

**Modeling Heterogeneity in Relationships
Between Initial Status and Rates of Change:
Latent Variable Regression in a
Three-Level Hierarchical Model**

CSE Report 647

Kilchan Choi and Michael Seltzer
National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)/
University of California, Los Angeles

March 2005

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 Methodologies for Assessing Student Progress, Strand 1
Project Directors: Michael Seltzer, Yeow Meng Thum, and Kilchan Choi

Copyright © 2005 The Regents of the University of California

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R305B960002, 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 Center for Education Research, the Institute of Education Sciences, or the U.S. Department of Education.

MODELING HETEROGENEITY IN RELATIONSHIPS BETWEEN INITIAL STATUS AND RATES OF CHANGE: LATENT VARIABLE REGRESSION IN A THREE-LEVEL HIERARCHICAL MODEL

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

Abstract

In studies of change in education and numerous other fields, interest often centers on how differences in the status of individuals at the start of a time period of substantive interest relate to differences in subsequent change. In this report we present a fully Bayesian approach to estimating three-level hierarchical models in which latent variable regression coefficients capturing the relationship between initial status and rates of change within each of J schools ($Bw_j, j = 1, \dots, J$) are treated as varying across schools. Through analyses of data from the Longitudinal Study of American Youth, we show how modeling differences in Bw_j as a function of school characteristics can broaden the kinds of questions we can address in school effects research. We illustrate the possibility of conducting sensitivity analyses employing t distributional assumptions at each level of such models (termed LVR-HM3s) and present results from a simulation study that focuses on the coverage properties of marginal posterior intervals for fixed effects in LVR-HM3s. We outline extensions of LVR-HM3s to settings in which growth is nonlinear, and discuss the use of LVR-HM3s in other types of research including multi-site evaluation studies in which time-series data are collected during a pre-intervention period, and cross-sectional studies in which within-cluster latent variable regression slopes are treated as varying across clusters.

Key questions in studies of change often center on how differences in the status of individuals at the start of a time period of substantive interest relate to differences in subsequent change. Examples can be found in an array of fields and in studies of various types (e.g., large-scale panel studies; longitudinal studies of interventions). For example, in the area of public health, Svardsudd and Blomqvist (1978) focused on the question of whether men with higher levels of blood pressure at age 50 tend to experience more rapid increases in blood pressure in subsequent years (see also Blomqvist, 1977, and Wu, Ware, & Feinlab, 1980). In gerontology, researchers have studied whether levels of cognitive functioning at certain points in adulthood are related to subsequent changes in cognitive functioning over time (see, e.g., Adler, Adam, & Arenberg, 1990). In education, there has been a long-standing interest in questions concerning the extent to which growth in particular skills or content areas is related to differences in status at the start of a series of grades of interest (see, e.g.,

Bloom, 1964; Werts & Hilton, 1977). Furthermore, in longitudinal studies of interventions, investigators have been interested in baseline \times treatment interactions, that is, whether expected differences in rates of change between individuals in treatment and control conditions depend crucially on initial status (see, e.g., Khoo, 2001, Muthén & Curran, 1997).

To overcome attenuation problems that can arise, for example, in a regression of ordinary least squares (OLS) estimates of growth rate parameters ($\hat{\pi}_{1i}$) on OLS estimates of initial status parameters ($\hat{\pi}_{0i}$) for a sample of individuals ($i = 1, \dots, N$), Blomqvist (1977) focused attention on the latent variable regression (LVR) of π_{1i} on π_{0i} . In particular, he derived the maximum likelihood (ML) estimate of the latent variable regression coefficient (b) relating differences in π_{0i} to π_{1i} . In the computational formulae that Blomqvist presented, the ML estimate of b is obtained through an adjustment of the coefficient resulting from a regression of $\hat{\pi}_{1i}$ on $\hat{\pi}_{0i}$.

Specifying relationships among latent variables is a hallmark of the Structural Equation Modeling (SEM) framework. Thus, employing initial status parameters as predictors of rates of change and obtaining ML estimates of LVR coefficients can be accomplished readily in numerous implementations of the SEM framework, for example, Mplus (Muthén & Muthén, 2002), EQS (Bentler, 2002), LISREL (Joreskog & Sorbom, 2000), and AMOS (Arbuckle & Wothke, 1999).

In an important extension of the hierarchical modeling framework, Raudenbush and Sampson (1999) presented an ML-based strategy for incorporating latent variable regressions into hierarchical models (HMs), which has been implemented in the HLM 5 software program (Raudenbush, Bryk, Cheong, & Congon, 2000). Raudenbush and Bryk (2002) presented several examples of this approach, including an analysis comparing growth in math achievement for girls and boys during high school controlling for differences in status in eighth grade.

Seltzer, Choi, and Thum (2003) employed a fully Bayesian approach to estimating two-level HMs involving latent variable regressions (e.g., settings in which time-series observations are nested within students) and discussed various advantages of such an approach (e.g., the ability to employ t distributional assumptions). In this report, we wish to explore valuable extensions of this approach to three-level settings (e.g., settings in which time-series observations are nested within students, who in turn are nested within a sample of J schools). In particular, we focus on three-level HMs in which latent variable regression

coefficients capturing the relationship between initial status and rates of change within each of J schools ($Bw_j, j = 1, \dots, J$) are treated as varying across schools. We illustrate how models of this kind (which we term LVR-HM3) can help broaden the kinds of questions we are able to address in school effects research, and we discuss various possibilities that arise in applications of such models in other research settings.

Specifically, in the case of school effects research based on analyses of longitudinal data from such studies as NELS, interest typically centers on differences among schools in their mean rates of change, and investigating how differences in various school policies and practices relate to differences in school mean rates of change. An advantage of the models we propose is that the magnitude of LVR initial status/rate of change slopes (Bw_j) provides information regarding the distribution of growth in achievement within schools. In some schools, those students with relatively high initial status may progress extremely rapidly, whereas those with lower initial achievement levels may display markedly slower progress rates. In such cases, initial differences in achievement become magnified over time. In other schools, however, initial status may be weakly related to subsequent rates.

Thus, formally investigating differences among schools in the magnitude and direction of LVR initial status/rate of change slopes, as well as differences in school mean rates of change, enables us to address questions of the following kind: How do differences in various school policies, practices, and intake characteristics relate to differences between schools in their mean rates of change *and* in the magnitude of their initial status/rate of change slopes? Are there particular school factors that are associated with both high mean rates of progress *and* more equitable distributions of growth in achievement (i.e., schools in which initial status is relatively weakly related to subsequent progress).

Such an approach can also help broaden the kinds of questions that we are able to address in multi-site studies of interventions. Consider, for example, a study in which an innovative remedial reading program has been implemented in one set of schools, and a more standard form of remedial instruction has been implemented in another set. In certain sites, those children with very low initial reading skills may continue to do poorly, while those with less serious problems tend to make substantial progress. In other sites, those students with very low initial reading skills might exhibit rapid rates of progress, while rates of progress may be less rapid

for those with less serious problems at the outset. Thus it would be valuable in such settings to conduct post hoc analyses in which differences in Bw_j , as well as site mean rates of change, are modeled as a function of the type of program a school was assigned to, differences in implementation across schools, and the like. Extensions to settings in which time-series observations are collected during a pre-intervention phase are discussed below.

The LVR-HM3 framework that we present also has important applications in the case of cross-sectional studies. Consider, for example, a sample that consists of students ($i = 1, \dots, n_j$) nested within different schools ($j = 1, \dots, J$). Suppose that similar to the kinds of analyses of the High School and Beyond data presented in Raudenbush and Bryk (2002), we wish to model student achievement in 12th grade as a function of student SES. Note importantly that if standard errors of measurement for these variables are available, we can pose a three-level model that contains a measurement model at level 1 in which the observed achievement and SES values for each student (i.e., Y_{ij}, X_{ij}) are modeled as a function of their corresponding true scores (i.e., $\theta_{ach(ij)}, \theta_{ses(ij)}$). In a within-school (level-2) model, we can then model $\theta_{ach(ij)}$ as a function of $\theta_{ses(ij)}$. The LVR coefficients capturing the relationship between SES and achievement within the J schools ($Bw_j, j = 1, \dots, J$) can then be modeled, along with school mean achievement parameters (β_{0j}), as a function of key school-level characteristics in a between-school (level-3) model.

While LVR-HM3s are extremely complex from an estimation standpoint, we show that simulating the marginal posterior distributions of the parameters in such models via the Gibbs sampler is highly feasible. This report consists of the following sections. We first outline a fairly simple two-level HM involving a regression of rate of change on initial status (LVR-HM2) and then extend this model to a basic LVR-HM3 setting. After discussing the strengths and limitations of various approaches to estimating latent variable regressions in analyses of longitudinal data (e.g., Chou, Bentler, & Pentz, 2000; Muthén & Curran, 1997; Raudenbush & Sampson, 1999), we present the logic of a fully Bayesian approach based on the use of the Gibbs sampler. (See, e.g., Carlin & Louis, 1996; Gelfand, Hills, Racine-Poon, & Smith, 1990; Gelman, Carlin, Stern, & Rubin, 1995; Gilks, Richardson, & Spiegelhalter, 1996; and Seltzer, Wong, & Bryk, 1996 for discussions and illustrations of the use of the Gibbs sampler in fitting HMs.) Through analyses of the data from a subsample of the Longitudinal Study of American Youth (LSAY; Miller, Kimmel, Hoffer, & Nelson, 1999), we then illustrate the use of LVR-HM3 models in school effects research. Next, we conduct a

sensitivity analysis employing t distributional assumptions at each level of our final model. This has the effect of downweighting possible outliers at each level (i.e., outlying time-series observations, students and schools). We then focus on factors that improve mixing of the Gibbs sampler in LVR-HM3 settings, and present results from a simulation study that focuses on the coverage properties of marginal posterior intervals, and the degree of bias of marginal posterior means, for fixed effects in the LVR-HM3. In the final section, we discuss other applications of the LVR-HM3 and extensions of these models to settings in which growth is curvilinear.

Latent Variable Regression in a Hierarchical Modeling Framework

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

For heuristic purposes, we first specify a simple two-level HM for longitudinal analysis in which rate of change is regressed on initial status (LVR-HM2). In the following within-person (level-1) model, we model the time series of outcome values for each of N individuals ($i = 1, \dots, N$) as a simple linear function of time:

$$Y_{it} = \pi_{0i} + \pi_{1i}a_{it} + \varepsilon_{it} \quad \varepsilon_{it} \sim N(0, \sigma^2), \quad (1)$$

where Y_{it} is the outcome score for individual i at measurement occasion t ($t = 1, \dots, T_i$), and a_{it} represents, for example, an individual's age or grade at measurement occasion t . In the case of an intervention study, a_{it} might represent the amount of time that has elapsed since the start of treatment. π_{1i} represents the growth rate for individual i , and π_{0i} is the status for individual i when $a_{it} = 0$. If we want π_{0i} to represent status at the start of a span of time of substantive interest (i.e., initial status), which is what we desire here, then a_{it} must be coded in such a way that $a_{it} = 0$ corresponds to the start of the time period. The ε_{it} are residuals assumed normally distributed with mean 0 and variance σ^2 .

We now pose the following between-person (level-2) model:

$$\pi_{0i} = \beta_{00} + r_{0i} \quad r_{0i} \sim N(0, \tau_{00}) \quad (2a)$$

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

where β_{00} and β_{10} are, respectively, the population means for initial status and growth rates, and r_{0i} and r_{1i} are random effects assumed multivariate normally distributed with mean $\mathbf{0}$ and variance-covariance matrix

$$\mathbf{T}_r = \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}.$$

The variance terms τ_{00} and τ_{11} capture the extent to which individuals vary in their initial status and rates of change, and τ_{01} represents the covariance between initial status and growth rates.

We now model π_{1i} as a function of π_{0i} as follows:

$$\pi_{0i} = \beta_{00} + r_{0i} \quad r_{0i} \sim N(0, \tau_{00}) \quad (3a)$$

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

Thus b is a latent variable regression coefficient capturing the amount of change that we expect in π_{1i} when π_{0i} increases one unit (see, for example, Muthén & Curran, 1997; Raudenbush & Bryk, 2002; Seltzer et al., 2003). As in the previous model, τ_{00} represents the extent to which individuals vary in their initial status. τ_{11} , however, now represents the amount of variance in growth rates that remains after taking into account differences in initial status. Since we are conditioning on initial status in Equation 3b, we assume that $Cov(r_{0i}, r_{1i}) = 0$.

Note finally that in Equation 3b we have centered π_{0i} around the population mean for initial status (β_{00}) (see also Blomqvist, 1977). There are two reasons for doing so. First, this gives β_{10} a useful interpretation; that is, β_{10} now represents the expected rate for an individual whose initial status is equal to β_{00} . And secondly, as will be discussed in a later section, centerings of this kind can greatly reduce the degree of autocorrelations among sampled values generated by the Gibbs sampler in applications involving latent variable regressions.

