{00j}, u_{10j}) = \tau_{\beta 00,\beta 10}, \; Cov(u_{00j}, u_{\mathrm{Bw}j}) = \tau_{\beta 00,\mathrm{Bw}}, \; Cov(u_{10j}, u_{\mathrm{Bw}j}) = \tau_{\beta 10,\mathrm{Bw}}$$

Thus, a key parameter of interest in the above level-3 model is Bw_0, which represents the overall mean of the within-school initial status / rate of change slope (i.e., Bw_j). In addition, $u_{\mathrm{Bw}j}$ captures each school's deviation from the overall mean within-school initial status / rate of change slope. The variability of $u_{\mathrm{Bw}j}$ is captured by τ_{Bw}, and $u_{\mathrm{Bw}j}$ is assumed normally distributed with mean 0 and variance τ_{Bw}.

We also need to pay attention to covariances among the three level-3 random effects. First, the between-school relationship between school mean initial status and school mean rate of change is represented by the covariance, i.e., $Cov(u_{00j}, u_{10j}) = \tau_{\beta 00,\beta 10}$. Second, the covariance between u_{00j} and $u_{\mathrm{Bw}j}$, (i.e., $Cov(u_{00j}, u_{\mathrm{Bw}j}) = \tau_{\beta 00,\mathrm{Bw}}$) provides us with the information about how the within-school initial status / rate of change slopes are associated with the school mean initial status: For example, do within-school initial status / rate of change slopes (Bw_j) increase or decrease as school mean initial status (β_{00j}) increases? Furthermore, $Cov(u_{10j}, u_{\mathrm{Bw}j}) = \tau_{\beta 10,\mathrm{Bw}}$ captures the relationship between school mean growth rates and within-school initial status / rate of change slopes.

For each of the models presented in this paper, we specified normal priors with extremely low precision for the fixed effects, which are functionally equivalent to uniform priors. For example, for γ_{000} in the above model, we specified a normal prior with mean 0 and precision 1.0E-5. Thus the data will dominate the prior in drawing inferences concerning the fixed effects. For the variance components at levels 1, 2 and 3, we employed gently data-determined inverse gamma priors and inverse Wishart priors based on a strategy outlined in Seltzer, Novak, Choi, and Lim (2002) (see Endnote 1).

Upon convergence, the Gibbs sampler essentially provides us with draws from the joint posterior distribution of all unknowns in a given model. The empirical distribution of the deviates generated for a parameter of interest over a large number of iterations provides us with an accurate approximation of the marginal posterior distribution of that parameter. A marginal posterior distribution provides us with a summary of the plausibility of different values for a parameter of interest given the data.

For assessing convergence and mixing, we examined trace plots and autocorrelation function (ACF) plots. Note that for each model we ran two or more chains using different starting values with a burn-in period of 2,000 and then each chain was run for an additional 30,000 iterations. We compared results based on the two chains and found them to be extremely similar. Using a Pentium IV 2.5 GHz machine, approximately 10 minutes of CPU time were required to complete 30,000 iterations of our algorithms. To help ensure results with high degrees of accuracy, we employed a burn-in period of 2,000 iterations, and used the output from 60,000 subsequent iterations of the Gibbs sampler to simulate marginal posteriors of interest.

In Table 1, we present the mean (labeled *estimate*), median, standard deviation, and 95% interval of the marginal posterior distribution of each parameter in Model 1. The .025 and .975 quantiles of a marginal posterior distribution for a parameter of interest provides the basis of constructing a 95% interval for that parameter. In addition, we calculated the proportion of the posterior distribution greater than 0 for various parameters.

The average within-school initial status / rate of change slope (Bw_0) is equal to 0.113 and its 95% interval includes only positive values. This estimate can be

Table 1

Model 1: 3-level unconditional LVR Model – Estimating Within-School Initial Status / Rate of Change Slope (Bw_j)

	Estimate	SE	95% Interval	Median	Prop. >0
Fixed effect:					
Mean Init. Status(γ_{000})	49.72	0.63	(48.48, 50.96)	49.72	1.000
Mean Rate of Change(γ_{100})	3.85	0.14	(3.57, 4.12)	3.85	1.000
Mean Init. Status / Rate of Change Slope(Bw_0)	0.113	0.012	(0.089, 0.136)	0.113	1.000
Variance Components:					
Level-1 Error(σ^2)	17.16	0.37	(16.44, 17.90)	17.16	
Level-3 Variance					
Sch. Initial Status($\tau_{\beta 11}$)	16.63	3.952	(10.48, 25.86)	16.080	
Sch. Rate of Change($\tau_{\beta 22}$)	0.736	0.198	(0.430, 1.201)	0.708	
Init. Status / Rate of Change Slope (τ_{BW})	0.003	0.001	(0.002, 0.006)	0.003	
Correlations:					
Sch. Init. Status, Sch. Rate of Change ($Corr_{\beta 00j,\beta 10j}$)	0.33	0.15	(0.00, 0.61)	0.34	.9760
Sch. Init. Status, Init / Rate of Change Slope($Corr_{\beta 00j,\, Bwj}$)	-0.39	0.19	(-0.71, 0.01)	-0.41	.0288
Sch. Rate, Init / Rate of Change Slope($Corr_{\beta 10j,\, Bwj}$)	-0.56	0.16	(-0.81, -0.20)	-0.58	.0025

interpreted as follows: on average, a one-unit increases in initial status is expected to eventuate in an increase of 0.113 points in rate of change. This applies to students with low initial status. As result, on average, the differences in math achievement at grade 7 are magnified as grade increases.

Furthermore, these within-school initial status / rate of change slopes (Bw_j) vary considerably across the 45 schools. The 95% intervals of the Bw_j parameters for the schools in our sample are displayed in Figure 1. Specifically, the top, middle line, and bottom of each bar are equal to, respectively, the .975 quantile, the mean, and .025 quantile of the marginal posterior distribution of the initial status / rate of change slope for a given school. Schools 9, 19, 33, 44 have relatively high Bw_j values. The corresponding estimates (posterior mean) for these schools are, respectively, 0.19, 0.18, 0.17, and 0.16, and the correlation coefficients capturing the relationship between initial status and rate of change are above 0.95 for all four of these schools. In contrast, schools 14, 18, 22, and 28 tend to have low Bw_j values. The 95% intervals

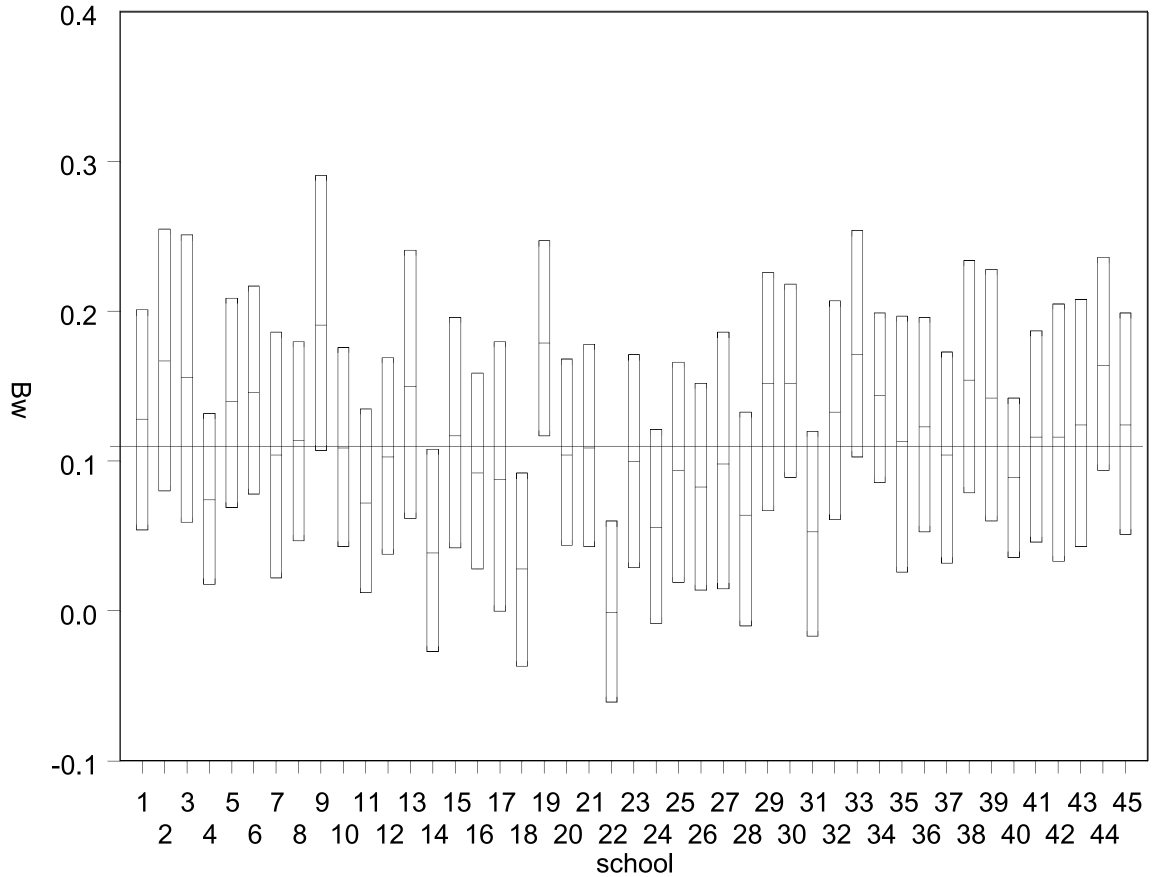


Figure 1. Within-school initial status / rate of change slopes (bw_j) for the 45 schools (Model 1).