Latent Variable Regression in a Three-Level Hierarchical Model (LVR-HM3)

We now extend the above two-level model and specify a simple LVR-HM3 in the context of a longitudinal school effects study. For heuristic purposes, we do not include any observed student- or school-level predictors in the model at this juncture. At level 1 (see Equation 4 below), the outcome value for student i , in school j ($j = 1, \dots, J$), at time t (Y_{ijt}), is modeled as a function of time (a_{ijt}). In a level-2 (within-school/between-student) model, the rate of change for student i in school j (π_{1ij}) is modeled as a function of initial status (π_{0ij}).

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

$$\pi_{0ij} = \beta_{00j} + r_{0ij} \quad r_{0ij} \sim N(0, \tau_{\pi_{0j}}) \quad (5a)$$

$$\pi_{1ij} = \beta_{10j} + Bw_j(\pi_{0ij} - \beta_{00j}) + r_{1ij} \quad r_{1ij} \sim N(0, \tau_{\pi_{1j}}). \quad (5b)$$

A key parameter in this model is Bw_j , which is a latent variable regression coefficient capturing the relationship between initial status and rates of change in school j . We refer to these coefficients as within-school initial status/rate of change slopes (Seltzer et al., 2003). By virtue of centering π_{0ij} around the mean initial status value for school j (i.e., β_{00j}), β_{10j} represents the mean rate of change for school j . Further, the level-2 random effects for the students in school j (i.e., r_{0ij} and r_{1ij}) are assumed independently and normally distributed with mean 0 and variances $\tau_{\pi_{0j}}$ and $\tau_{\pi_{1j}}$, respectively. Note importantly that we allow the variances of r_{0ij} and r_{1ij} to differ across schools.

School mean initial status (β_{00j}), school mean rate of change (β_{10j}), and the within-school initial status/rate of change slope for school j (Bw_j) are treated as outcomes in a level-3 (between-school) model. To convey some of the possibilities of our modeling approach, we can pose a level-3 model in which school mean rate of change (β_{10j}) and within-school initial status/rate of change slopes (Bw_j) are modeled as a function of school mean initial status (β_{00j}):

$$\beta_{00j} = \gamma_{000} + u_{00j} \quad u_{00j} \sim N(0, \tau_{\beta_{00}}) \quad (6a)$$

$$\beta_{10j} = \gamma_{100} + Bb(\beta_{00j} - \gamma_{000}) + u_{10j} \quad u_{10j} \sim N(0, \tau_{\beta_{10}}) \quad (6b)$$

$$Bw_j = Bw_0 + Bw_1(\beta_{00j} - \gamma_{000}) + u_{Bwj} \quad u_{Bwj} \sim N(0, \tau_{Bw}) \quad (6c)$$

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

In Equation 6a, γ_{000} represents the grand mean for initial status. In Equation 6b, a latent variable regression coefficient, Bb , captures the relationship between school mean initial status and school mean rates of change. In contrast to the Bw_j parameters we term this a between-school initial status/rate of change slope. By centering school mean initial status around the grand mean for initial status (γ_{000}), γ_{100} represents the expected rate of change when school mean initial status (β_{00j}) is equal to γ_{000} . In Equation 6c, Bw_1 is a latent variable regression coefficient relating differences in school mean initial status to differences in within-school initial

status/rate of change slopes, and Bw_0 is the expected within-school initial status/rate of change slope when school mean initial status is equal to γ_{000} .

The random effects in the above model (i.e., u_{00j} , u_{10j} , u_{Bwj}) are assumed multivariate normally distributed with mean $\mathbf{0}$ and variance-covariance matrix

$$\mathbf{T}_u = \begin{pmatrix} \tau_{\beta_{00}} & 0 & 0 \\ 0 & \tau_{\beta_{10}} & \tau_{\beta_{10},Bw} \\ 0 & \tau_{Bw,\beta_{10}} & \tau_{Bw} \end{pmatrix} \quad (6d)$$

The variance term $\tau_{\beta_{00}}$ captures the extent to which schools vary in their school mean initial status, and $\tau_{\beta_{10}}$ and τ_{Bw} are, respectively, residual variances for school mean rates of change and within-school initial status/rate of change slopes after taking into account school mean initial status. With respect to the off-diagonal elements of \mathbf{T}_u , we assume that $Cov(u_{00j}, u_{10j}) = 0$ and $Cov(u_{00j}, u_{Bwj}) = 0$, since β_{00j} is employed as a predictor in Equations 6b and c. $\tau_{\beta_{10},Bw}$ captures the covariance between u_{10j} and u_{Bwj} . Note that for later use we define the submatrix:

$$\mathbf{T}_{2,3} = \begin{pmatrix} \tau_{\beta_{10}} & \tau_{\beta_{10},Bw} \\ \tau_{Bw,\beta_{10}} & \tau_{Bw} \end{pmatrix} . \quad (6e)$$

As will be seen, observed predictors can easily be incorporated into the above LVR-HM3. In particular, β_{10j} and Bw_j can be modeled as a function of school-level characteristics of interest.

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

Recently, several pioneering researchers have developed strategies for incorporating latent variable regressions in hierarchical modeling settings (Chou et al., 2000; Muthén, 1997; Raudenbush & Sampson, 1999). First, Raudenbush and Sampson's approach (1999), which has been implemented in the latest release of the HLM software program (version 5; Raudenbush et al., 2000), can be applied in a variety of LVR-HM2 settings. In this strategy, the regression coefficients for latent predictors are estimated by a two-stage procedure. Consider, for example, the HM depicted in Equations 1, 3a and 3b. We first fit a more standard model that does not contain latent variable regressions, which in this case would be the HM specified in Equations 1, 2a and 2b. The resulting ML estimate of the level-2 variance-covariance matrix in Equation 2 would then be used to estimate b in Equation 3b, that is, $\hat{b} =$

$\hat{\tau}_{01} / \hat{\tau}_{00}$. A two-step procedure would also be employed in more complex LVR-HM2 settings. Raudenbush and Bryk (2002) presented an example in which they wish to estimate an expected difference in rates of change in mathematics achievement between high school girls and boys, controlling for differences in status in Grade 8 mathematics achievement. Thus the estimand of primary interest would be the coefficient for gender in the following level-2 model: $\pi_{1i} = \beta_{10} + \beta_{11} \text{GENDER}_i + b \pi_{0i} + r_{1i}$. Implementing the Raudenbush and Sampson (1999) approach would first entail fitting a standard growth model in which initial status and rates of change are modeled as a function of gender, thus yielding estimates of the difference in initial status between girls and boys, the unadjusted gender difference in rates of change, and a level-2 variance-covariance matrix conditional on gender. These quantities provide the basis of estimating the LVR coefficient for π_{0i} and an adjusted difference in growth rates (i.e., the direct effect of gender on growth rates) (see Raudenbush & Bryk, 2002, chapter 11).

In terms of three-level models, one could, for example, use HLM 5 to fit a three-level model in which school mean growth rates are modeled as a function of school mean initial status. Similar to the procedure for fitting a LVR-HM2 described above, this would first entail fitting a standard three-level model that did not contain latent variable regressions, and then using the ML estimates of the covariance between school mean growth rates and initial status, and the variance in school mean initial status to obtain an estimate of the desired LVR coefficient.

It is not possible, however, to use Raudenbush and Sampson's (1999) two-step approach in fitting LVR-HM3 models in which student growth rates are modeled as a function of initial status in a level-2 (within-school) model, and within-school initial status/rate of change slopes are treated as varying across schools. One possibility, however, would entail using HLM 5 to fit, for example, an LVR-HM2, such as the one specified in Equations 1, 3a and 3b, to the data for each of J schools in a sample. Using the resulting ML estimates of b and their standard errors, we could then employ a meta-analytic modeling framework, thus enabling us to obtain estimates of the variance in within-school initial status/rate of change slopes across schools and model differences in such slopes as a function of school-level predictors. Note, however, that when the number of individuals per cluster is small (i.e., in situations where large sample properties of ML estimators likely do not apply), the resulting ML estimates of b and their standard errors may be problematic.

A second approach can be found in Muthén's work on multilevel SEM (Muthén, 1994; see also Hox, 1993; Hox & Maas, 2000; Muthén, 1997; Muthén & Satorra, 1995). Essentially, this approach summarizes unbalanced nested data structures by means of within- and between-group variance-covariance matrices (see Muthén, 1994). Based on these matrices, the structural relations among within- and between-group variables can be estimated simultaneously using a multiple-group modeling technique.

In three-level hierarchical modeling settings where the numbers of level-2 and level-3 units are fairly large, the current implementation of the multilevel SEM approach allows us to estimate a pooled within-school initial status/rate of change coefficient and a between-school initial status/rate of change coefficient. However, as in the case of Raudenbush and Sampson's (1999) approach, because a separate within-school variance-covariance matrix for each school is not estimated, we cannot treat latent variable regression coefficients as varying across schools.

Finally, Chou et al. (2000) presented a two-stage approach to multilevel SEM. Their approach is somewhat analogous to Burstein's (1980) "slopes-as-outcomes" framework for analyzing multilevel data. The first step of their approach entails fitting a very general mean and/or covariance structure model to the data for each cluster (e.g., school). Parameter estimates for each cluster are then treated as outcomes in a between-school regression model. In the case of the LVR-HM3 posed in Equations 4, 5 and 6, the first step would entail fitting the LVR-HM2 specified in Equations 1, 3a and 3b to the data for each school. In the second step, estimates of b for the set of J schools would be regressed on estimates of school mean initial status; similarly, estimates of school mean growth rates would be regressed on estimates of school mean initial status as well. Note that the second step is carried out via OLS. In contrast to employing a random effects meta-analytic model at this step, such an approach would not, for example, provide us with an estimate of the variance in initial status/rate of change slopes (τ_{bw}), nor would those clusters with more precise estimates of such slopes receive more weight in the analysis. Note also that in applications in which cluster sizes (n_j) are small, regressing estimates of b , or school mean rates, on estimates of school mean initial status ($\hat{\beta}_{00j}$) could result in attenuated estimates of the slopes relating school mean initial status to these outcomes.

The Logic of the Fully Bayesian Approach in Estimating LVR-HMs

We now discuss the logic of the fully Bayesian approach in LVR-HM settings and the use of the Gibbs sampler in implementing such an approach. In doing so, we first focus on the LVR-HM2 specified in Equations 1, 3a, and 3b. The fully Bayesian approach entails calculating the marginal posterior distributions of parameters of interest (e.g., b) from the joint posterior distribution of all unknowns in the model. The joint posterior distribution is proportional to the product of the likelihood for the observed data and the prior specification for all unknowns. The level-1 model depicted in Equation 1 provides the basis of the likelihood for the observed data, Equation 3a provides the prior specification for the π_{0i} , and Equation 3b provides the prior specification for the π_{1i} . To complete the prior specification we must also place priors on the variance components and fixed effects. We assume that these parameters are independent *a priori*, which is a commonly employed assumption in applications of HMs, and for each of these parameters, we assume that all possible values are equally probable *a priori* and hence proportional to a constant (e.g., $p(\tau_{11}) \propto 1, (0 < \tau_{11} < \infty)$). Such non-informative uniform priors are common in the literature (see, e.g., Browne & Draper, 2000; Gelman et al., 1995; Rubin, 1981; Seltzer, Novak, Choi, & Lim, 2002). Alternative prior specifications are discussed below. Thus

$$\begin{aligned}
 & p(\beta_{00}, \beta_{10}, b, \tau_{00}, \tau_{11}, \sigma^2, \pi_0, \pi_1, | Y) \propto \\
 & \prod_{i=1}^N \prod_{t=1}^{T_i} (1/\sigma^2) \exp \left[- (1 / 2\sigma^2) (Y_{it} - [\pi_{0i} + \pi_{1i} a_{it}])^2 \right] \\
 & \times \prod_{i=1}^N (1 / \tau_{00})^{1/2} \exp \left[- (1 / 2\tau_{00})(\pi_{0i} - \beta_{00})^2 \right] \\
 & \times \prod_{i=1}^N (1 / \tau_{11})^{1/2} \exp \left[- (1 / 2\tau_{11})(\pi_{1i} - [\beta_{10} + b(\pi_{0i} - \beta_{00})])^2 \right] \tag{7}
 \end{aligned}$$

where $\pi_0 = (\pi_{0(1)}, \pi_{0(2)}, \dots, \pi_{0(N)})'$, and $\pi_1 = (\pi_{1(1)}, \pi_{1(2)}, \dots, \pi_{1(N)})'$. Suppose, for example, that interest centers on drawing inferences concerning b . Integrating over all other unknowns in the joint posterior yields the marginal posterior of b . The difficulty is that the required integrations for obtaining marginal posteriors of unknowns in all but the simplest of HMs are intractable. The virtue of the Gibbs sampler is that it provides an alternative means of obtaining marginal posterior distributions of

parameters of interest in complex modeling settings. The Gibbs sampler, upon convergence, essentially yields draws from the joint posterior of all unknowns in a model. From these sampled values, as will be seen below, we are able to simulate the marginal posterior of any parameter of interest (see, e.g., Gelfand et al., 1990; Morris, 1987).

To implement the Gibbs sampler, we subdivide the unknowns in the joint posterior into a number of subsets, such that it is easy to sample from the conditional posterior of each subset given the data and the current values (i.e., the most recently sampled values) for the other subsets of unknowns.

To illustrate, we now consider the form of the full conditional posterior distribution of the fixed effects in the level-2 equation in which rates of change are modeled as a function of initial status (i.e., $\beta_R = (\beta_{10}, b)'$). Given the data (Y) and all other unknowns, the parts of the joint posterior corresponding to the likelihood for the data and to the prior for the π_{0i} reduce to constants. Thus

$$p(\beta_R \mid Y, \pi_1, \pi_0, \beta_{00}, \tau_{11}, \tau_{00}, \sigma^2) \propto \prod_{i=1}^N (1 / \tau_{11})^{1/2} \exp [- (1 / 2\tau_{11})(\pi_{1i} - [\beta_{10} + b(\pi_{0i} - \beta_{00})])^2] . \quad (8)$$

Note that given π_1 , π_0 and τ_{11} , we are essentially in a linear modeling setting in which we are modeling a set of outcomes (i.e., the π_{1i}) as a function of a predictor variable deviated around its mean (i.e., $\pi_{0i} - \beta_{00}$), and where the residual variance (i.e., τ_{11}) is known. Based on standard Bayesian results for linear models (see, e.g., Box & Tiao, 1973, chapter 2):

$$\beta_R \mid Y, \pi_1, \pi_0, \beta_{00}, \tau_{11}, \tau_{00}, \sigma^2 \sim MVN_2 (\hat{\beta}_R, \tau_{11}(X'X)^{-1}) \quad (9)$$

where

$$\hat{\beta}_R = (X'X)^{-1} X'\pi_1, \quad (10)$$

and

$$\mathbf{X} = \begin{pmatrix} 1 & (\pi_{0(1)} - \beta_{00}) \\ 1 & (\pi_{0(2)} - \beta_{00}) \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & (\pi_{0(N)} - \beta_{00}) \end{pmatrix} .$$

For this particular application, we would sample sequentially from a series of six full conditional posteriors: $(\beta_R | \cdot)$; $(\pi_1 | \cdot)$; $(\pi_0 | \cdot)$; $(\beta_{00} | \cdot)$; $(\tau_{00} | \cdot)$; and $(\tau_{11} | \cdot)$, where the notation “ $| \cdot$ ” indicates that we are given the data and all other unknowns in the joint posterior (see, for example, Gilks et al., 1996). Each of these conditional posteriors has a known distributional form (e.g., normal, inverse gamma) and can be sampled from readily.

In implementing the Gibbs sampler, we proceed sequentially through the steps of the algorithm, sampling once from each conditional posterior given the data and the current values of the other unknowns in the joint posterior. Thus, for example, after completing the k^{th} iteration or cycle of the algorithm, we would sample from the conditional posterior for β_R (see Equation 9) with $\pi_1 = \pi_1^{(k)}$, $\pi_0 = \pi_0^{(k)}$, $\beta_{00} = \beta_{00}^{(k)}$, $\tau_{11} = \tau_{11}^{(k)}$, $\tau_{00} = \tau_{00}^{(k)}$ and $\sigma^2 = \sigma^{2(k)}$, thus yielding $\beta_R^{(k+1)}$. Upon convergence ($k = C$), the values generated for the parameters in the model in M subsequent iterations (i.e., $(\beta_R^{(k)}, \pi_1^{(k)}, \pi_0^{(k)}, \beta_{00}^{(k)}, \tau_{11}^{(k)}, \tau_{00}^{(k)}, \sigma^{2(k)})$, ($k = C + 1, \dots, C + M$)) would constitute M draws from the joint posterior. Thus, with M set to a large value, the empirical distribution of the M values generated for any parameter of interest in the model (e.g., b ($b^{(k)}$, $k = C + 1, \dots, C + M$)) would provide us with an accurate approximation of the marginal posterior distribution of that parameter (e.g., $p(b | Y)$).

In LVR-HM3 modeling settings, we can readily use the Gibbs sampler to obtain the marginal posterior distribution of any unknown of interest (e.g., Bw_{-1} in Equation 6c). For example, a key step in implementing the Gibbs sampler in the case of the LVR-HM3 specified above would entail sampling from the full conditional posterior distribution of the fixed effects in the level-3 equations in which school mean rates of change (β_{10j}) and within-school initial status/rate of change slopes (Bw_j) are modeled as a function of school mean initial status (β_{00j}): $p(\theta | \cdot)$, where $\theta = (\gamma_{10} \ Bb \ Bw_{-0} \ Bw_{-1})'$. In this step, we are given β_{10j} and Bw_j , that is, the outcomes in the second and third equations at level 3 ($\beta_{(2,3)j} = (\beta_{10j} \ Bw_j)'$ ($j = 1, \dots, J$)). Furthermore, we are given the centered predictor values (i.e., $(\beta_{00j} - \gamma_{000})$), and we are given the variance-covariance matrix connected with the second and third equations in the level-3 model, that is, $T_{2,3}$ in Equation 6e. Thus, we are in a linear modeling setting where for each of J schools we have a set of outcomes $(\beta_{(2,3)j})$ assumed bivariate normally distributed with conditional mean $W_j \theta$ and known variance-covariance $T_{2,3}$, where

$$\mathbf{W}_j = \begin{pmatrix} 1 & (\beta_{00j} - \gamma_{000}) & 0 & 0 \\ 0 & 0 & 1 & (\beta_{00j} - \gamma_{000}) \end{pmatrix}. \quad (11)$$

Again placing uniform priors on the fixed effects and variance components over the range of possible values for these parameters, it can be shown that:

$$\theta \mid . \sim MVN_4(\hat{\theta}, D_\theta), \quad (12)$$

where

$$\hat{\theta} = \left[\sum_{j=1}^J \mathbf{W}_j' (\mathbf{T}_{2,3})^{-1} \mathbf{W}_j \right]^{-1} \sum_{j=1}^J \mathbf{W}_j' (\mathbf{T}_{2,3})^{-1} \beta_{(2,3)j} \quad (13)$$

and

$$D_\theta = \left[\sum_{j=1}^J \mathbf{W}_j' (\mathbf{T}_{2,3})^{-1} \mathbf{W}_j \right]^{-1}. \quad (14)$$

As in the case of the LVR-HM2 model discussed above, the joint posterior of all unknowns for the above LVR-HM3 can be subdivided into a number of subsets such that the full conditional posterior of each subset has a known distributional form and can be sampled from readily. Such Gibbs sampling formulations can be easily extended to more complex LVR-HM3 settings in which observed covariates are also included at any level of the model.

An Illustrative Example Using Data From the LSAY

To help illustrate various possibilities that arise in employing LVR-HM3 in school effects research, we turn to analyses of the time series data for 2,628 students nested within 45 schools in the LSAY sample. We focus on mathematics achievement time series data collected at the start of Grades 7, 8, 9 and 10. Mathematics achievement scores are based on a scale that was developed using IRT methods. The mean achievement scores at Grades 7, 8, 9 and 10 are, respectively, 50.09, 53.74, 58.57, and 63.45 (see Table 1). The corresponding standard deviations are 9.97, 10.89, 12.56, and 13.63.

Extensive preliminary analyses of the data indicate that a linear model of individual growth (see Equation 4) tends to provide an adequate representation of growth (see Appendix for details). We discuss extensions of our framework to settings in which growth is nonlinear in a later section of our paper.

Table 1
Descriptive Statistics Based on the LSAY Data

	Variable name	<i>N</i>	Mean	<i>SD</i>
Outcomes				
Math achievement at Grade 7	Y_{1ij}	2,592	50.09	9.97
Math achievement at Grade 8	Y_{2ij}	2,257	53.74	10.89
Math achievement at Grade 9	Y_{3ij}	1,952	58.57	12.56
Math achievement at Grade 10	Y_{4ij}	1,784	63.45	13.63
Student-level variables				
Home resources	$HOMERES_{ij}$	2,628	2.54	1.14
Behavioral problems	$BEHAV_PBLMS_{ij}$	2,628	0.17	0.37
Educational expectations	ED_EXPEC_{ij}	2,628	4.00	1.41
School-level variables				
School mean home resources	$\overline{HOMERES}_{.j}$	45	2.53	0.30
Teacher effort	$\overline{TCHEFF}_{.j}$	45	4.41	0.55

We first discuss the specification of priors for the fixed effects and variance components in our models. Next, to help explore the heterogeneity in within-school initial status/rate of change slopes, we fit the two-level LVR model defined by Equations 1, 3a and 3b to each school's data. We then fit the LVR-HM3 specified in Equations 4, 5 and 6 (Model 1), and then expand this model by including key observed student- and school-level covariates (Model 2).

Note that all examples were carried out using WinBUGS1.4 (Spiegelhalter, Thomas, Best, & Lunn, 2003). Annotated copies of our programs are available upon request.

Specification of priors. For the fixed effects and variance components in our models, we employed diffuse priors, that is, priors that allow the data to dominate our inferences. Specifically, for the fixed effects, we specified normal priors with a mean of 0 and extremely low precision (i.e., 1.0E-5). Such priors are functionally equivalent to uniform priors.

We placed uniform priors on all scalar variances (e.g., σ^2 in Equation 4; τ_{π_0j} , $\tau_{\pi_{1j}}$ ($j = 1, \dots, J$) in Equations 5a and 5b; $\tau_{\beta_{00}}$ in Equation 6a). Models in WinBUGS are parameterized in terms of precisions (e.g., $1/\tau_{\beta_{00}}$) rather than variances (e.g., $\tau_{\beta_{00}}$). Spiegelhalter et al. (2003) noted that placing uniform priors on scalar variances

translates to placing particular Pareto priors on scalar precisions. Thus we employed Pareto priors of the form *Pareto*(1, .0001).

Finally, we consider the prior specification for the random effects variance-covariance matrix connected with the level-3 equations for β_{10j} and Bw_j in our models (see, e.g., $T_{2,3}$ in Equation 6e). In specifying priors for a random effects variance-covariance matrix of dimension m , a common choice in the literature is to employ inverse Wishart (*IW*) priors with small degrees of freedom (ν) and scale matrix \mathbf{S} . (Note that ν must be at least as large as m .) Such priors translate to Wishart priors with ν degrees of freedom and scale matrix \mathbf{S}^{-1} for precision matrices (e.g., $T_{2,3}^{-1}$). (In WinBUGS, one is limited to specifying Wishart (*W*) priors for precision matrices connected with random effects assumed multivariate normally distributed.)

It is important to note that the mode of an *IW* or *W* prior will, for a given value of ν , depend critically upon one's choice of \mathbf{S} . Hence \mathbf{S} must be chosen with care. In situations where there is little prior information concerning the variance components, one may try one's best to use such information in choosing sensible values for \mathbf{S} . However, one may find retrospectively that the mode of one's prior conflicts substantially with the mode of the likelihood (e.g., the mode of $L(T_{2,3})$). This can have some undesirable consequences, especially when the number of clusters in a sample is small or moderate. For example, focusing on two-level HMs, Seltzer et al. (2002) noted that if the prior modes for random effects variance parameters are substantially smaller than the ML estimates of such parameters, the marginal posterior intervals of fixed effects of interest may in fact be appreciably narrower than those based on an empirical Bayes (EB) approach (i.e., an approach in which the ML estimates of the variance components are treated as the known, true values of such parameters).

An increasingly common approach that helps avoid such difficulties and attendant calibration problems involves using information based on the data at hand to help specify \mathbf{S} (see Seltzer et al., 2002, and Browne & Draper, 2000, 2003). Such an approach is now the default in MLWin's MCMC routines developed by Browne and Draper (see Rasbash et al., 2002). This approach gives rise to *IW* and *W* priors that are data-dependent, but with ν set to a small value, their information content is very low in relation to the likelihood. Thus we employed diffuse data-dependent priors for $T_{2,3}$ in our analyses. See Endnote 1 for specific details, along with details concerning an alternative strategy that we also employed. (For important work on

the use and value of data-dependent priors in other modeling settings, see Wasserman, 2000, and Natarajan & Kass, 2000.)

As Wasserman (2000) and Browne and Draper (2003) noted, in settings in which little information is available *a priori*, it is useful to seek diffuse priors that yield posterior intervals for parameters of interest whose actual levels of coverage are close to nominal, and point estimates (e.g., posterior means) with low bias. We later present results from a simulation study that suggest that our strategy for specifying priors for the fixed effects and variance components in our models gives rise to results with such properties.

Assessing convergence. For assessing convergence and mixing of the Gibbs sampler, we examined trace plots and autocorrelation function (ACF) plots. For each analysis, we ran two chains using different starting values with a burn-in period of 2,000 iterations, and ran each chain for an additional 30,000 iterations. We then compared results based on the two chains (e.g., posterior means and 95% intervals) and found them to be extremely similar in each of the analyses presented below. To help ensure results with high degrees of accuracy in each analysis, we pooled the two chains. Thus all results are based on a sample size of 60,000 deviates. Using a Pentium IV 2.5 MHz machine, approximately 10 minutes of CPU time were required to complete 30,000 iterations of the algorithms for our LVR-HM3 analyses; considerably less time was required in the case of our LVR-HM2 analyses.