* The horizontal line represents the overall average of the within-school initial status/rate of change slopes.

** The top, middle line, and the bottom of each bar corresponds, respectively, to the .975 quantile, mean, and the .025 quantile of the marginal posterior distribution of the initial status/rate of change slope for a given school.

for these schools contain a value of 0, which indicates that initial status and rates of change within these schools are unrelated. In addition, correlation coefficients for these schools do not exceed a value of 0.06.

In addition to Figure 1, we can use our estimates of the grand mean of the Bw_j's (Bw_0) and our estimate of the variance in the Bw_j's across schools (τ_{Bw}). As noted above, our estimate of Bw_0 is .113, and as can be seen in Table 1, our estimate of τ_{Bw} is .003. A school whose initial status / rates of change slope is two standard deviations above the grand mean is equal to: .113 + $2\sqrt{.003}$ = .223. A school whose initial status/rates of change slope is two standard deviations below the grand mean is equal to: .113 - $2\sqrt{.003}$ = .012. In the first school, where the initial status/rate of change slope is equal to .223, student's initial status is very consequential in terms of his or her rate of change in the sense that a 10-point initial difference is expected to widen to 16.69 points at grade 10 (i.e., the expected difference at the first year is equal to 2.33 (.223 * 10 = 2.23), and the expected difference after 3 years is equal to 6.69 (2.23 * 3 = 6.69). In contrast, in the second school, a student's initial level of math achievement is not consequential in terms of how fast he or she progresses. For example, in that school, students 10 points apart in math scores at grade 7 are expected to be only 10.36 points apart at the end of grade 10, since the expected difference at the first-year period is .12 (10 * .012 =.12). And after 3 years, the expected difference is equal to .36 (.12 *3 = .36). As such, we can conclude that there is appreciable variability in Bw_j across schools.

Taking things one step further, how are the within-school initial status/rate of change slopes related to school mean initial status? For example, does the relationship between initial status and rate of change tend to be weak or strong in schools with high mean initial status values? And how strongly are they associated with each other? The posterior mean of the correlation between the within-school slopes and school mean initial status ($Corr_{Bwj,\beta 00j} \mid y$) computed in WinBUGS based on the equation $\frac{\mathrm{cov}(Bw_j, \beta_{00j})}{\sqrt{\mathrm{var}(Bw_j)\,\mathrm{var}(\beta_{00j})}}$ is equal to -0.38 (see Table 1). The 95% interval of this correlation is -.71 to .01 and approximately 2.9% of the mass of the posterior distribution for $Corr_{Bwj,\beta 00j}$ lies above 0. This sizable negative coefficient implies that the initial gaps at grade 7 among students tend to widen appreciably more in low mean initial status schools than in high mean initial status schools.

In addition, we estimated the correlation coefficient between school mean initial status and school mean growth rates. The resulting positive correlation coefficient ($Corr_{\beta 00j,\, \beta 10j}$ = 0.33, $p(Corr_{\beta 00j,\, \beta 10j} > 0 \mid y)$ = .976) suggests that school mean growth trajectories for schools with low mean initial status are flatter than those for schools with high mean initial status.

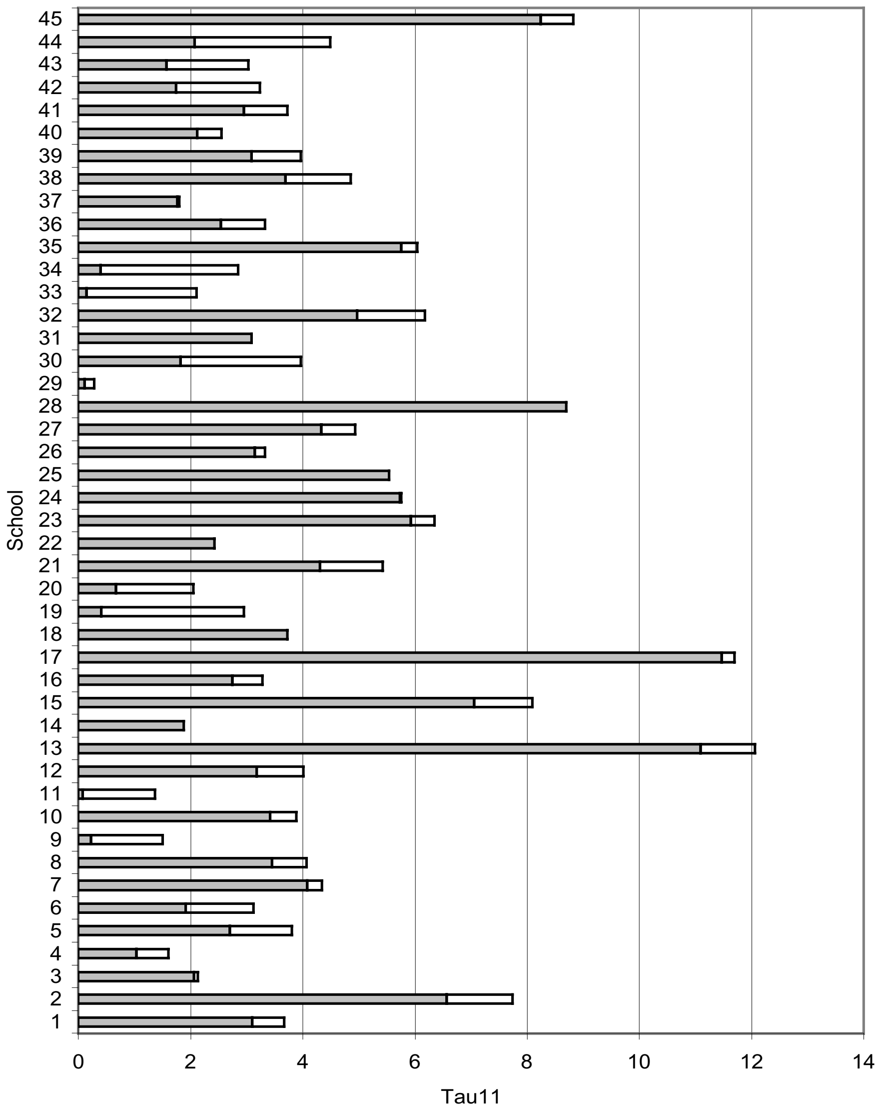
Are the within-school relationships more positive in schools where students on average grow at a faster rate than in schools where their students progress on average at a slower rate? The correlation coefficient reveals that there is a highly significant negative correlation ($Corr_{\beta 10j,Bwj}$ = -0.56; $p(Corr_{\beta 10j,\, Bwj} > 0 \mid y)$ = .025). This coefficient implies that schools in which students on average progress rapidly from grade 7 to grade 10 have weaker within-school relationships.

Next, how important is initial status as a predictor for rate of change? In other words, we consider how much the variance in rate of change in the within-school model is reduced after initial status is included as a predictor. Figure 2 displays the amount of reduction in rate of change variance after initial status is included as a predictor in the level-2 model. In each bar plot, the white portion of the bar indicates the variance accounted for by initial status, while the dark portion of the bar represents the remaining variance.

We can see that there are marked differences across schools in terms of the percentage of reduction in the variance of growth rates. For example, schools 14, 18, 22, 25, 28, and 31 have 0% reduction even after initial status is included in the equation for rates of change. However, in schools 9, 11, 19, 33, and 34, initial status has exceptionally high power in accounting for the differences in rates of change. In these schools, 85.3%, 94.1%, 86.1%, 92.9%, and 86.0%, respectively, of variability is attributed to initial status. Thus, the considerably large variability in the percentage of reduction in the rate of change variance demonstrates that the relationships between initial status and rates of change vary to a large extent across schools.

Model 2 (Latent Variable Regressions in Within-School and Between-School Models)

In the following model (Model 2), we specify latent variable regressions in the between-school (level-3) model. By doing this, we can estimate latent variable regression coefficients instead of estimating correlation coefficients as in Model 1. While the correlation coefficient provides only information on the strength of association, regression coefficient provides information on the expected amount of increase or decrease in an outcome of interest in a model when a predictor variable increases by one unit, holding constant the other variables in the model. The within-individual (level-1) and within-school (level-2) models are the same as in Model 1 (see Equations 7 and 8). In the following between-school (level-3) model, school mean rate of change parameters are modeled as a function of school mean initial status parameters. Furthermore, the within-school initial status/rate of change slopes are modeled as a function of school mean initial status parameters as well.



Sch#	Reduct.%
#45	6.58%
#44	53.90%
#43	48.18%
#42	46.13%
#41	20.91%
#40	16.86%
#39	21.97%
#38	23.92%
#37	2.22%
#36	23.72%
#35	4.64%
#34	85.96%
#33	92.89%
#32	19.58%
#31	0.00%
#30	54.16%
#29	60.71%
#28	0.00%
#27	12.17%
#26	5.71%
#25	0.00%
#24	0.52%
#23	6.61%
#22	0.00%
#21	20.48%
#20	67.32%
#19	86.10%
#18	0.00%
#17	1.88%
#16	16.16%
#15	12.86%
#14	0.00%
#13	8.04%
#12	20.70%
#11	94.12%
#10	11.86%
#9	85.33%
#8	15.23%
#7	5.99%
#6	38.78%
#5	29.13%
#4	35.40%
#3	3.29%
#2	15.25%
#1	15.53%

Figure 2. Percentage of reduction in growth rate variance after initial status is included as a predictor

* The white part of each bar represents amount of variance accounted for by initial status. The shaded part of each bar represents amount of residual variance that is not accounted for by initial status.

We pose the following between-school model:

$$\beta_{00j} = \gamma_{000} + u_{00j} \qquad u_{00j} \sim N\,(0, \tau_{\beta 00})$$
$$\beta_{10j} = \gamma_{100} + \mathrm{Bb}^*(\beta_{00j} - \gamma_{000}) + u_{10j} \qquad u_{10j} \sim N\,(0, \tau_{\beta 10})$$
$$\mathrm{Bw}_j = \mathrm{Bw}_0 + \mathrm{Bw}_1{}^*(\beta_{00j} - \gamma_{000}) + u_{\mathrm{Bw}j} \qquad u_{\mathrm{Bw}j} \sim N\,(0, \tau_{\mathrm{Bw}}) \tag{10}$$
$$Cov\,(u_{00j}, u_{10j}) = 0,\ Cov\,(u_{00j}, u_{\mathrm{Bw}j}) = 0,\ Cov\,(u_{10j}, u_{\mathrm{Bw}j}) = \tau_{\beta 10,\mathrm{Bw}},$$

where Bb is a latent variable regression coefficient that relates school mean initial status to school mean rate of change. We term this coefficient the between-school mean initial status / mean rate of change slope (Bb), in contrast to the within-school initial status / rate of change slope (Bw_j). This coefficient represents the expected increase or decrease in school mean rate of change when school mean initial status increases by one unit. Likewise, Bw_1 is also a latent variable regression coefficient capturing the relationship between within-school initial status / rate of change slopes and school mean initial status. This latent variable regression coefficient is termed the school mean initial status / within-school relationship slope. In addition, it is interpreted as the expected amount of increase or decrease in the within-school initial status / rate of change slopes when school mean initial status increases by one unit.

Turning to the fixed effects and variance-covariance components, γ_{000} retains the same meaning as in Model 1 (i.e., grand mean initial status). By virtue of centering school mean initial status around the grand mean initial status, γ_{100} represents the expected rate of change when school mean initial status (β_{00j}) is equal to the grand mean (γ_{000}). The random effect, u_{10j}, represents the residual for school *j*, after taking into account school mean initial status, and this random effect is assumed normally distributed with mean 0 and variance $\tau_{\beta 10}$. In addition, because of the centering of school mean initial status around the grand mean, Bw_0 represents the within-school initial status / rate of change slope for schools where school mean initial status is equal to the grand mean. The residual of Bw_j (i.e., $u_{\mathrm{Bw}j}$,) in the third equation above is assumed normally distributed with mean 0 and variance τ_{Bw}.

The covariance between u_{00j} and u_{10j} is set equal to 0, since school mean rate of change is modeled as a function of school mean initial status. For the same reason,

$Cov(u_{00j}, u_{Bwj})$ is also assumed to be 0. However, the covariance between school mean rate and the within-school initial status / rate of change slopes is $\tau_{\beta 10,Bw}$.

The results for Model 2 are presented in Table 2. The posterior mean for the between-school mean initial status / school mean rate of change slope (Bb) is equal to 0.076, and its 95 % interval ranges from -0.001 to 0.151. Though the 95 % interval includes a value of 0, the proportion of posterior distribution below 0 is only approximately 2.7%. The results suggest that schools with high mean initial status tend to have steeper growth trajectories than schools with low mean initial status. Based on these results, we expect an increase in school mean rate of change of about .076 points when school mean initial status increase one unit.

Table 2

Model 2: 3-Level unconditional LVR HM: Estimating Within-School Initial Status / Rate of Change Slopes (Bw_j) and Between-School Mean Initial Status / School Mean Rate of Change Slope (Bb)

	Estimate	SE	95% Interval	Median	Prop.>0
Fixed effects:					
Model for School Mean Initial Status (β_{00j}):					
Mean Init. Status (γ_{000})	49.72	0.63	(48.48, 50.96)	49.72	1.0000
Model for School Mean Rate of Change (β_{10j}):					
Mean Rate of Change (γ_{100})	3.86	0.15	(3.57, 4.14)	3.86	1.0000
School Mean Init. Status (Bb)	0.076	0.039	(-0.001, 0.151)	0.076	.9734
Model for Init. Status / Rate of Change Slope (Bw_j):					
Mean Init. Status / Rate of Change Slope (Bw_0)	0.089	0.012	(0.067, 0.112)	0.089	1.0000
School Mean Init. Status (Bw_1)	-.005	0.003	(-0.011, 0.001)	-0.005	.0601
Variance Components:					
Level-1 Error(σ^2)	16.36	0.36	(15.67, 17.07)	16.36	
Level-3 Variance:					
Initial Status($\tau_{\beta 00}$)	16.22	3.782	(10.310, 25.010)	15.710	
Rate of Change($\tau_{\beta 10}$)	0.656	0.189	(0.364, 1.100)	0.630	.0314
Init. / Rate of Change Slope (τ_{BW})	0.002	0.001	(0.001, 0.004)	0.002	
Cov. (Sch. Rate of Change, Init. / Rate of Change Slope)	-.016	0.010	(-0.037, 0.001)	-0.015	
Correlations:					
Sch. Rate of Change , Init / Rate of Change Slopes ($Corr_{\beta 10j, Bwj}$)	-0.56	0.16	(-0.81, -0.20)	-0.58	. 0029

We often have a negative estimate of the latent variable coefficient (Bw_1) relating differences in school mean initial status to the within-school initial status/rate of change slopes. The resulting posterior mean (Bw_1) is -0.005 and approximately 6% of mass of the posterior distribution is above the value of 0 ($p(Bw_1 > 0 \mid y) = .060$). This result shows that as school mean initial status increases one unit, the within-school initial status/rate of change slope decreases approximately 0.005. Thus, it is very likely that there is a weaker relationship between initial status and rates of change in higher mean initial status schools than in lower mean initial status schools.

We now consider the meaning of these results from a practical standpoint and what they might disclose about the process of student growth and the distribution of students' achievement within schools. As an illustration, Figure 3 displays the expected growth trajectories for three different schools: a school where mean initial status is two standard deviations (8 points) above the grand mean, a school where mean initial status is equal to the grand mean, and a school where mean initial status is two standard deviations (8 points) below the grand mean. Each school has three expected growth trajectories for students whose initial status is equal to 15 points above the school mean (circle lines), at the school mean (box lines), and 15 points below the school mean (triangle lines). Thus, in all three schools, initial gaps among the students are equal to 30 points as presented in the figure. Note that all of the expected growth trajectories presented in Figure 3 are based on the fitted Model 2.

In this Figure 3, we can clearly see two growth patterns in terms of overall growth rate and the underlying distribution of achievement. First, with respect to overall growth rates, as school mean initial status increases, the overall school rate of change gets faster. The school starting off at 8 points above the grand mean has the fastest average rate, the school at the grand mean has the second fastest rate, and the school 8 points below the grand mean has the slowest rate among the three schools. These mean trajectories, which are labeled with boxes, show that students in these three schools start off on average with values of 57.2, 49.7, and 41.72, and they end up with values of 71.1, 61.2, and 51.5, respectively. Thus, the expected gains are approximately 13.4, 11.6, and 9.8 points, respectively.