Heterogeneity in within-school initial status/rate of change slopes. To gauge the degree of heterogeneity in within-school initial status/rate of change slopes, we now fit the two-level LVR model defined by Equations 1, 3a and 3b to each school's data using WinBUGS1.4. In these analyses we placed diffuse normal priors on the fixed effects and diffuse Pareto priors on the precision parameters $1/\sigma^2$, $1/\tau_{00}$ and $1/\tau_{11}$ (i.e., *Pareto*(1, .001)). Note that at level 1 (Equation 1), $a_{ii} = (GRADE_{ii} - 7)$ where $GRADE_{ii}$ takes on values from 7 – 10. As such, π_{0i} represents the status of student i at the start of grade 7.

Figure 1 shows the resulting posterior mean and 95% interval for each school's initial status/rate of change slope (b). Note that the horizontal reference line corresponds to the average of the posterior means for the 45 schools in our sample (i.e., 0.11).

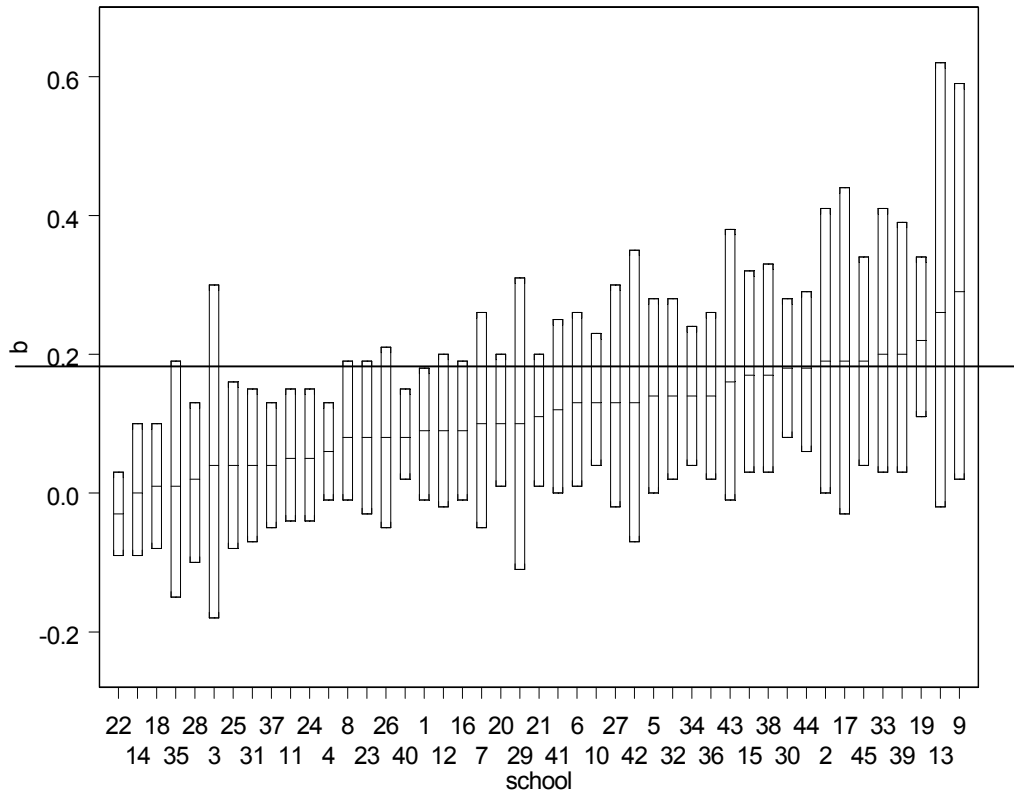


Figure 1. School-by-school two-level LVR analyses: The marginal posterior mean and 95% interval for each school's initial status/rate of change slope (b). The horizontal line represents the average of the initial status/rate of change slopes for the 45 schools in our sample. The top, middle, and bottom lines of each bar correspond, respectively, to the .975 quantile, the mean, and the .025 quantile of the marginal posterior distribution of the initial status/rate of change slope for a given school.

The range of the posterior means is considerable; the smallest value is -0.03 (School 22) and the largest is approximately 0.26 (School 13). Note further that the 95% posterior intervals for 4 schools (i.e., approximately 9% of the sample) lie above or below the reference line (i.e., Schools 14, 18, 19 and 22). As can also be seen, the posterior intervals for 25 schools contain a value of 0, whereas the posterior intervals for the remaining 20 lie above a value of 0.

Note that we also obtained an estimate of b for each school and its standard error using HLM 5 (Raudenbush et al., 2000). The resulting point estimates and 95% intervals for b (i.e., $\hat{b} \pm 1.96 \text{ SE}(\hat{b})$), which are based on the ML estimation strategy outlined in Raudenbush and Sampson (1999), are very similar to those obtained via WinBUGS1.4. For example, while the 95% intervals based on the Raudenbush and Sampson approach are slightly narrower than the fully Bayesian intervals shown in

Figure 1, the same four schools (i.e., 14, 18, 19 and 22) emerge as the ones whose intervals lie above or below the reference value of 0.11.

Results for Model 1

As can be seen in Table 2, the posterior mean for the grand mean of initial status (γ_{000}) is 49.72 points, and the posterior mean of the grand mean rate (γ_{100}) is 3.85 points per grade. Note that the posterior mean for the mean initial status/rate of change slope (Bw_0) is 0.101, which is similar to the value upon which the reference line in Figure 1 is based. Furthermore, we see that the lower and upper boundaries of the resulting 95% interval are 0.077 and 0.126, respectively.

Table 2

Model 1: Estimating Within-School Initial Status/Rate of Change Slopes (Bw_j) and a Between-School Initial Status/Rate of Change Slope (Bb)

	Mean	SD	95% Interval	Median	Prop. > 0
Fixed effects					
Model for school mean initial status (β_{00j})					
Grand mean (γ_{000})	49.72	0.63	(48.47, 50.96)	49.72	1.0000
Model for school mean rate of change (β_{10j})					
Grand mean (γ_{100})	3.85	0.15	(3.56, 4.14)	3.85	1.0000
School mean init. status (Bb)	0.076	0.038	(0.001, 0.152)	0.076	.9759
Model for within-school init. status/Rate of change slope (Bw _j)					
Mean init. status/Rate of change slope (Bw_0)	0.101	0.012	(0.077, 0.126)	0.100	1.0000
School mean init. status (Bw_1)	-.006	0.003	(-0.013, 0.000)	-0.006	.0275
Variance components					
Level-1 error (σ^2)	16.69	0.37	(15.97, 17.43)	16.68	
Level-3 variance					
Initial status ($\tau_{\beta_{00}}$)	16.35	4.083	(10.060, 25.890)	15.780	
Rate of change ($\tau_{\beta_{10}}$)	0.665	0.191	(0.370, 1.109)	0.638	
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)	-.017	0.010	(-0.040, 0.001)	-0.016	

Note. Proportion of the values generated for a given parameter via the Gibbs Sampler that lie above a value of 0. These proportions provide estimates of the marginal posterior probability that a given parameter exceeds a value of 0.

The results for Bb indicate a positive relationship between school mean initial status and school mean rates of change, whereas the results for Bw_1 point to a negative relationship between school mean initial status and within-school initial status/rate of change slopes. Note that the upper boundary of the 95% interval for Bw_1 just barely includes a value of 0. However, the posterior probability that Bw_1 exceeds a value of 0 is extremely small (0.028), and a 90% marginal posterior interval for Bw_1 (i.e., -0.012, -0.001) spans only negative values (see Figures 2 through 6 for the density plots of posteriors for γ_{000} , γ_{100} , Bb, Bw_0 and Bw_1).

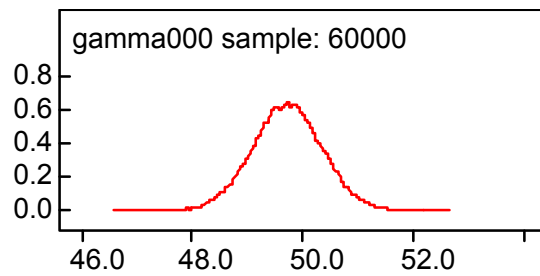


Figure 2. Density plot for γ_{000} .

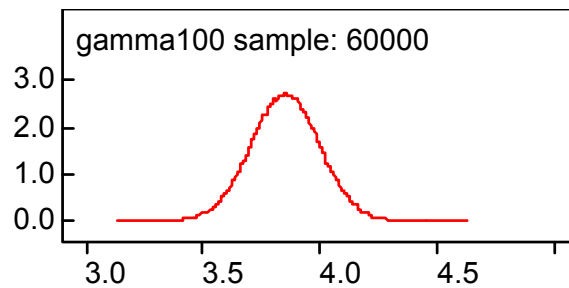


Figure 3. Density plot for γ_{100} .

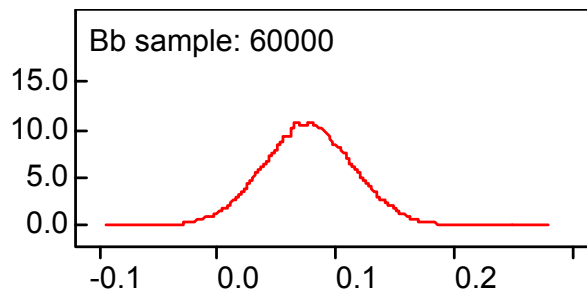


Figure 4. Density plot for Bb.

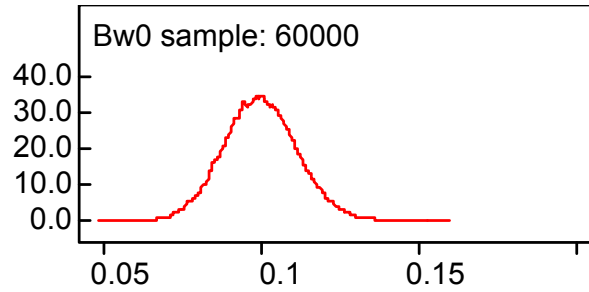


Figure 5. Density plot for Bw_0.

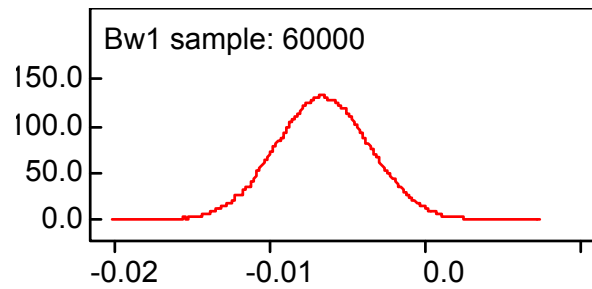


Figure 6. Density plot for Bw_1.

We now consider the implications of these results from a practical standpoint and what they might disclose about patterns of change within and between schools. Note that in Table 2, the posterior mean of the variance in school mean initial status ($\tau_{\beta_{00}}$) is equal to a value of 16.35. Though not shown in Table 2, the posterior mean of the standard deviation in school mean initial status ($\sqrt{\tau_{\beta_{00}}}$) is approximately 4 points. As noted above, the posterior mean of the grand mean for initial status is 49.72 points. In Figure 7 we display the expected trajectories for three schools with mean initial status values that are, respectively, 8 points (i.e., approximately 2 *SDs*) above a value of 49.72 (School A), equal to a value of 49.72 (School B), and 8 points below a value of 49.72 (School C). The expected trajectories for these schools are labeled with a darkened box.