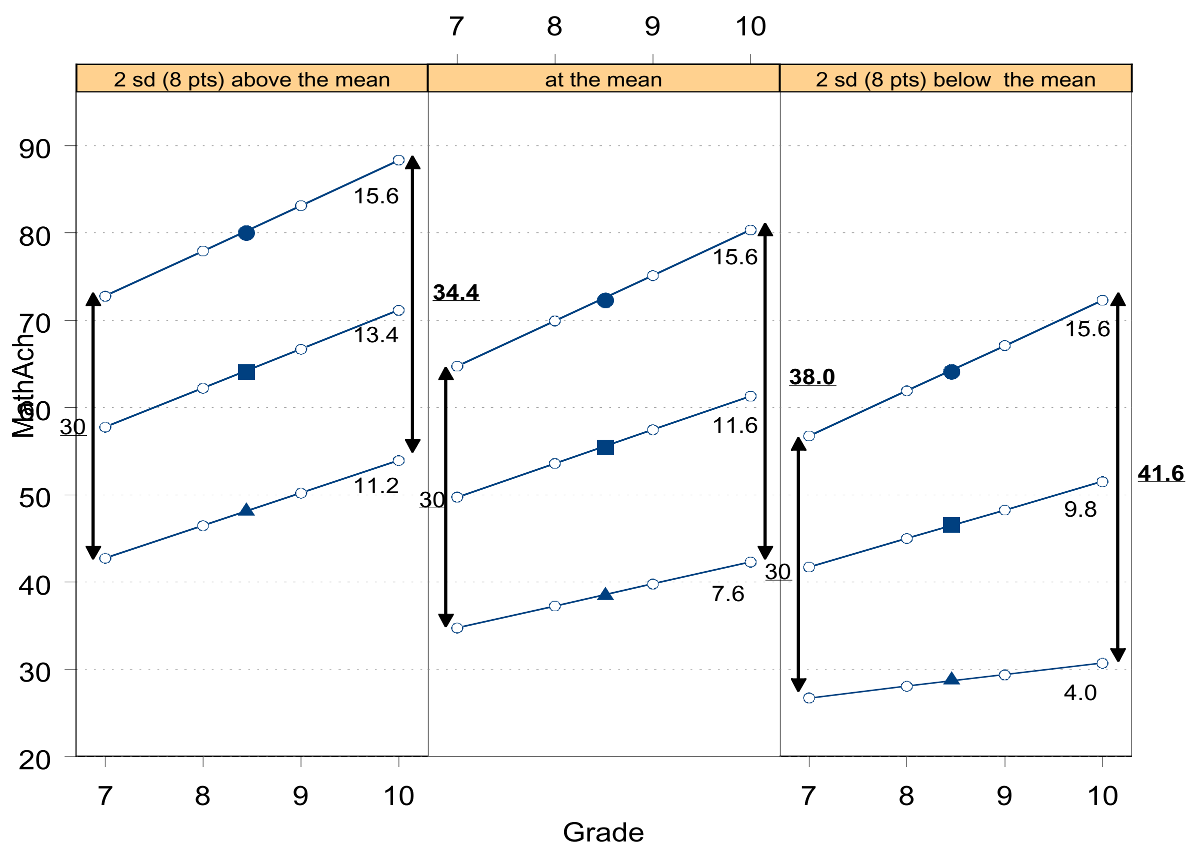


Figure 3. Expected growth trajectories for three different schools based on the results from Model 3. Circle line: expected rate for students 15 points above the school mean initial status value. Box line: expected rate for students with initial status values equal to the school mean. Triangle line: expected rate for students 15 points below the school mean initial status value.

* The numbers attached to each line denote the differences between expected scores at grade 10 and initial status values. In other words, these are gain scores based on the fitted model.

** The boldfaced number denotes the expected difference at grade 10 between a student who starts 15 points above the school mean initial status value and a student who starts 15 points below the school mean initial status value.

Second, we now examine how much initial gaps among students within each of the three schools increase by grade 10. Given that the within-school initial status/rate of change slopes for these three schools are all positive, we can expect that initial gaps at grade 7 get wider over time. However, in the school whose mean initial status is 8 points above the grand mean, the initial gap (i.e., 30 points) becomes 34.4 points by grade 10; the gap increases by only 4.4 points. This indicates that students in this school tend to progress at fairly similar rates regardless of initial status. Comparing this school to the school starting off 8 points below the grand mean, we can see that the initial gap (30 points) gets magnified to 41.6 points by grade 10. This large difference at grade 10 is due to the fact that a student who starts off low initially grows at a slow rate, gaining only 4.0 points by grade 10, while a student who starts off high grows at a considerably faster rate, gaining 15.6 points by grade 10.

Third, Figure 3 clearly illustrates how consequential school membership can be with respect to a student's initial status and his or her subsequent rate of progress. Among the students who start off 15 points above their school's mean initial status value (see the lines labeled with circles), the gains from grade 7 to grade 10 are almost identical (15.6 points) in each of the three schools. In contrast, students who start off with 15 points below their school's mean initial status value (triangle lines) progress very differently from each other. The student in the first panel gains 11.2 points, while the student in the third panel gains only 4.0 points.

In summary, two important things that come to light regarding school mean growth rates and the distribution of student achievement within schools are: a) students in schools with high mean initial status tend to grow faster on average than students in schools with low mean initial status, and b) students in schools in which mean initial status is high not only progress at faster rates, but growth is also more equitable as well. In other words, the expected gaps at grade 10 among students in these schools do not get magnified as much as the expected gaps among students in schools with low mean initial status values.

Model 3 (Latent Variable Regressions in Within-School and Between-School Models Including Student Characteristics and School Characteristics)

We now pose a Model 3 in which we include student-level predictors and school-level predictors in the model. Specifically, we specify a within-school (level-2) model in which two student-background characteristics are included as predictors both for initial status and rates of change. And we specify a between-school model to explore how differences in particular school-level factors are related to where students in a school start on average (i.e., school mean initial status), how fast students progress on average (i.e., school mean rates of change), and how student growth is distributed within schools (i.e., within-school initial status / rate of change slopes). For Model 3, we have the same within-individual (level-1) model as in Models 1 and 2. The within-school model is as follows:

$$\pi_{0ij} = \beta_{00j} + \beta_{01j}(SEDEX_{ij} - SEDEX_{..}) + \beta_{02j}(BPBLM - BPBLM_{..}) + r_{0ij}$$

$$r_{0ij} \sim N\ (0, \tau_{\pi 0j})$$

$$\pi_{1ij} = \beta_{10j} + \mathrm{Bw}_j(\pi_{0ij} - \beta_{00j}) + \beta_{11j}(SEDEX_{ij} - SEDEX_{..})$$

$$+ \beta_{12j}(BPBLM - BPBLM_{..}) + r_{1ij} \qquad r_{1ij} \sim N\ (0, \tau_{\pi 1j}) \qquad (11)$$

$$Cov\ (r_{0ij}, r_{1ij}) = 0.$$

In the above model, two student-level variables are included (see Endnote 2 for more details regarding these variables). First, the variable student educational expectations (*SEDEX*) at grade 7 reflects the highest level of education a student believes that he or she will attain. There are six possible responses for *SEDEX*: 1 for a high school degree, 2 for vocational training, 3 for 2-year college, 4 for 4-year college, 5 for a master's degree and 6 for a doctorate or professional degree. Second, the variable student behavioral problems (*BPBLM*) take on a value 1 if a student reported behavioral problems at grade 7 (e.g., if a student had been suspended, had been arrested by the police, or had considered dropping out), and 0 otherwise. Among the 2,628 sampled students, 436 students (16.6%) indicated behavioral problems.

The main reason to include these student characteristics in the model is to adjust for differences among schools in these key student characteristics before examining the relationship between school mean initial status (β_{00j}) and school mean rate of change (β_{10j}). Analogous to ANCOVA, if behavioral problems are negatively related to initial status, and the proportion of students with behavioral problems in school *j* is above the grand mean (*BPBLM*), the expected initial status value for school *j* would be adjusted upwards. By virtue of centering each predictor around its grand mean, β_{00j} represents the adjusted mean initial status for school *j*. β_{01j} is a fixed effect coefficient that relates a student's educational expectations to his or her initial status, holding the other level-2 predictor constant. And β_{02j} is a fixed effect coefficient capturing the difference in initial status in school *j* between students reporting behavioral problems and those who do not, holding the other level-2 predictor constant. Likewise, β_{11j} and β_{12j} are fixed effects coefficients capturing the effects of student educational expectations and student behavioral problems on rates of change, respectively, holding constant a student's initial status. Note that these four student-level fixed effects are treated as not varying across schools at level 3.

The between-school (level-3) model is specified as follows.

$$\beta_{00j} = \gamma_{000} + \gamma_{001}(MHOMERS_j — MHOMERS.) + u_{00j} \quad u_{00j} \sim N(0,\tau_{\beta 00})$$

$$\beta_{01j} = \gamma_{010}$$

$$\beta_{02j} = \gamma_{020}$$

$$\beta_{10j} = \gamma_{100} + Bb^*(\beta_{00j} — \gamma_{000}) + \gamma_{101}(MHOMERS_j — MHOMERS.) +$$
$$\gamma_{102}(MTEACARE_j — MTEACARE.) + u_{10j} \qquad u_{10j} \sim N(0, \tau_{\beta 10})$$

$$\beta_{11j} = \gamma_{110}$$

$$\beta_{12j} = \gamma_{120}$$

$$\mathrm{Bw}_j = \mathrm{Bw}_0 + \mathrm{Bw}_1(\beta_{00j} — \gamma_{000}) + \mathrm{Bw}_2(MHOMERS_j — MHOMERS.) +$$
$$\mathrm{Bw}_3(MCOLLEAGE_j — MCOLLEAGE.) +$$
$$\mathrm{Bw}_4(MCOMHW_j — MCOMHW.) +$$
$$\mathrm{Bw}_5(MPERSUCC_j — MPERSUCC.) + u_{\mathrm{Bw}j} \qquad u_{\mathrm{Bw}j} \sim N(0, \tau_{\mathrm{Bw}}) \quad (12)$$

$$Cov(u_{00j}, u_{10j}) = 0, Cov(u_{00j}, u_{\mathrm{Bw}j}) = 0, Cov(u_{10j}, u_{\mathrm{Bw}j}) = \tau_{\beta 10,\mathrm{Bw}}$$

First, school mean initial status is modeled as a function of only one predictor variable (i.e., school mean home resources [*MHOMERS*]). This variable is created by aggregating student home resources, which is the sum of students responses to the following questions: whether a student has his or her own place to do homework, owns a computer, has his or her own room, and has more than 50 books in his or her home. Thus, the range of this variable is from 0 to 4.

Second, school mean rate of change is regressed on school mean initial status, school mean home resources, and school mean teacher care (*MTEACARE*). The teacher care variable is based on teachers' responses to the following questions: "I sometimes feel it is a waste of time to try to do my best as a teacher" and "The teachers in this school push the students pretty hard in their academic subject." Each question is measured on a 5-point Likert scale ranging from 1 *(Strongly Disagree)* to 6

(Strongly Agree). Note that the teacher care variable was measured based on the responses of the science and math teachers in the sampled schools to a questionnaire administered in spring 1988, when students in the sample were in grade 7. The number of teachers who completed this questionnaire is 349, and the number of teachers per school ranges from 3 to 15. The average number of teachers per school in the sample is approximately 7.5 and the median is 7. The sample mean of this teacher care variable is equal to 4.41, and its standard deviation is equal to .55.

Finally, within-school initial status/rate of change slopes are modeled as a function of school mean home resources, the school mean percentage of students completing homework on time (*MCOMHW*), the school mean percentage of students who their teachers expect to graduate from college (*MCOLLEAGE*), and school mean teachers' self-evaluation of their success in educating students (*MPERSUCC*). *MCOMHW*, *MCOLLEAGE*, and *MPERSUCC* are measured based on teachers' answers to questions on a class-specific questionnaire that was administered to the math and science teachers who had one or more of the LSAY students in their classes. In particular, *MCOLLEAGE* reflects teacher expectations of the students in their classes and not of particular LSAY students. The sample mean of this variable is equal to 78.7, and its standard deviation is equal to 10.82. Teachers' self-evaluations of their success are based on their responses to the following question: "To what extent do you feel successful in providing the kind of education you would like to provide for the students in your class?" The measurement scale for this variable ranges from 1 *Not very successful* to 4 *Very successful*. The sample mean and its standard deviation are equal to 2.89 and .43, respectively.

We present the results for this model in Table 3. The grand mean for initial status is close to 50 points. School mean home resources (γ_{001}) shows a significant positive effect on school mean initial status. The estimate for γ_{001} is equal to 7.14 and its 95% interval contains only positive values ranging from 4.20 to 10.07. Thus we expect approximately a 7-point increase in school mean initial status given a one unit increase in school mean home resources.

We now consider results for school mean rates of change. The expected growth rate is approximately 3.9. This means that we can expect on average 3.9 points improvement in math achievement at each grade. School mean home resources (γ_{101}) is positively related to school mean rate of change, even after holding school mean initial status and the teacher care variable constant. The estimate for school mean home resources (γ_{101}) is equal to 1.182, and $p(\gamma_{101} > 0 \mid y) = .9883$, which indicates that

Table 3

Model 3: 3-Level LVR HM (Latent Variable Regressions in Within-School and Between-School Models Including Student Characteristics and School Characteristics)

	Estimate	SE	95% Interval	Median	Prop. >0
Fixed effects:					
Model for School Mean Initial Status (β_{00j}):					
Mean Init. Status (γ_{000})	49.76	0.44	(48.90, 50.63)	49.76	1.0000
Home Resources (γ_{001})	7.14	1.50	(4.20, 10.07)	7.14	1.0000
Model for School Mean Rate of Change (β_{10j}):					
Mean Rate of Change (γ_{100})	3.861	0.113	(3.638, 4.084)	3.861	1.0000
School Mean Init. Status(Bb)	-.009	0.049	(-0.106, 0.088)	-0.009	.3972
Home Resources (γ_{101})	1.182	0.516	(0.173, 2.202)	1.180	.9883
Teacher's Care (γ_{102})	0.679	0.202	(0.282, 1.075)	0.679	.9994
Model for Within-School Init. Status / Rate of Change Slope (Bw_j):					
Mean of Init. Status / Rate of Change Slope (Bw_0)	0.076	0.010	(0.056, 0.097)	0.076	1.0000
School Mean Init. Status (Bw_1)	0.002	0.004	(-0.007, 0.010)	0.001	.6348
Home Resources (Bw_2)	-.032	0.050	(-0.131, 0.067)	-0.032	.2643
% Grad. College (Bw_3)	-.001	0.001	(-0.002, 0.000)	-0.001	.0623
% Complt. Homework (Bw_4)	-.002	0.001	(-0.005,-0.000)	-0.002	.0239
Percpt of Succ. (Bw_5)	0.055	0.028	(-0.001, 0.108)	0.055	.9726
Effects of Student-Level Characteristics on Init. Status (π_{0ij}):					
SEDEX (γ_{010})	1.892	0.123	(1.649, 2.134)	1.892	1.0000
BPBLM (0 / 1) (γ_{020})	-3.55	0.465	(-4.450,-2.635)	-3.546	.0000
Effects of Student-Level Characteristics on Rate of Change (π_{1ij}):					
SEDEX (γ_{110})	0.202	0.053	(0.097, 0.305)	0.202	.9999
BPBLM (0 / 1) (γ_{120})	-.256	0.196	(-0.641, 0.128)	-0.256	.0954
Variance Components:					
Level-1 Error(σ^2)	16.49	0.36	(15.80, 17.21)	16.48	
Level-3 Variance:					
Initial Status($\tau_{\beta 00}$)	7.325	1.863	(4.443,11.710)	7.072	
Rate of Change($\tau_{\beta 10}$)	0.359	0.119	(0.181, 0.643)	0.341	
Init. / Rate of Change Slopes (τ_{BW})	0.001	0.000	(0.000, 0.002)	0.001	
Cov. (Sch. Rate of Change, Init. / Rate of Change Slopes)	-.004	0.005	(-0.016, 0.005)	-0.004	
Correlations:					
Sch. Rate of Change, Init / Rate of Change Slopes ($Corr_{\beta 10j,\, Bwj}$)	-.258	0.301	(-0.745, 0.383)	-0.289	.2060

when school mean resources increases one unit we expect a 1.182 increase in school mean rate of change, holding constant the teacher care variable and school mean initial status as well. The teacher care variable also shows a significant positive effect on school mean rate of change, after holding constant school mean home resources and school mean initial status. The estimate (γ_{102} = .679) and the corresponding 95% interval (.282, 1.075) suggest that students who attend schools in which teachers try to do their best as a teacher and push their students fairly hard in their academic subjects grow at faster rates on average. Note that school mean initial status (Bb) is not significantly related to school mean rate of change, holding constant school mean home resources and the teacher care variable.

Regarding within-school initial status/rate of change slopes (Bw_j), two predictors have significant negative estimates, while one variable has a positive estimate. Holding constant the other variables in the model, the school mean percentage of students completing homework on time has a significant negative effect on Bw_j. The coefficient (Bw_4) takes on a value of -.002, which suggests that a 10% increase in the percentage of students completing homework on time is associated with a .02 decrease in Bw_j. Note that the corresponding 95% interval includes only negative values and $p(Bw_4 > 0 \mid y)$ is equal to .0239. Based on this result, schools in which the percentages of students completing their homework on time are high tend to have more equitable growth patterns in the sense that those schools have smaller Bw_j values. Thus for students in such schools, differences in math achievement among students at grade 7 are not magnified over time as much as in schools in which the percentages of students completing homework are low. The percentage of students who teachers expect will graduate from college with a baccalaureate, which is a proxy measure of *teacher expectations* of students, is also negatively related to Bw_j, holding constant all the variables in the model. However, the interval for the estimate contains 0 and more than 5% of the posterior distribution is greater than 0 (95% interval = (-.002, .000), $p(Bw_3 > 0)$ = .062).

In contrast, teachers' perceptions of how successful they are in educating students has a significant positive effect on Bw_j, holding constant school mean initial status, *MCOLLEAGE*, and *MCOMHW* in the model. The estimate (Bw_5) is equal to .055, and the proportion of posterior distribution greater than 0 is .9726 (its 95% interval = [-0.001, 0.108]). The positive coefficient indicates the more teachers in schools feel themselves to be successful in providing the kind of education they would like to provide, the larger the expected value of Bw_j.

One possibility or speculation is that the kind of education many of the teachers would like to provide is the kind of instruction that is challenging and that requires students to draw on a variety of skills and to use their creativity. One can imagine that in classes where this kind of instruction is occurring, there may be a tendency for those students with relatively high initial status to benefit a great deal. However, in classes where the emphasis is on making sure that the students are by and large prepared for district-wide exams—classrooms in which teachers may not be providing the kind of challenging and stimulating instruction they might like to provide—it seems that this might work to weaken somewhat the relationship between initial status and rate of change. Note that *MCOMHW*, *MCOLLEAGE* and *MPERSUCC* are positively correlated. Thus high values on *MCOMHW* and *MCOLLEAGE* appear to have the effect of reducing the magnitude of the initial status/rate of change slope in a school, whereas *MPERSUCC* would appear to work in the direction of increasing it somewhat.

We now consider the amount of variance accounted for by school-level predictors. The remaining variance in school mean initial status ($\tau_{\beta 11}$) is equal to 7.325, which indicates that 56.2% of variance in school mean initial status is accounted for by using only school mean home resources as a predictor. Furthermore, 44.5% of the variance in school mean rate of change is explained by using school mean initial status, school mean home resources, and the school mean teacher care variable as predictors. Finally, the half of the variance of the within-school initial status and rate of change slopes is accounted for by the five predictors in the model.

Sensitivity Analyses

Note that different sets of predictors are employed in the three equations in the level-3 model in Equation 12. Drawing from Raudenbush and Bryk (2001, p. 271), in such situations, it is possible that misspecifications in one of the three equations can impact the estimates of the fixed effects in another (Raudenbush & Bryk). To help investigate this issue, we re-fit Model 3 with all level-3 covariance terms set to a value of 0. All of the resulting estimates and intervals were nearly identical to those reported in Table 3.