We computed expected growth rates for students in these schools based on the posterior means for the grand mean rate (3.85) and for Bb (0.076). Thus, the expected growth rate for students in school A is 4.47 points per grade (i.e., $3.85 + 0.076 \cdot (8)$), the expected rate for school B is 3.85, and the expected rate for school C is 3.25. In Figure 7 we see that expected rates decrease as school mean initial status decreases. However, even for schools whose initial status values differ substantially (e.g., Schools A and C), the difference in expected rates is fairly modest. The expected

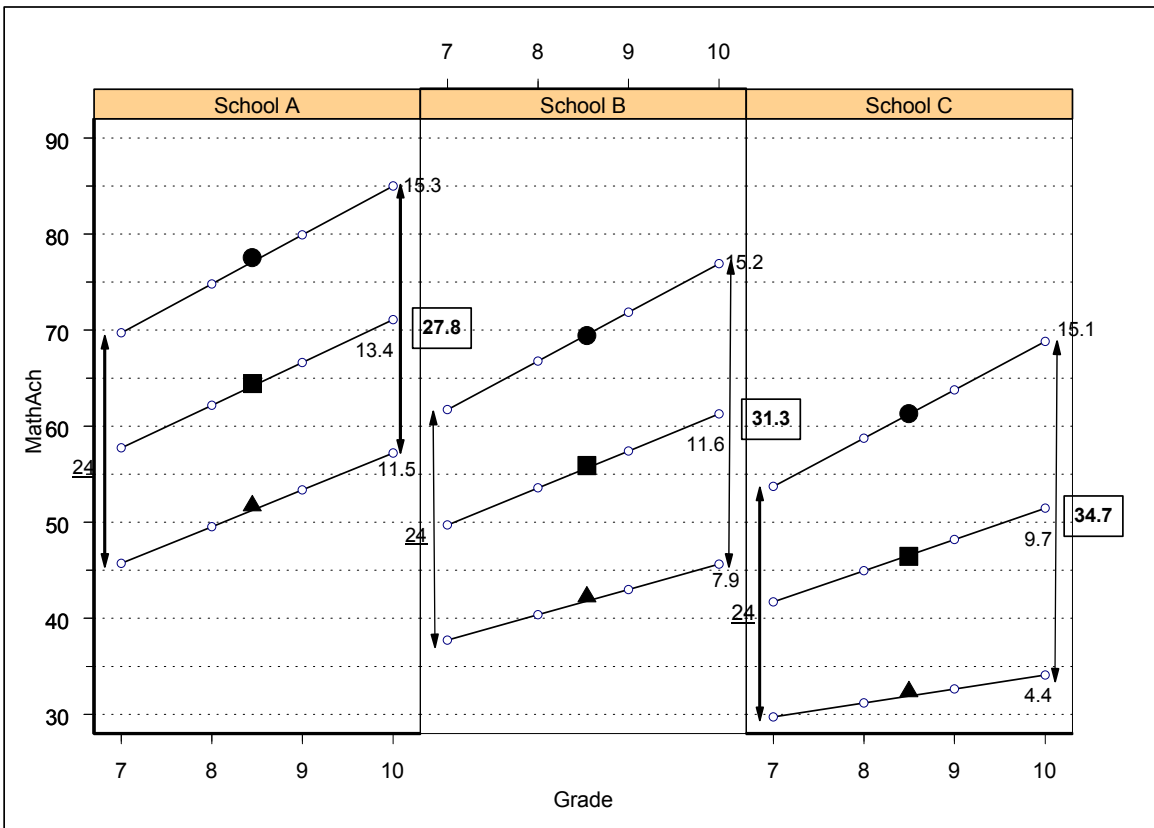


Figure 7. Expected growth trajectories for three schools with mean initial status values that are, respectively, 8 points (i.e., approximately 2 SD) above a value of 49.72 (School A), equal to a value of 49.72 (School B), and 8 points below a value of 49.92 (School C). The numbers attached to each line 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 number in the box denotes the expected difference at Grade 10 between a student who starts 12 points above a given school's mean initial status value and a student who starts 12 points below the mean initial status value.

gain in achievement for a student in School A over the course of 3 years is $3 \times 4.47 = 13.4$ points, whereas the expected gain in achievement for a student in School C over the course of 3 years is $3 \times 3.25 = 9.7$ points.

Turning to the distribution of achievement within schools, we computed expected initial status/rate of change slopes based on the posterior means for Bw_0 (0.101) and Bw_1 (-0.006). The expected initial status/rate of change slope for School A is 0.053 (i.e., $0.101 + (-0.006) \times (8)$), and the expected slopes for Schools B and C are 0.101 and 0.149, respectively. This indicates that as school mean initial status decreases, differences in initial status within schools appear to be more consequential with respect to subsequent achievement.

We now consider two students in School A whose initial status values are, respectively, 12 points above and below the mean initial status value for School A. Based on the expected rate and expected initial status/rate of change slope for this school, the initial gap of 24 points between two such students is expected to increase slightly by Grade 10 to a value of 27.8 points (see Figure 7). In School C, however, we see that the initial gap of 24 points for students with initial status values 12 points above and below the mean initial status value for school C, is expected to increase by more than 10 points by Grade 10 (i.e., 34.7 points).

It is instructive to examine the increases in the initial gaps in achievement for these schools more closely. We first consider expected gains in achievement by Grade 10 for students with initial status values that are 12 points above the mean initial status values of their respective schools. That is, we first focus on students with initial levels of achievement that are high in relation to the other students in their respective schools. Based on the expected rates and expected initial status/rate of change slopes for these schools, the expected gains for such students are approximately 15 points in all three schools (see Figure 7).

The results are quite different, however, when we consider the expected gains in achievement for students with initial status values that are 12 points below the mean initial status values of their respective schools. In the case of School A, the expected gain is substantial (11.5 points), but for School C, we see that the expected gain is rather minimal (i.e., 4.4 points) (see Figure 7). Thus, the differences in the expected gains for such students would appear to underlie the differences that we see in the extent to which initial gaps in achievement become magnified over time in these schools.

Results for Model 2

We now expand the level-2 (within-school) model as follows:

$$\begin{aligned}
 \pi_{0ij} &= \beta_{00j} + \beta_{01j}(HOMERES_{ij} - \overline{HOMERES..}) + \beta_{02j}(ED_EXPEC_{ij} - \overline{ED_EXPEC..}) + \\
 &\quad \beta_{03j}(BEHAV_PBLMS_{ij} - \overline{BEHAV_PBLMS..}) + r_{0ij} \quad r_{0ij} \sim N(0, \tau_{\pi_{0j}}) \\
 \pi_{1ij} &= \beta_{10j} + Bw_j(\pi_{0ij} - \beta_{00j}) + \beta_{11j}(HOMERES_{ij} - \overline{HOMERES..}) + \\
 &\quad \beta_{12j}(ED_EXPEC_{ij} - \overline{ED_EXPEC..}) + \beta_{13j}(BEHAV_PBLMS_{ij} - \overline{BEHAV_PBLMS..}) + r_{1ij} \\
 &\quad r_{1ij} \sim N(0, \tau_{\pi_{1j}}) \tag{15}
 \end{aligned}$$

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

where $HOMERES_{ij}$, ED_EXPEC_{ij} and $BEHAV_PBLMS_{ij}$ are student-level measures of home resources, educational expectations and behavioral problems, respectively (see Endnote 2). Note importantly that Bw_j is now the initial status/rate of change slope for school j holding constant the student characteristics that have been added to the model.

As can be seen in Equation 15, $HOMERES_{ij}$, ED_EXPEC_{ij} and $BEHAV_PBLMS_{ij}$ have been centered around their grand means. Following the terminology used in Raudenbush and Bryk (2002), β_{10j} and β_{00j} , by virtue of this centering, represent adjusted means for rate of change and initial status for school j . For example, if the relationship between student educational expectations (ED_EXPEC_{ij}) and student rate of change (π_{1ij}) is positive, and if the educational expectations of the students in school j are on average lower than the grand mean, the mean rate for school j would be adjusted upwards. Raudenbush and Bryk pointed out that such adjustments are analogous to the computation of adjusted treatment group means in ANCOVA analyses (2002, pp. 32-33). Note that such adjustments for the effects of student-level characteristics can facilitate, for example, the search for school practices and policies and other characteristics that may be related to rates of change.

The level-3 (between-school) model for Model 2 is as follows:

$$\begin{aligned}
\beta_{00j} &= \gamma_{000} + \gamma_{001}(\overline{HOMERES}_{.j} - \overline{HOMERES}_{..}) + u_{00j} & u_{00j} &\sim N(0, \tau_{\beta_{00}}) \\
\beta_{01j} &= \gamma_{010} \\
\beta_{02j} &= \gamma_{020} \\
\beta_{03j} &= \gamma_{030} \\
\beta_{10j} &= \gamma_{100} + Bb^*(\beta_{00j} - \gamma_{000}) + \gamma_{101}(\overline{HOMERES}_{.j} - \overline{HOMERES}_{..}) + \\
&\quad \gamma_{102}(\overline{TCHEFF}_{.j} - \overline{TCHEFF}_{..}) + u_{10j} & u_{10j} &\sim N(0, \tau_{\beta_{10}}) \\
\beta_{11j} &= \gamma_{110} \\
\beta_{12j} &= \gamma_{120} \\
\beta_{13j} &= \gamma_{130} \\
Bw_j &= Bw_0 + Bw_1(\beta_{00j} - \gamma_{000}) + Bw_2(\overline{HOMERES}_{.j} - \overline{HOMERES}_{..}) + \\
&\quad Bw_3(\overline{TCHEFF}_{.j} - \overline{TCHEFF}_{..}) + u_{Bwj} & u_{Bwj} &\sim N(0, \tau_{Bw}) \quad (16) \\
Cov(u_{00j}, u_{10j}) &= 0, Cov(u_{00j}, u_{Bwj}) = 0, Cov(u_{10j}, u_{Bwj}) = \tau_{\beta_{10}, Bw}
\end{aligned}$$

where $\overline{TCHEFF}_{.j}$ is a measure of teacher effort and concern for school j (see Endnote 3 for details), and $\overline{HOMERES}_{.j}$ is the average home resources measure for the sample of students in school j , thus constituting a measure of school j 's compositional characteristics. As can be seen, random effects are specified in the equations for β_{00j} , β_{10j} and Bw_j . The variance in the coefficients of the observed student predictors in the level-2 model is constrained to 0 (i.e., $\beta_{01j} = \gamma_{010}$; $\beta_{02j} = \gamma_{020}$; $\beta_{03j} = \gamma_{030}$; $\beta_{11j} = \gamma_{110}$; $\beta_{12j} = \gamma_{120}$; $\beta_{13j} = \gamma_{130}$).

We first focus on results concerning parameters of interest in the level-3 equation for Bw_j . Bw_j , as noted above, now represents the initial status/rate of change slope for students in school j holding constant the student-level predictors that we have included in the level-2 model, and Bw_0 represents the grand mean or population mean of Bw_j . As can be seen in Table 3, the posterior mean of Bw_0 is 0.08, which is somewhat smaller than the result we obtained when student background variables were not included at level 2 (i.e., 0.101; see Table 2). Note, however, that the lower boundary of the resulting 95% posterior interval (i.e., 0.058) is appreciably larger than a value of 0.

The results for Bw_2 signal that larger initial status/rate of change slopes are expected in schools with low mean home resource values. As $\overline{HOMERES}_{.j}$ decreases, initial gaps in achievement within a school are expected to become magnified over time. It is important to note that $\overline{HOMERES}_{.j}$ is likely a proxy for a number of factors that potentially underlie this relationship. In many schools with low $\overline{HOMERES}_{.j}$ values, appreciable numbers of students enter school with poor prior preparation. Such schools typically have limited resources and, in particular, may have insufficient resources for mounting the kinds of programs that could help promote more rapid growth among entering students with very low levels of initial achievement. Ability grouping (e.g., tracking) may also be much more prevalent in such schools. Probing these potential explanations would require conducting studies that attend carefully to within-school processes and practices.

We now turn to the results concerning key parameters in the model for school mean adjusted rates. As can be seen in Table 3, the posterior mean and resulting 95% interval for γ_{102} suggest that after adjusting for the student-level effects of $\overline{HOMERES}_{ij}$, $\overline{ED_EXPEC}_{ij}$ and $\overline{BEHAV_PBLMS}_{ij}$, and holding constant school mean initial status and $\overline{HOMERES}_{.j}$, a 1-unit increase in teacher effort implies an increase in school mean rate of change of approximately two-thirds of a point per grade.

Table 3

Model 2: Including Observed Student- and School-Level Predictors in the 3-Level LVR-HM

	Estimate	SE	95% Interval	Median	Prop. > 0
Fixes effects					
Model for school mean initial status (β_{00j})					
Grand mean (γ_{000})	50.02	0.45	(49.14, 50.92)	50.02	1.0000
School mean home resources (γ_{001})	6.45	1.56	(3.39, 9.51)	6.45	.9999
Model for school mean rate of change (β_{10j})					
Grand mean (γ_{100})	3.905	0.115	(3.677, 4.131)	3.905	1.0000
School mean init. status (Bb)	-.022	0.049	(-0.120, 0.074)	-0.022	.3250
School mean home resources (γ_{101})	1.191	0.504	(0.207, 2.195)	1.189	.9906
Teacher effort (γ_{102})	0.657	0.207	(0.251, 1.064)	0.655	.9990
Model for within-school init. status/Rate of change slope (Bw_j)					
Mean of init. status/Rate of change slope (Bw_0)	0.080	0.011	(0.058, 0.103)	0.080	1.0000
School mean initial status (Bw_1)	0.002	0.005	(-0.007, 0.011)	0.002	.6491
School mean home resources (Bw_2)	-.095	0.043	(-0.180, -0.010)	-0.095	.0143
Teacher effort (Bw_3)	-.016	0.018	(-0.051, 0.020)	-0.016	.1875
Effects of student-level characteristics on init. status (π_{0ij})					
Home resources (γ_{010})	0.998	0.157	(0.690, 1.305)	0.998	1.0000
ED_EXPEC(γ_{020})	1.773	0.128	(1.522, 2.026)	1.772	1.0000
BEHAV_PBLMS(γ_{030})	3.224	0.480	(-4.165, -2.283)	-3.224	.0000
Effects of student-level characteristics on rate of change (π_{1ij})					
Home resources (γ_{110})	0.063	0.060	(-0.055, 0.180)	0.063	.8521
ED_EXPEC (γ_{120})	0.183	0.053	(0.076, 0.288)	0.183	.9998
BEHAV_PBLMS (γ_{130})	-.225	0.202	(-0.621, 0.168)	-0.224	.1348
Variance components					
Level-1 error (σ^2)	16.67	0.38	(15.95, 17.43)	16.67	
Level-3 variance					
Initial status($\tau_{\beta_{00}}$)	7.768	2.144	(4.501, 12.84)	7.465	
Rate of change($\tau_{\beta_{10}}$)	0.376	0.124	(0.191, 0.668)	0.357	
Initial/rate of change slopes (τ_{Bw})	0.001	0.000	(0.000, 0.002)	0.001	
Cov. (School rate of change, initial/Rate of change slopes)	-0.004	0.006	(-0.017, 0.007)	-0.004	

(Note that a 1-unit increase in teacher effort represents an increase of slightly less than 2 standard deviations; see Table 1).