In hierarchical modeling settings, normality assumptions are commonly employed at each level. In fitting models under normality assumptions, parameter estimates are potentially sensitive to outlying cases. Specifically, in the LVR-HM3, an unusually high or low time series observation for an individual given the overall trend of that person's data (i.e., level-1 outliers) can impact the estimation of initial status and rate of change for that person. A potential problem connected with level-1 outliers is that level-1 outliers can impact summaries of the data for a given cluster, which in turn, can impact the estimation of the coefficients of level-2 predictors (Rachman-Moore & Wolfe, 1984; Seltzer & Choi, 2002). For example, individuals with unusually high observations at the first time point but unusually slow rates of change, or individuals with unusually low observations at the first time point but unusually rapid rates of change (i.e., level-2 outliers), could strongly influence the estimation of within-school initial status / rate of change slopes. Similarly, level-3 outliers—schools with unusually high or low mean initial status values, or unusually slow or rapid mean rates of change—especially in small sample settings, can impact the estimation of level-3 latent variable regression coefficients (e.g., the between-school mean initial status / mean rate of change slope (Bb) and the school mean initial status / within-school relationship slope (Bw_1) and the coefficients of observed level-3 predictors.

Thus we examine how the results based on Model 3 may change when we re-fit Model 3 under *t* distributional assumptions at levels 1, 2, and 3 with the degrees of freedom parameters of the *t* distributions set to a value of 4. Fitting HMs under *t* distributional assumptions can be readily carried out in WinBUGS (see Seltzer & Choi, 2002). Re-fitting HMs under heavy tailed distributional assumptions has the effect of downweighting possible outliers (Seltzer, 1993; Seltzer et al., 2002; Seltzer & Choi).

The results show that all the point estimates and intervals based on normality assumptions are extremely close to the corresponding estimates and intervals based on *t* distributional assumptions. The largest change occurs for the effect of student behavioral problems on rate of change (γ_{120}). The point estimate changes from -.256 to -.324. However, the resulting interval based on *t* distributional assumptions still contains a value of 0 (-.719, .075). In conclusion, the sensitivity analyses conducted under *t* distributional assumptions do not lead to changes in conclusions based on Model 3.

Improving the Performance of the Gibbs Sampler: Comparing Among Different Formulations and Parameterizations of the LVR-HM3

In implementing the Gibbs sampler in HM settings, we need to be alert to situations in which mixing is poor, i.e., situations where successive values in the chains generated for one or more parameters in the model are highly autocorrelated. When mixing is extremely poor, it can be difficult to assess convergence of the Gibbs sampler. Even if one is reasonably confident that the sampler has converged, it is difficult to know whether all regions of the joint posterior have been adequately traversed.

Poor mixing can stem from high correlations among particular parameters in the joint posterior (see, e.g., Gilks et al. [1996, p. 91]). One strategy to reduce high posterior correlations and improve mixing is to center the covariates in one's model. Thus in the case of our models, one option would be to center student initial status around its school mean (i.e., $\pi_{0ij} - \beta_{00j}$) as in Equation 11, and to center school mean initial status around the grand mean (i.e., $\beta_{00j} - \gamma_{000}$) as in Equation 12.

In addition, another important option is to explore differences in mixing when we implement our models using a mixed model formulation versus a hierarchical model formulation. In the mixed model formulation, we write the LVR-HM3 as a single equation. In other words, Equations 7, 11, and 12 are collapsed into one level by substituting the level-3 equation (Equation 12) into the level-2 equation (Equation 11), and in turn, substituting the resulting equation into the level-1 equation (Equation 7). The parameters of the mixed model consist of the fixed effects (e.g., γ_{000}, γ_{100}, Bb, Bw_0, Bw_1, etc.), random effects (e.g., r_{0ij}, r_{1ij} ; u_{00j}, u_{10j}, $u_{\text{BW}j}$) and the variance components in the LVR-HM3. In contrast, in the hierarchical formulation, parameters in lower levels (e.g., π_{0ij}, π_{1ij}; β_{00j}, β_{10j}, Bw_j) are viewed as arising from distributions of parameters specified at higher levels. Note that all results that we presented above are based on the hierarchical model formulation. Even though the two formulations are mathematically equivalent, the joint posterior densities are different. Notably, random effects appear in the joint posterior in the mixed model formulation (e.g., r_{0ij}, r_{1ij} ; u_{00j}, u_{10j}, $u_{\text{BW}j}$), while random intercepts and coefficients (e.g., π_{0ij}, π_{1ij}; β_{00j}, β_{10j}, Bw_j) appear in the joint posterior in the hierarchical formulation. In addition, while the random effects are modeled as a function of means of 0, parameters such as π_{0ij}, π_{1ij}, for example, are modeled as a function of expected values based on equations in the next level of the hierarchy.

In this paper, we are especially interested in how we can improve the performance of the Gibbs sampler in estimating models that involve latent variable regressions in three-level hierarchical models. We explored the following implementations of the Gibbs sampler for Model 3.

1. Hierarchical formulation with no centering involved in the latent variable regressions either at level 2 or level 3.
2. Hierarchical formulation with β_{00j} centered around γ_{000} (level-3 centering; uncentered at level-2).
3. Hierarchical formulation with π_{0ij} centered around β_{00j} (level-2 centering; uncentered at level-3).
4. Hierarchical formulation employing both level-2 and level-3 centering (i.e., π_{0ij} centered around β_{00j}; β_{00j} centered around γ_{000}).
5. Mixed formulation algebraically equivalent to implementation 4.

The performance of the Gibbs sampler for each of the above implementations is monitored by means of computing autocorrelations among the deviates in a chain. Plots of a series of autocorrelations (i.e., autocorrelation functions (ACF) plots) are important tools for describing the serial (or temporal) dependence structure. The ACF at lag k (ρ(k)) is estimated by dividing the covariance between X_t and X_{t+k} by the variance of X_t. Thus, the ACF at lag k is as follows: $\hat{\rho}(k) = \dfrac{\sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}$,

where $\bar{x} = \dfrac{\sum_{t=1}^{n} x_t}{n}$ (Smith, 2000). When there is little serial dependency, the ACF will drop rapidly to values close to 0 as k increases.

Table 4 summarizes estimates of ACF at a series of different lag k values. As can be seen, we categorize parameters in each implementation depending upon whether their ACF values are smaller than .10 at a series of different lag values. Note that the autocorrelation functions (ACF) in Table 4 are constructed based on 4,000 deviates for each parameter generated over iterations 2,001 to 6,000 of the Gibbs sampler.

Table 4

Comparison of Performance of Gibbs Sampler Among Different Formulations and Parameterizations of the LVR-HM3: Parameters With Autocorrelation Function (AFC) Values Below .10 at lag = k ($\rho(k) < .10$)

	Hierarchical model formulation				Mixed model formulation
	No centering	UnCentered at lev-2	UnCentered at lev-3	Centered at lev-2 and 3	
$\rho(k) < .10$ k = 5	γ_{000}, γ_{001}, γ_{010}, γ_{020}, σ^2, $\tau_{\beta 00}$	γ_{001}, γ_{010}, γ_{020}, σ^2, $\tau_{\beta 00}$	γ_{000}, γ_{001}, γ_{010}, γ_{020}, γ_{102}, σ^2, $\tau_{\beta 00}$	γ_{000}, γ_{001}, γ_{010}, γ_{020}, $\tau_{\beta 00}$, γ_{100}, γ_{102}, σ^2	
$\rho(k) < .10$ k = 10	Bw_5, γ_{102}	γ_{000}, Bw_3, Bw_4, Bw_5	Bw_3, Bw_4, Bw_5	$\tau_{\beta 10}$	σ^2
$\rho(k) < .10$ k = 50	γ_{110}, γ_{120}, τ_{Bw}	γ_{110}, γ_{120},	γ_{110}, γ_{120}	Bb, Bw_0, Bw_1, Bw_2, Bw_3, Bw_4,,, γ_{101}, γ_{110}, γ_{120}, τ_{Bw}, Bw_5	γ_{020}, γ_{110}, γ_{120}, γ_{010}, $\tau_{\beta 00}$, $\tau_{\beta 10}$, τ_{Bw}, Bw_5
$\rho(k) < .10$ k = 100	Bw_2, Bw_3, Bw_4, γ_{101}, $\tau_{\beta 10}$	Bb, Bw_0, Bw_1, Bw_2, γ_{100}, γ_{101}, γ_{102}, $\tau_{\beta 10}$, τ_{Bw}	γ_{101}, $\tau_{\beta 10}$, τ_{Bw}		Bw_0, γ_{100}, Bw_3, Bw_4, γ_{102}
$\rho(k) < .10$ k = 150			Bw_2		Bw_1, Bw_2
$\rho(k) < .10$ k = 200					
$\rho(k) < .10$ k = 250					
$\rho(k) < .10$ k = 300	Bb, Bw_0, Bw_1, γ_{100}		Bb, Bw_0, Bw_1, γ_{100}		Bb, γ_{000}, γ_{001}, γ_{101}

First, for implementation 1 (i.e., no centering either at levels 2 or 3) ACF values for γ_{000}, γ_{001}, γ_{010}, γ_{020}, σ^2, and $\tau_{\beta 00}$ are smaller than .10 at lag = 5, while Bb, Bw_0, Bw_1, and γ_{100} show very high ACF values even at lag = 300. The ACF values at lag = 300 for these four parameters are .54, .59, .62, and .56, respectively. We present the ACF plot for γ_{000} in Figure 4, which is similar to the plots for γ_{001}, γ_{010}, γ_{020}, σ^2, and $\tau_{\beta 00}$. As can be seen, ACF values are decreasing very quickly and are close to 0 before lag = 5. In contrast, Figure 5 shows the ACF plot for γ_{100}. As can be seen, the ACF values hardly decrease as the lags increase. The ACF plots for Bb, Bw_0, and Bw_1 are very similar to Figure 5.

ACF values for all parameters in the model are smaller than .10 at lag = 100 for the case of no centering at level 2 and centering at level 3 (implementation 2). However, fixed effects in the level-3 equation for Bw_j (i.e., Bw_0, Bw_1, Bw_2, Bw_3) and fixed effects in the level-3 equation for school mean rates of change (i.e., γ_{100}, γ_{101}, γ_{102})

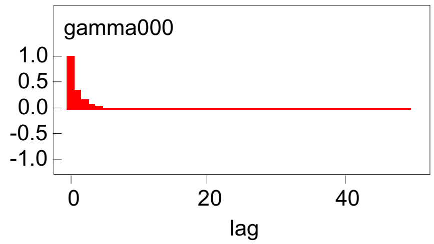


Figure 4. ACF 1: An example of good mixing (low ACF values).

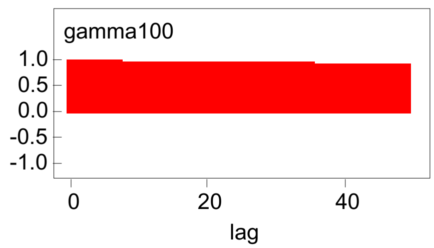


Figure 5. ACF 2: An example of bad mixing (high ACF values).

show relatively higher ACF values than the other parameters in the model. Third, we can see very different ACF values for implementation three in which we employ centering at level 2 but not at level 3. One group of parameters appears to have very good mixing in the sense that the ACF values dip below a value of .10 by lag 50. In contrast, the performance of the Gibbs sampler is very poor for the level-3 fixed effects, Bb, Bw_0, Bw_1, and γ_{100}.

Specifically, the ACF values at lag = 300 for those four parameters are .55, .59, .62, and .56, respectively. Thus, it is very obvious that Bb, Bw_0, Bw_1, and γ_{100} have very high autocorrelations. Fourth, the degree of autocorrelation tends to be far smaller based on implementation 4. Specifically, autocorrelations for all parameters in the model under implementation 4 dip below .1 before lag 50. Finally, we can see high ACF values of the parameters (i.e., Bw_1, Bw_2, Bb, γ_{000}, γ_{001}, and γ_{101} in the mixed model formulation with centering at levels 2 and 3).

Based on these results, we can conclude that the hierarchical formulation with centering at both levels 2 and 3 performs very well in terms of mixing, and that the mixing is far superior to the other implementations. Moreover, centering at level 3 is more crucial than centering at level 2 in terms of obtaining better mixing. We have found this to be the case for other LVR-HM3s in our analyses of the LSAY data.

Discussion

In this paper, we presented three-level hierarchical models that contain latent variable regressions in levels 2 and 3. Specifically, student initial status was included as a predictor of student growth at level 2. With this latent variable regression at level 2, we have possibly three random variables (i.e., latent variables) as outcomes at level 3: school mean initial status, school mean rates of change, and within-school initial status / rate of change slopes. All three of these outcomes can be modeled as a function of various school policy and practice variables. Furthermore, various useful latent variable regressions are possible at level 3. In a between-school (level-3) model, school mean rates of change can be modeled as a function of school mean initial status. In addition, we can model within-school initial status / rate of change slopes as a function of school mean initial status. We illustrated key ideas and various distinctive features of the LVR-HM3 by fitting a series of LVR-HM3s to the data from LSAY using MCMC methods.

LVR-HM3s help to broaden the kinds of questions we can address in longitudinal studies of school effectiveness / school accountability, in areas of school indicators, and in longitudinal multi-site intervention studies. As illustrated, in longitudinal studies of school effectiveness, in addition to focusing on differences among schools in their mean growth rates, we can examine how differences in the strength and direction of the relationship between initial status and rates of change relate to differences in various school characteristics, policies and practices. As such, employing LVR-HM3 enables us to explore how equitably student achievement is distributed within school and explain why the distribution of growth in achievement might be more equitable in some schools than others. In addition, by using school mean initial status as a covariate for school mean rates of change, we can examine the unique effects of school-level variables on school mean rates of change after controlling for school mean initial status.

Second, in studies on school indicators, interest may center on comparing schools based on students' achievement scores. However, schools are very different from each other in various aspects that make it difficult to obtain fair comparisons among schools. "Value-added" approaches try to estimate net gain or the amount of growth that can be considered to be solely attributable to a school's teachers, curriculum, policies, and the like. To accomplish this, it is necessary to control or adjust for various student intake factors—student SES, home resources, and other family background factors—and overall school intake characteristics such as school mean SES, available school resources, facilities, and so on. However, it is highly likely that student initial status is associated with rate of progress, and that school mean initial status is associated with school mean rate of change. As illustrated in this paper, the LVR-HM3 can be readily applied to settings in which we wish to adjust for differences in initial status in within-school (level-2) or between-school (level-3) models.

Third, applying LVR-HM3 to longitudinal multi-site intervention studies provides us with important information about the kinds of individuals who might tend to benefit from a treatment. In longitudinal studies of program effectiveness, interest often centers on the difference in growth rates between students in treatment sites and students in comparison sites. However, it is also important to consider whether those students who are most in need of help are those who are benefiting most. To address this question, we can focus on the relationship between initial status and rates of change. In sites where the relationship is positive, this suggests that students with milder difficulties are making more progress. In contrast, a negative relationship suggests that those students with severe initial difficulties are making more progress. Exploring this issue entails regressing rates of change on initial status in each site (level 2). Then in a between-site (level 3) model, we can examine whether site mean growth rates tend to be more rapid in treatment sites after controlling for differences among sites in their mean initial status. Furthermore, we can also explore whether initial status/rate of change coefficients tend to be negative, for example, in treatment sites versus positive in comparison sites. Conducting analyses of this kind requires estimating latent variable regression coefficients at level 2 and regressing those coefficients on site-level variables (e.g., site mean initial status, level of implementation, and so on) at level 3.

A fully Bayesian approach involving the use of MCMC techniques enables us to extend the LVR-HM3 in several useful directions. For example, the LVR-HM3 can be readily extended to settings in which we wish to examine whether the size and direction of expected differences in rates of change between various demographic groups (e.g., gender, race, type of school, etc.) or between individuals in different treatment groups, varies with initial status (Seltzer, Choi, & Thum, 2001a, 2001b). That is, we can investigate interactions between initial status and various categorical predictors on rates of change. In addition, it allows us to investigate further whether the magnitude of these interaction effects varies across schools, and is related to various school characteristics. The latent variable interaction model in the 3-level hierarchical modeling framework (LVI-HM3) can be specified for a study of gender differences in math achievement as follows. For example, the rate of change for student *i* in school *j* can be defined in the following way:

$$\pi_{1ij} = \beta_{10j} + \beta_{10j}Gender_{ij} + b_{1j}(\pi_{0ij} - \beta_{00j}) + b_{2j}[Gender_{ij} \times (\pi_{0ij} - \beta_{00j})] + r_{1ij} \quad (12)$$

Coding gender to a value of 0 for boys and a value of 1 for girls, the latent variable regression coefficient (b_{1j}) associated with the initial status term represents the initial status effect on rates of change for boys, while the other latent variable regression coefficient (b_{2j}) associated with the interaction term represents the differences between boys and girls in the effect of initial status on rates of change. If the interaction effect significantly differs from 0, then we can say that the relationship between initial status and rates of change differs for boys and girls. For example, among students with relatively low initial status, rates of change might be considerably steeper for girls than for boys, while among students with relatively high initial status, rates of change might be substantially more rapid for boys (see this example, Seltzer, Choi, & Thum, 2001a).

Furthermore, we might be interested in investigating why this is so. To what extent are these differences in initial status / rate of change slopes between boys and girls due to differences in school polices and practices such as students' course-taking patterns, math course requirements and the like? Or to what extent are these differences related to the differences in school characteristics such as school climate, academic press, etc.? Attending to these kinds of questions involves specifying latent variable interaction models in three-level hierarchical modeling settings. In other words, the coefficients b_1 and b_2 are treated as outcome variables in a level-3 model and in turn, regressed on school-level variables.

Note that the models and results we have presented are based on the tenability of the assumption that growth in achievement is linear. If growth is curvilinear (e.g., quadratic), one concern is that estimates of initial status under a linear model may differ substantially from estimates based on a quadratic model. Such assumption can be checked via the inspection of plots of student- and school-level growth trajectories, and through comparisons of the fit of linear and quadratic models. Our analyses indicate that the linear growth assumption is tenable for all but a small set of schools in the LSAY sample. For this small set of schools, however, estimates of student initial status based on both a linear model and a quadratic model for growth were extremely similar.

We are currently working on extending the LVR-HM3s presented in this paper to setting where measures of student achievement at only two points in time, along with their standard errors, are available. In addition, we are working on extension to settings in which a quadratic model for student growth is employed at level 1. This would enable us to examine, for example, how differences in initial status among students relate to differences in acceleration or deceleration in achievement.