By virtue of the fact that we are adjusting for the student-level effects of $HOMERES_{ij}$ (via the use of grand-mean centering at level 2), γ_{101} represents the contextual effect of home resources. (See Raudenbush and Bryk, 2002, for a detailed discussion of the estimation and interpretation of contextual effects, especially pp. 135–141.) Thus, for example, consider two students who are similar in terms of $HOMERES_{ij}$ but who attend schools that differ by one unit with respect to $\overline{HOMERES}_{.j}$. γ_{101} represents the expected difference in rate of change for two such students, holding constant all other predictors in the model. The results suggest a contextual effect of home resources on a rate of approximately 1.2 points per grade. (As can be seen in Table 1, a 1-unit increase in $\overline{HOMERES}_{.j}$ represents an increase of nearly 3 standard deviations.) Note that a number of factors potentially underlie contextual effects, including differences in the quality of instructional programs, in the normative climate of schools, and in the quality of help that peers are able to provide one another with school work.

In contrast to Model 1, we see that school mean initial status, after the inclusion of the observed student- and school-level predictors in our model, is no longer a significant predictor of school mean rate of change and within-school initial status/rate of change slopes. Finally, comparing the posterior means of the level-3 variance components with those in Model 1, we see that the inclusion of the observed predictors in Model 2 results in reductions in the random effects variance of β_{00j} , β_{10j} , and Bw_j of 52.5%, 43.4%, and 50.0%, respectively.

Re-Fitting Model 2 Under t Distributional Assumptions

In contrast to conducting analyses under normality assumptions, parameter estimates and intervals obtained under heavy-tailed distributional assumptions are less vulnerable to outliers. To gauge the sensitivity of our results to possible outliers (e.g., outlying time-series observations, students or schools), we re-fit Model 2 employing t distributional assumptions with small degrees of freedom ($\nu = 4$) at each level: $\varepsilon_{ij} \sim t_4(0, \sigma^2)$; $r_{0ij} \sim t_4(0, \tau\pi_{0j})$, $r_{1ij} \sim t_4(0, \tau\pi_{1j})$; $u_{00j} \sim t_4(0, \tau\beta_{00})$, $(u_{10j}, u_{Bwj}) \sim MVT4(0, T2,3)$ (see Endnote 4). Note that fitting HMs under t distributional assumptions can be carried out via the Gibbs sampler using the scale mixture of normals representation of the t (e.g., $\varepsilon_{ij} \sim N(0, \sigma^2 / \omega_{ij})$, $\omega_{ij} \sim \chi_{\nu}^{-2} / \nu$) (see e.g., Carlin, 1992;

Racine-Poon, 1992; Seltzer, 1993; Seltzer et al., 2002; Spiegelhater, Best, Gilk, & Inskip, 1996). This strategy provides the basis of t modeling in WinBUGS1.4.

We found that the resulting posterior means and 95% intervals for the parameters in Model 2 based on t assumptions tended to be extremely close to the corresponding posterior means and intervals based on normality assumptions. The largest change occurred for the fixed effect capturing the relationship between $\overline{HOMERES}_j$ and Bwj (i.e., Bw_2). Specifically, there is a small upward shift in the posterior mean (-.080) and 95% interval (-.165, .002) for Bw_2 (versus a posterior mean of -.095 and an interval of (-.177, -.014) under normality). Thus the evidence of a negative relationship between $\overline{HOMERES}_j$ and Bwj is slightly weaker under heavy-tailed assumptions. Note, however, that the marginal posterior probability that Bw_2 < 0 is still very high under t assumptions, that is, 0.975 (versus 0.989 under normality assumptions).

Improving the Performance of the Gibbs Sampler: The Use of Different Centerings in the LVR-HM3

In implementing the Gibbs sampler in HM settings, we need to be alert to the possible occurrence of poor mixing, that is, situations in which successive values in the chains generated for one or more parameters in a given model are highly autocorrelated. When mixing is 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 in a given set of M iterations.

One strategy to reduce high posterior correlations and improve mixing is to center the covariates in one's model (see Gilks et al., 1996, chapter 6; Spiegelhalter et al., 2003). In this section we focus on the importance of centering latent variable predictors. Thus in the case of our models, one option would be to center student initial status around its school mean (i.e., $\pi_{0ij} - \beta_{00j}$) and to center school mean initial status around the grand mean (i.e., $\beta_{00j} - \gamma_{000}$). We explored the effects of using various centerings on the performance of the Gibbs sampler in estimating Model 2.

No centering of latent variable predictors at levels 2 and 3

$\pi_{0ij} - \beta_{00j}$ (level-2 centering)

$\beta_{00j} - \gamma_{000}$ (level-3 centering)

$\pi_{0ij} - \beta_{00j} ; \beta_{00j} - \gamma_{000}$ (centering at levels 2 and 3).

Note that the performance of the Gibbs sampler for each of the above conditions was monitored by means of computing autocorrelations among the deviates in a chain. Plots of a series of autocorrelations (i.e., autocorrelation functions [ACF] plots) provide an important tool for assessing mixing.

In Table 4 for each condition, we categorize parameters based on whether their ACF values are smaller than .10 at a series of different lag (k) 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 the Performance of the Gibbs Sampler Employing Different Centerings in Model 2: Parameters With Autocorrelation Function (AFC) Values Below .10 at Lag = k ($\rho(k) < .10$)

	No centering	Level-2 centering	Level-3 centering	Centering at levels 2 and 3
$\rho(k) < .10$ $k = 5$	$\gamma_{000}, \gamma_{001}, \gamma_{010}, \gamma_{020},$ $\gamma_{030}, \sigma^2, \tau_{\beta 00}$	$\gamma_{000}, \gamma_{001}, \gamma_{010}, \gamma_{020}, \gamma_{030},$ $\gamma_{102}, \sigma^2, \tau_{\beta 00}$	$\gamma_{001}, \gamma_{010}, \gamma_{020}, \gamma_{030}, \sigma^2,$ $\tau_{\beta 00}$	$\gamma_{000}, \gamma_{001}, \gamma_{010}, \gamma_{020}, \gamma_{030},$ $\tau_{\beta 00}$
$\rho(k) < .10$ $k = 10$		$\gamma_{110}, \tau_{\beta 10}$	γ_{000}	$\gamma_{102}, \gamma_{100}, \tau_{\beta 10}$
$\rho(k) < .10$ $k = 50$	$\gamma_{110}, \gamma_{120}, \gamma_{130},$ τ_{Bw}	Bw_3, $\gamma_{120}, \gamma_{130}, \tau_{\beta 10}$	$\gamma_{110}, \gamma_{120}, \gamma_{130}, \tau_{Bw}$	Bb, Bw_0, Bw_1, Bw_2, Bw_3, $\gamma_{101},$ $\gamma_{110}, \gamma_{120}, \gamma_{130}, \sigma^2, \tau_{Bw}$
$\rho(k) < .10$ $k = 100$	Bw_2, Bw_3, $\gamma_{101}, \gamma_{102}, \tau_{\beta 10}$		Bb, Bw_0, Bw_1, Bw_2, Bw_3, $\gamma_{100},$ $\gamma_{101}, \gamma_{102}, \tau_{\beta 10}$	
$\rho(k) < .10$ $k = 150$				
$\rho(k) < .10$ $k = 200$				
$\rho(k) < .10$ $k = 250$				
$\rho(k) > .50$ $k = 300$	Bb, Bw_0, Bw_1, γ_{100}	Bb, Bw_0, Bw_1, Bw_2, $\gamma_{100}, \gamma_{101}$		

First, for condition 1, (i.e., no centering at levels 2 or 3), ACF values for γ_{000} , γ_{001} , γ_{010} , γ_{020} , γ_{030} , γ_{020} , σ^2 , and $\tau_{\beta_{00}}$ are smaller than .10 at $k = 5$, while the ACF values for Bb, Bw_0, Bw_1, and γ_{100} are very high even at $k = 300$. Note that γ_{100} and Bb are fixed effects in the level-3 equation for β_{10j} (school rates of change) and Bw_0 and Bw_1 are fixed effects in the level-3 equations for Bw_j. The ACF values at $k = 300$ for those four parameters are .59, .63, .63, and .62, respectively. Thus for γ_{000} , for example, the autocorrelations are decreasing very quickly and are close to 0 before $k = 5$ (see Figure 8). In contrast, for γ_{100} , the autocorrelations hardly decrease as k increases (see Figure 9).

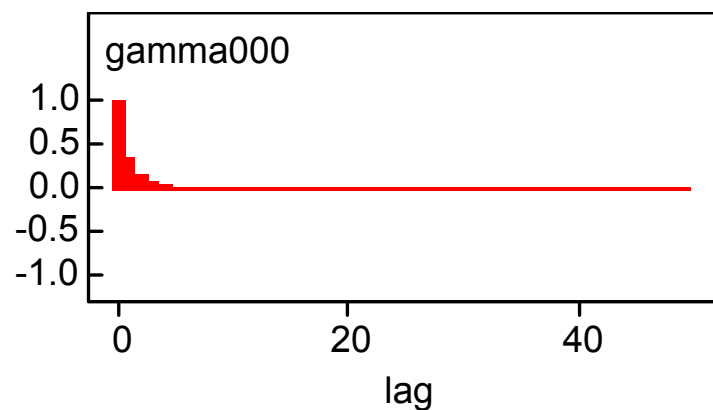


Figure 8. Plot for γ_{000} : An example of good mixing.

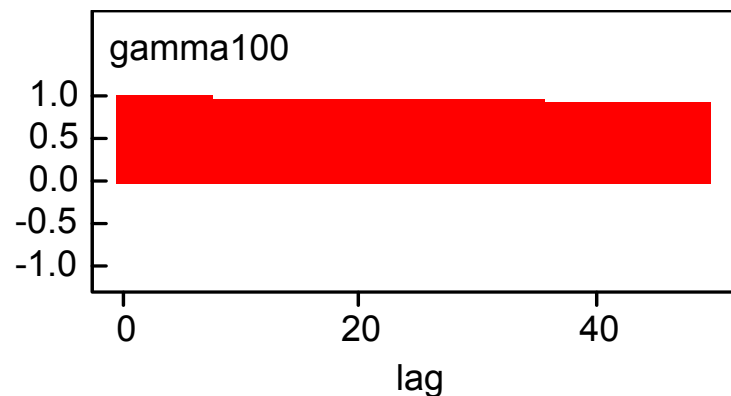


Figure 9. Plot for γ_{100} : An example of poor mixing.

With centering at level 2, we see that there is still a set of fixed effects in the level-3 equations for β_{10j} and Bw_j for which mixing is extremely poor. With centering at level 3, autocorrelations for all parameters are less than .10 by lag 100. Finally, we see a further improvement in mixing for fixed effects in the level-3 equations for β_{10j} and Bw_j when centering is employed at levels 2 and 3. As can be seen, $\rho < .10$ by $k = 50$. We have found that this general patterns tends to hold in fitting numerous LVR-HM3s to the LSAY data (Choi, 2002).

Simulation Study

We conducted a small-scale simulation study to examine issues of coverage and bias connected with our fully Bayesian approach to estimating LVR-HM3s. In particular, we focused on the actual levels of coverage of nominal 95% posterior intervals for the fixed effects in the LVR-HM3, and on the degree of bias of the marginal posterior means for the fixed effects, employing diffuse priors for the fixed effects and variance components as outlined above (see also Endnote 1).

The design of this simulation study is based on Model 1. Specifically, we generated 300 data sets using our results based on Model 1 (i.e., the posterior means in Table 2) as true values for the parameters in the model. Note that the posterior means of $\tau_{\pi_{0j}}$ and $\tau_{\pi_{1j}}$ ($j = 1, \dots, J$) were used as true values for the level-2 variance components. Mirroring the structure of the LSAY sample, each data set consists of 8,585 simulated time series observations nested within 2,628 students, who in turn are nested within 45 schools. We fit Model 1 to each simulated data set using WinBUGS1.4. For each data set we employed a burn-in period of 2,000 iterations and then ran the Gibbs sampler for an additional 80,000 iterations. Thus the posterior mean and (nominal) 95% posterior interval for each fixed effect (i.e., γ_{000} , γ_{100} , Bw_0 , Bw_1 , Bb) were based on a sample of 80,000 deviates.