Endnote

1. Specification of priors for the variance components requires care in situations where little information is available a priori. For example, inverse gamma and inverse Wishart priors with small degrees of freedom and scale parameters S and **S**, respectively, are commonly specified for variance components in hierarchical models. The modes of these priors will depend upon one's choice of S or **S**. In situations where little information concerning the variance components is available a priori, an attempt to choose sensible values for S or **S** can be made. However, it may be found retrospectively that the mode of the prior conflicts substantially with the mode of the likelihood. This can, for example, result in intervals for fixed effects of interest in HMs that are not well calibrated, particularly when the number of clusters in a sample is small or moderate (e.g., see Browne & Draper [2002] and Seltzer et al. [2002]). (The term calibration is used here in a frequentist sense. Thus, for example, if a simulation study shows that actual levels of coverage of nominal 95% intervals are far from 95%, calibration is poor.)

Alternatively, based on analysis of the data at hand, S or **S** could be chosen such that the modes of the prior and the likelihood are more or less in agreement. When the data are used to specify S and **S**, and the degrees of freedom parameters of the priors are set to small values, such priors would be termed gently data-determined priors (e.g., see Browne, 1998; Browne & Draper (2002); Rasbash et al., 1999, p. 201; Seltzer et al. (2002); Seltzer, Wong, & Bryk, 1996; see also Natarajan & Kass, 2000).

In our work, we used a strategy for specifying gently data-determined priors that is outlined in Seltzer et al. (2002). A simulation study presented in Seltzer et al. focusing on the coverage properties of posterior intervals for fixed effects suggests that such an approach is very promising with respect to producing intervals for fixed effects that are well calibrated. Note that in WinBUGS, models are parameterized in terms of precisions rather than variances. The gently data-determined priors for variances discussed in this endnote translate to gently data-determined gamma and Wishart priors for scalar precisions and precision matrices, respectively (see Seltzer & Choi, 2002).

2. These variables are selected as covariates for student initial status and rate of change based on exploratory data analyses. First, students without behavioral problems have higher mean scores at each grade by two thirds of the corresponding pooled standard deviation than those with behavioral problems. Specifically, the mean differences between the two groups from grades 7 to 10 are 6.2, 7.6, 8.7, and 9.9, respectively. The corresponding pooled standard deviations are 10.0, 10.1, 12.6, and 13.6. Second, across all grades, the mean math scores become higher as students' educational expectations (*SEDEX*) go up. For example, at grade 7, the mean math achievement scores for each of the categories in the *SEDEX* variable are 43.4, 44.9, 46.9, 50.2, 52.5, and 55.0. The correlations between *SEDEX* and mean math achievement scores at each grade are approximately .35 ~ .40. These results strongly suggest that the differences in these two student characteristics should be related to differences in initial status and rates of change.

References

Blomqvist, N. (1977). On the relation between change and initial value. *Journal of the American Statistical Association, 72*, 746-749.

Browne, W. J. (1998). *Applying MCMC methods in multilevel models.* Unpublished doctoral dissertation, University of Bath, UK.

Browne, W. J., & Draper, D. (2002). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. Submitted.

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

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

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

Carlin, B., & Louis, T. (1996). *Bayes and empirical Bayes methods for data analysis.* London: Chapman and Hall.

Chou, C. P., Bentler, P. M., & Pentz, M. A. (1998). Comparison of two statistical approaches to study growth curves: The multilevel model and latent curve analysis. *Structural Equation Modeling: A Multidisciplinary Journal, 5*(3), 247-266.

Chou, C. P., Bentler, P. M., & Pentz, M. A. (2000). A two-stage approach to multilevel structural equation models: Application to longitudinal data. In T.D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 33-49). Mahwah, New Jersey: LEA.

Gelfand, A. E., & Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Associations, 85,* 398-409.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (1995). *Bayesian data analysis.* London: Chapman & Hall.

Gibbons, R. D., Hedeker, M. A., Waternaux, C., & Davis, J. M. (1988). Random regression models: A comprehensive approach to the analysis of longitudinal psychiatric data. *Psychopharmacology Bulletin, 24*(3), 438-443.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in practice.* London: Chapman & Hall.

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

Goldstein, H. I. (1995). *Multilevel statistical models*. London: Edward Arnold.

Hox, J. J. (1993). Factor analysis of multilevel data: Gauging the Muthen model. In J. H. J. Oud & R. A. W. van Block-Vogelesang (Eds.), *Advances in longitudinal and multilevel analysis in the behavioral sciences* (pp. 141-156). Nijmegen: ITS.

Hox, J. J., & Maas, C. J. M. (2000). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling, 8*(2), 157-174.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics, 38*, 963-975.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*, 13-22.

Longford, N. T. (1993). *Random coefficient models*. Oxford, England: Clarendon Press.

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

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

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.

Miller, J. D., Kimmel, L., Hoffer, T. B., & Nelson, C. (2000). *Longitudinal study of American youth: User's manual*. Chicago: International Center for the Advancement of Scientific Literacy, Academy of Sciences.

Muthen, O. 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.

Muthen, O. B. (1994). Multilevel covariance structural analysis. *Sociological Methods and Research, 22*(3), 376-398.

Muthen, O. B. (1997). Latent variable modeling with longitudinal and multilevel data. In A. Raftery (Ed.), *Sociological Methodology 1997* (pp. 453-480). Boston: Blackwell Publishers.

Muthen, O. B. & Satorra, A. (1995). Complex sample data in structural equation modeling. In P. Marsden (Ed.), *Sociological methodology* (pp. 267-316). Washington, DC: American Psychological Association.

Muthen, O. B., & Curran, J. P. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods, 2*(4), 371-402.

Muthen, O. B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing at random. *Psychometrika, 42*, 431-462.

Muthen, L., & Muthen, O. B. (2002). *Mplus User's Guide (version 2.1)*. Los Angeles, CA: Muthen & Muthen.

Muthen, O. B. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika, 29,* 81-117.

Natarajan, R., & Kass, R. (2000). Reference Bayesian methods for generalized linear mixed models. *Journal of American Statistical Association, 95*, 227-237.

Rasbash, J., Browne, W. J., Goldstein, H., Yang, M., Plewis, I., Draper, D., et al. (1999). *A user's guide to MlwiN, Version 2.0*. London: Institute of Education, University of London.

Raudenbush, S. W., & Sampson, R. (1999). Assessing direct and indirect associations in multilevel designs with latent variables. *Sociological methods and research, 28*(2), (pp. 123-153). Newbury Park, CA: Sage.

Raudenbush, S. W., Bryk, A. S., Cheong, Y., & Congdon, R. (2000). *HLM5: Hierarchical linear and non-linear modeling*. Homewood, IL: Scientific Software International.

Raudenbush, S. W., & Bryk, A. S. (2001). *Hierarchical linear models: Applications and data analysis methods, 2nd edition*. Newbury Park, CA: Sage Press.

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

Rachman-Moore, D., & Wolfe, R. (1984). Robust analysis of a nonlinear model for multilevel education survey data. *Journal of Educational Statistics, 9*, 277-293.

Raudenbush, S. W. (1998). To*ward a coherent framework for comparing trajectories of individual change*. Paper presented for the New Methods for the Analysis of Change conference, Pennsylvania State University.

Rogosa, D. R., & Willet, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika, 50,* 203-228.

Seltzer, M. (1993). Sensitivity analysis for fixed effects in the hierarchical model: A Gibbs sampling approach. *Journal of Educational Statistics, 18,* 207-235.

Seltzer, M., & Choi, K. (2002). Sensitivity analysis for multilevel models. In N. Duan & S. Reise (Eds.), *Multilevel modeling: Methodological advances, issues, and applications.* Mahwah, NJ: Earlbaum.

Seltzer, M., Choi, K., & Thum, Y. (2001a). *Examining relationships between where students start and how rapidly they progress: Implications for constructing indicators that help illuminate the distribution of achievement within schools* (CSE Tech. Rep. No. 560). Los Angeles: University of California, National Center for Research on Evaluation, Standards and Student Testing (CRESST).

Seltzer, M., Choi, K., & Thum, Y. (2001b). *Latent variable modeling in the hierarchical modeling framework: Exploring initial status $\times$ treatment interactions in longitudinal studies of intervention* (CSE Tech. Rep. No. 559). Los Angeles: University of California, National Center for Research on Evaluation, Standards and Student Testing (CRESST).

Seltzer, M., Novak, J., Choi, K., & Lim, N. (2002). Sensitivity analysis for hierarchical models employing *t* level-1 assumptions. *Journal of Educational and Behavioral Statistics, 27,* 181-222.

Seltzer, M., Wong, W., & Bryk, A. (1996). Bayesian analysis in applications of hierarchical models: Issues and methods. *Journal of Educational and Behavioral Statistics, 18,* 131-167.

Smith, B. (2000). *Bayesian output analysis program (BOA) version 0.5.0 user manual.* Iowa City, IA: The University of Iowa College of Public Health.

Spiegelhalter, D., Thomas, A., & Best, N. (2000). *WinBUGS: Windows version of Bayesian inference using Gibbs sampling, version 1.3, user manual.* Cambridge, MA: MRC Boistatistics Unit, Cambridge University.

Tanner, M. (1996). *Tools for statistical inference (2nd ed.).* New York: Springer-Verlag.

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