From the 300 95% posterior intervals obtained for a given fixed effect (e.g., Bw_1) we calculated the percentage of intervals that captured the corresponding true value used to generate the data. We also used the 300 posterior means obtained for a given fixed effect to compute the degree of bias and relative bias.

Table 5 summarizes the simulation results. We can see that the actual levels of coverage for nominal 95% fully Bayesian intervals for all of the fixed effect parameters are very close to 95%. In addition, the resulting marginal posterior means are on average very close to the corresponding true values. Note that the bias and relative bias of the posterior means for γ_{000} , γ_{100} , Bw_1 , and Bb are negligible. As

Table 5

Summary of the Simulation Study: Posterior Means, Bias, Relative Bias, and Actual Levels of Coverage of Nominal 95% Intervals

Parameter	True value (θ)	Average posterior mean(s)	Bias (s- θ)	Relative bias (s- θ)/(θ)	Actual coverage of nominal 95% intervals
γ_{000}	49.72	49.709	-.0011	-.000	.948
γ_{100}	3.86	3.857	-.0028	-.001	.957
Bw_0	.089	.084	-.005	-.056	.943
Bw_1	-.005	-.005	.000024	.005	.950
Bb	.076	.0763	.0003	.004	.953

can be seen, the resulting posterior means of Bw_0 are on average slightly negatively biased (.084 vs. .089), with a relative bias of 5.6%. Though this is a small-scale simulation, the results suggest that our fully Bayesian approach to estimating LVR-HM3s is extremely promising with respect to coverage and bias, especially when we consider that the number of schools in our simulated data sets and the number of students per school were moderate in size.

Discussion

There are many research settings, particularly in education, in which individuals are observed on only two occasions (e.g., at the start and end of a school year), and interest centers on individual change (e.g., learning) during that span of time. When standard errors of measurement are available, it is possible to implement LVR-HM3s in such situations. For example, let Y_{1ij} and Y_{2ij} represent, respectively, the math achievement scores at the start and end of the school year for student i in school j , with standard errors of measurement $SE(Y_{1ij})$ and $SE(Y_{2ij})$. We pose a level-1 model of the following form:

$$\begin{pmatrix} Y_{1i} \\ Y_{2i} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \pi_{0i} \\ \pi_{1i} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \quad (17)$$

and scale the left- and right-hand sides of Equation 17 by the inverse of the standard errors of measurement for student i such that $\varepsilon_{1i} \sim N(0, 1)$ and $\varepsilon_{2i} \sim N(0, 1)$. In a level-2 (e.g., within-school) model, we can then model the latent gains (π_{1i}) for the students in school j as a function of initial status (π_{0i}). Initial status/latent gain

slopes, along with school mean gains and initial status, can then be treated as varying across schools in a level-3 model.

In settings in which individuals are observed on four or more occasions and individual growth exhibits curvature, several options are available. First, in an approach that can be viewed as an extension of the above strategy for two time-point settings, one could fit a quadratic model to each student's time-series data via OLS, and, based on an individual's fitted trajectory, obtain estimates of initial status (\hat{Y}_{1ij}) and final status (\hat{Y}_{2ij}). Note that residuals from the fitted trajectories ($\hat{Y}_{tij} - [(\hat{\pi}_{0ij} - \hat{\pi}_{1ij} a_{tij} + \hat{\pi}_{2ij} a_{tij}^2)]$) could be used to obtain a pooled estimate of within-person variance ($\hat{\sigma}^2$), which, in turn, could be used in computing standard errors for \hat{Y}_{1ij} and \hat{Y}_{2ij} ($SE(\hat{Y}_{1ij}), SE(\hat{Y}_{2ij})$). Replacing $(Y_{1i} \ Y_{2i})'$ with $(\hat{Y}_{1ij} \ \hat{Y}_{2ij})'$ in Equation 17 and scaling the left- and right-hand sides by the inverse of the corresponding standard errors, we can then model latent gains as a function of initial status and pose level-2 and level-3 models similar to those illustrated above.

Another possibility, which requires further research, entails posing a quadratic model for individual growth at level 1, employing initial status (π_{0ij}) as a predictor of initial rate (π_{1ij}) and acceleration (π_{2ij}) at level 2, and treating initial status/initial rate slopes and initial status/acceleration slopes as varying across schools. A possible strategy for studying the relationship between initial status and instantaneous rates at later points in time is briefly discussed in Endnote 5.

As discussed at the outset, LVR-HM3s have valuable applications in longitudinal multi-site studies of programs. Of particular interest are multi-site studies in which data are collected at several points in time prior to the start of a treatment, and then at several points during the treatment phase. In such instances, one could pose a within-site model in which rates of change during the treatment phase are modeled as a function of final status or rates of change during the pre-treatment phase. It may be the case that individuals with extremely low final status values or with markedly slow rates of change in the pre-intervention phase make little subsequent progress in some sites, but substantial progress in others. Are such differences related to differences in certain aspects of program implementation across sites? Question of this kind could be addressed via the kinds of models discussed in this report.

In the case of cross-sectional studies, and two-time-point studies in which student pretest measures are employed as covariates, the heterogeneity in within-cluster slopes is often of substantive interest. In analyses of the High School and Beyond (HSB) data, for example, investigations have focused on the question of whether within-school SES/achievement slopes are flatter in Catholic schools than public schools (see, e.g., Raudenbush & Bryk, 1986). In studies of instruction, researchers have examined whether class mean achievement is higher and pretest/posttest slopes flatter when teachers employ particular instructional practices (see, e.g., Cronbach, 1976; Seltzer, 1995). When standard errors of measurement are available, we can, in the LVR-HM3 framework, employ a measurement model at level 1, pose within-cluster models in which latent variable outcomes are modeled as a function of latent variable predictors (level 2), and treat the corresponding LVR coefficients as varying across clusters (level 3). This strategy can help overcome possible attenuation problems connected with employing observed predictors at level 2. In summary, the LVR-HM3 framework that we have presented would appear to be of potential value in a variety of research settings.

Endnotes

1. In the case of each of our LVR-HM3s, we first treated u_{10j} and u_{Bwj} as univariate normally distributed (i.e., $u_{10j} \sim N(0, \tau_{\beta_{10}})$; $u_{Bwj} \sim N(0, \tau_{Bw})$; $\tau_{\beta_{10}, Bw} = 0$), and placed *Pareto* priors on $\tau_{\beta_{10}}^{-1}$ and τ_{Bw}^{-1} . (Note that re-fitting hierarchical models with random effects covariances set to 0 provides a means of checking for possible model misspecification; see Raudenbush & Bryk (2002, p. 272). In each of these analyses we found the modes of $p(\tau_{\beta_{10}} | y)$ and $p(\tau_{Bw} | y)$, which we denote $\hat{\tau}_{\beta_{10}}$ and $\hat{\tau}_{Bw}$.

From Zellner (1971), note that if *a priori* $\mathbf{T}_{2,3} \sim W^{-1}(3, \mathbf{S})$, then *a priori* $\tau_{\beta_{10}}$ and τ_{Bw} are inverse χ^2 distributed with 2 degrees of freedom: $\tau_{\beta_{10}} \sim \chi^2(v_{IC} = 2, S_{11})$; $\tau_{Bw} \sim \chi^2(v_{IC} = 2, S_{22})$, where S_{11} and S_{22} are the diagonal elements of \mathbf{S} . From the properties of the inverse χ^2 distribution, the modes of $p(\tau_{\beta_{10}})$ and $p(\tau_{Bw})$ are, respectively, $S_{11} / (v_{IC} + 2)$ and $S_{22} / (v_{IC} + 2)$. Based on a strategy detailed in Seltzer et al. (2002, pp. 214-218), we chose S_{11} and S_{22} so that the modes of $p(\tau_{\beta_{10}})$ and $p(\tau_{Bw})$ are equal to $\hat{\tau}_{\beta_{10}}$ and $\hat{\tau}_{Bw}$, respectively (i.e., $S_{11} = \hat{\tau}_{\beta_{10}} \times 4$; $S_{22} = \hat{\tau}_{Bw} \times 4$). Thus, in the case of Model 1:

$$T_{2,3} \sim W^{-1} \left[3, \begin{pmatrix} 2.57 & 0 \\ 0 & .005 \end{pmatrix} \right];$$

$$\text{for Model 2: } T_{2,3} \sim W^{-1} \left[3, \begin{pmatrix} 1.434 & 0 \\ 0 & .0036 \end{pmatrix} \right].$$

2. We have three observed student characteristics variables. First, the variable ED_EXPEC_{ij} captures a student's educational expectations at the start of Grade 7. This variable takes on 5 possible values: 1 = high school degree; 2 = vocational training; 3 = 2-year college; 4 = 4-year college; 5 = masters degree; 6 = doctorate or professional degree. Secondly, $BEHAV_PBLMS_{ij}$ is a measure of student behavioral problems obtained at the start of Grade 7. It takes on a value 1 if a student reported that he or she had ever been suspended, arrested by the police, or had considered dropping out (0 otherwise). Third, $HOMERES_{ij}$ is a measure of a student's home resources obtained at the start of Grade 7. It is the sum of a student's responses to questions concerning whether the 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. A response of "yes" to each question is coded 1 (0 otherwise). Thus possible values for this variable range from 0 to 4.

3. The teacher effort variable is based on averaging teachers' responses to the following two 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 effort 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.

4. Paralleling the strategy detailed in Endnote 1, we first treated u_{10j} and u_{Bwj} as univariate t distributed, and placed *Pareto* priors on $\tau_{\beta_{10}}^{-1}$ and τ_{Bw}^{-1} . The marginal posterior modes of $\tau_{\beta_{10}}$ and τ_{Bw} were then used to obtain values for S_{11} and S_{22} , as in the MVN case. Thus $T_{2,3} \sim W^{-1} \left[3, \begin{pmatrix} 1.025 & 0 \\ 0 & .0026 \end{pmatrix} \right]$.

5. Consider a quadratic model for individual growth:

$$Y_{tij} = \pi_{0ij} + \pi_{1ij} a_{tij} + \pi_{2ij} a_{tij}^2 + \varepsilon_{tij} \quad \varepsilon_{tij} \sim N(0, \sigma^2).$$

Suppose that each person is observed on 5 occasions, corresponding to a_{tij} values of 0, 1, 2, 3, 4. If we are interested in how differences in initial status relate to differences in change, one thing we could do is employ initial status as predictor of acceleration and initial rate:

$$\pi_{0ij} = \beta_{00j} + u_{0ij} \quad (a)$$

$$\pi_{1ij} = \beta_{10j} + B_{1wj} (\pi_{0ij} - \beta_{00j}) + u_{1ij} \quad (b)$$

$$\pi_{2ij} = \beta_{20j} + B_{2wj} (\pi_{0ij} - \beta_{00j}) + u_{2ij} \quad (c)$$

It can be shown that the instantaneous rate of change at a given point in time for person i in school j is : $\pi_{1ij} + 2\pi_{2ij} a_{tij}$. Thus, for example, student i 's rate at $a_{tij} = 2$ would be equal to $\pi_{1ij} + 4\pi_{2ij}$. Furthermore, if we were to modify Equation b as follows:

$$\pi_{1ij} = \beta_{10j} + B_{1wj} (\pi_{0ij} - \beta_{00j}) - 4\pi_{2ij} + u_{1ij}, \quad (d)$$

then B_{1wj} in Equation d would capture the relationship between initial status and rate at $a_{tij} = 2$ for students in school j . To help see this, note that if we were given π_{0ij} , π_{1ij} and π_{2ij} in a step of the Gibbs sampler, then Equation d could be re-written as:

$$\pi_{1ij} + 4\pi_{2ij} = \beta_{10j} + B_{1wj} (\pi_{0ij} - \beta_{00j}) + u_{1ij}. \quad (e)$$

In a personal communication, Spiegelhalter pointed out that an aspect of this approach that requires further investigation is the specification of the prior of the variance-covariance matrix for u_{1ij} and u_{2ij} .

Appendix

In our analyses, individual growth is modeled as a linear function of *GRADE*. Extensive preliminary analyses that we conducted pointed to this as being a fairly adequate representation of individual growth. Specifically, we first inspected displays of the time series data for the students in each school. Next, using the HLM 5 program (Raudenbush et al., 2000), we fit two different two-level HMs to each school's data. In the first HM (Model A), we specified a linear model for individual growth, and in the second we specified a quadratic model (Model B). Model A is identical to the HM specified in Equations 1, 2a and 2b, and Model B is as follows:

$$Y_{ii} = \pi_{0i} + \pi_{1i}a_{ii} + \pi_{2i} a_{ii}^2 + \varepsilon_{ii} \quad \varepsilon_{ii} \sim N(0, \sigma^2), \quad (\text{A1})$$

$$\pi_{0i} = \beta_{00} + r_{0i} \quad r_{0i} \sim N(0, \tau_{00}) \quad (\text{A2})$$

$$\pi_{1i} = \beta_{10} + r_{1i} \quad r_{1i} \sim N(0, \tau_{11}) \quad (\text{A3})$$

$$\pi_{2i} = \beta_{20} + r_{2i} \quad r_{2i} \sim N(0, \tau_{22}), \quad (\text{A4})$$

where, for the students in a given school, π_{0i} and π_{1i} represent, respectively, the status and rate of change for student i at the start of Grade 7 (i.e., *GRADE* = 0), and π_{2i} captures the acceleration or deceleration connected with student i 's trajectory (i.e., π_{2i} is the quadratic component for student i). At level 2, β_{00} , β_{10} and β_{20} represent, respectively, the mean initial status, mean initial rate and mean quadratic component for the students in a given school, and the variability among students in their initial status, initial rate and curvature is captured by τ_{00} , τ_{11} and τ_{22} . The level-2 model also includes the following covariance terms: $Cov(r_{0i}, r_{1i}) = \tau_{01} = \tau_{10}$; $Cov(r_{0i}, r_{2i}) = \tau_{02} = \tau_{20}$; and $Cov(r_{1i}, r_{2i}) = \tau_{12} = \tau_{21}$.

In fitting Models A and B to each school's data, we found that for 37 of 45 schools in our sample, the estimate of the fixed effect for the quadratic component ($\hat{\beta}_{20}$) was not statistically significant. Furthermore, for each of these 37 schools, a likelihood ratio test revealed no improvement in fit based on a model in which the quadratic component was allowed to vary across individuals versus a model in which the variance in the quadratic component was constrained to a value of 0 (i.e., $\tau_{22} = 0$).

Of the remaining 8 schools, $\hat{\beta}_{20}$ is statistically significant for schools S11, S28 and S43, and there is evidence that τ_{22} is non-zero in the case of schools S19, S37 and S45. For 2 schools, S10 and S12, there is evidence that both β_{20} and τ_{22} are non-zero.

For each of these schools, how different from a practical standpoint are the results regarding various key aspects of change based on the use of a linear model and a quadratic model for individual growth? To address this question, we will focus on differences in estimates of initial status under the two models, and differences in estimates of final status conditional on particular values of initial status.

Again conducting HM analyses school by school, for each student in a given school we computed an empirical Bayes (EB) estimate of initial status based on the HM for quadratic growth specified above, and an EB estimate of initial status based on the HM for linear growth specified in Equations 1, 2a and 2b; we term these estimates π_{0iQ}^* and π_{0iL}^* , respectively. We then computed the difference between these estimates: $DIFF_{0i} = \pi_{0iL}^* - \pi_{0iQ}^*$. Next, we computed the mean $DIFF_{0i}$ value for each of the 8 schools ($M(DIFF_{0i})$), as well as the SD of the differences ($SD(DIFF_{0i})$).

As can be seen in Table A1, the smallest mean difference is $-.05$ points (S19), and 7 of 8 schools have mean differences under 1 point. The largest mean difference is 1.59 points (S11). In terms of the standard deviations of the differences, the standard deviations for 5 of the schools range from 0.57 to 0.80; 2 schools (S12 and S28) have standard deviations of 1.09 and 1.21, respectively, and the largest standard deviation is 1.86 (S45).

Table A1

Comparing EB Estimates of Initial Status Based on Linear and Quadratic Models for Individual Growth

School #	$M(DIFF_{0i})$	$SD(DIFF_{0i})$	$\hat{\sigma}_L - \hat{\sigma}_Q$	$(\hat{\tau}_{00L})^{1/2} - (\hat{\tau}_{00Q})^{1/2}$
S10	-.72	.68	3.77 - 3.22 = 0.55	7.24 - 7.13 = 0.11
S11	-1.59	.79	4.38 - 3.56 = 0.82	8.83 - 8.63 = 0.2
S12	.51	1.09	4.20 - 3.62 = 0.58	9.40 - 10.22 = -0.82
S19	-.05	.61	3.91 - 3.10 = 0.81	7.16 - 6.93 = 0.23
S28	-.74	1.21	4.62 - 4.14 = 0.48	9.52 - 10.16 = -0.64
S37	.13	.80	3.21 - 2.63 = 0.58	7.16 - 7.53 = -0.37
S43	-.83	.57	4.28 - 4.06 = 0.22	5.80 - 5.51 = 0.29
S45	.15	1.86	4.51 - 3.82 = 0.69	7.05 - 8.18 = -1.13

To put these mean differences and SD ($DIFF_{oi}$) values in perspective, note that estimates of the level-1 (within-student) standard deviation parameters under these two models ($\hat{\sigma}_L; \hat{\sigma}_Q$) for the 8 schools range from 2.63 to 4.62, and estimates of the standard deviation in initial status under the two models ($(\hat{\tau}_{00L})^{1/2}; (\hat{\tau}_{00Q})^{1/2}$) range from 5.51 to 10.22. Against this backdrop, the differences between π_{oiQ}^* and π_{oiL}^* that we tend to see within each of the 8 schools are relatively small.

Comparing the estimates of the standard deviations in initial status under the two models (i.e., $(\hat{\tau}_{00L})^{1/2}; (\hat{\tau}_{00Q})^{1/2}$), we see that the estimates are extremely similar, with differences ranging between 0.11 and 1.13. In addition, comparing the estimates of σ obtained under the two models, we see that the differences are relatively small, ranging between values of 0.22 and 0.82.

We now focus on the relationship between initial status and subsequent change in these 8 schools. Specifically, we examine differences in estimates of final status conditional on particular values of initial status when we employ quadratic and linear models for individual growth. Employing the quadratic level-1 model specified in Equation A1, we can investigate the relationship between initial status and growth via a level-2 model of the following form:

$$\pi_{oi} = \beta_{00Q} + r_{oi} \quad r_{oi} \sim N(0, \tau_{00}) \quad (A5)$$

$$\pi_{1i} = \beta_{10Q} + b_1(\pi_{oi}) + r_{1i} \quad r_{1i} \sim N(0, \tau_{11}) \quad (A6)$$

$$\pi_{2i} = \beta_{20Q} + b_2(\pi_{oi}) + r_{2i} \quad r_{2i} \sim N(0, \tau_{22}) \quad (A7)$$

In Equation A6, b_1 relates differences in initial status to differences in initial rate, and in Equation A7, b_2 relates differences in initial status to differences in curvature.

Based on such a model, it is then useful to compare the expected final status (i.e., achievement in the fall of Grade 10 ($a_{ii} = 3$)) for a student whose initial status is relatively high versus a student whose initial status value is relatively low. Consider, for example, school S28. We fit the model defined by Equations A1, A5, A6, and A7 to the data for S28 using the latent variable modeling feature in HLM 5, and then computed the expected final status ($FS1_Q$) for a student who is 8 points above the mean initial status estimate for that school (i.e., $IS1 = \hat{\beta}_{00Q} + 8$), and the expected final status ($FS2_Q$) for a student who is 8 points below that school's mean initial status estimate (i.e., $IS2 = \hat{\beta}_{00Q} - 8$). (Note that the average $(\hat{\tau}_{00})^{1/2}$ value in Table A1 is approximately 8 points.) For school 28, we see that $IS1 = 58.12$ and $IS2 =$

42.12 (see Table A2). To compute $FS1_Q$ and $FS2_Q$, we substitute the right hand side of Equations A6 and A7 into Equation A1 and then proceed as follows:

$$FS1_Q = IS1 + (\hat{\beta}_{10Q} + b_1(IS1)) * 3 + (\hat{\beta}_{20Q} + b_2(IS1)) * 9 \quad (A8)$$

and

$$FS2_Q = IS2 + (\hat{\beta}_{10Q} + b_1(IS2)) * 3 + (\hat{\beta}_{20Q} + b_2(IS2)) * 9 \quad (A9)$$

As can be seen in Table A2, the $FS1_Q$ and $FS2_Q$ values for S28 are, respectively, 70.80 and 54.90 points. The expected difference in final status between two students in S28 who are, respectively, 8 points above and 8 points below the mean initial status estimate for that school is: $GAP_Q = FS1_Q - FS2_Q = 15.90$ points. Thus the initial difference between two such students at the outset (i.e., 16 points) remains virtually unchanged based on this analysis.

We proceeded in a similar manner for each school. Consider, for example, S19. Note that for this school $IS1 = 45.97 + 8 = 53.97$, and $IS2 = 45.97 - 8 = 37.97$. The corresponding expected final status values are $FS1 = 69.02$ points and $FS2 = 42.78$ points, respectively. Thus for two such students, who are 16 points apart at the outset, the expected difference by fall of Grade 10 is approximately 10 points wider (i.e., $GAP_Q = 26.24$). As can be seen in Table A2, the GAP_Q values for the set of 8 schools range from 15.90 to 26.24.

Using the same $IS1$ and $IS2$ values in Table A2, we then computed expected final status values for each school based on a linear model for individual change

Table A2
Comparisons of Final Status Values Based on Quadratic and Linear Models for Individual Growth

School	$\hat{\beta}_{00Q}$	IS1 ($\hat{\beta}_{00Q} + 8$)	IS2 ($\hat{\beta}_{00Q} - 8$)	$FS1_Q$	$FS2_Q$	GAP_Q ($FS1_Q - FS2_Q$)	$FS1_L$	$FS2_L$	GAP_L ($FS1_L - FS2_L$)	$DIFF_{GAP}$ ($GAP_L - GAP_Q$)
S10	56.89	64.89	48.89	79.32	60.09	19.23(4)	79.62	57.52	22.10(5)	2.87
S11	56.90	64.90	48.90	78.48	61.12	17.37(3)	78.82	59.93	18.89(3)	1.52
S12	48.02	56.02	40.02	70.55	50.19	20.37(5)	71.06	50.64	20.42(4)	0.06
S19	45.97	53.97	37.97	69.02	42.78	26.24(8)	69.62	42.26	27.36(8)	1.12
S28	50.12	58.12	42.12	70.80	54.90	15.90(1)	70.92	54.12	16.80(1)	0.91
S37	47.26	55.26	39.26	65.59	48.30	17.29(2)	65.76	48.10	17.67(2)	0.37
S43	46.59	54.59	38.59	72.51	48.19	24.32(7)	73.11	47.43	25.69(7)	1.37
S45	51.00	59.00	43.00	73.12	51.33	21.79(6)	74.83	49.88	24.95(6)	3.16

(i.e., $FS1_L$ and $FS2_L$). Specifically, using HLM 5, we fit an HM to the data for each school consisting of the linear model for individual change at level 1, and a level-2 model in which rates of change are modeled as a function of initial status:

$$Y_{ti} = \pi_{0i} + \pi_{1i}a_{ti} + \varepsilon_{ti} \quad \varepsilon_{ti} \sim N(0, \sigma^2) \quad (A10)$$

$$\pi_{0i} = \beta_{00L} + r_{0i} \quad r_{0i} \sim N(0, \tau_{00}) \quad (A11)$$

$$\pi_{1i} = \beta_{10L} + b(\pi_{0i}) + r_{1i} \quad r_{1i} \sim N(0, \tau_{11}) \quad (A12)$$

To compute $FS1_L$ and $FS2_L$, we substitute the right-hand side of Equations A11 and A12 into Equation A10 and then proceed as follows:

$$FS1_L = IS1 + (\hat{\beta}_{10L} + b(IS1)) * 3 \quad (A13)$$

and

$$FS2_L = IS2 + (\hat{\beta}_{10L} + b(IS2)) * 3 \quad (A14)$$

Thus for school S28, the $FS1_L$ and $FS2_L$ values for students with initial status values of $IS1 = 58.12$ and $IS2 = 42.12$, are, respectively, 70.92 and 54.12. Note that these expected final status values are extremely similar to those that were obtained using a quadratic model for individual growth (i.e., $FS1_Q$ and $FS2_Q$). Note further that $GAP_L = FS1_L - FS2_L = 16.80$ differs from GAP_Q by slightly less than a point ($DIFF_{GAP} = GAP_L - GAP_Q = 0.91$).

For each school, we can see that the resulting $FS1_L$, $FS2_L$ and GAP_L values are fairly similar to the corresponding $FS1_Q$, $FS2_Q$ and GAP_Q values. Note, in particular, that the differences between GAP_L and GAP_Q values range from 0.06 to 3.16 points, and that for 6 of the 8 schools, the differences are approximately a point and a half or less. Also, note that the rank order of the schools based on their FS_Q values is nearly identical to the rank order based on their FS_L values; as can be seen in Table A2, there is one reversal: S10 and S12 rank fourth and fifth, respectively, based on their GAP_Q values, and fifth and fourth based on their GAP_L values.

Thus predicted final status values conditional on initial status values 8 points above and below a school's mean initial status estimate seem to be relatively insensitive to the use of a linear model versus a quadratic model for our set of 8 schools. Note that, interestingly, in the schools with the largest $DIFF_{GAP}$ values (i.e., S10 and S45), a substantial number of students do not have fall Grade 10 test scores. This is the case for nearly half of the students in S10, and for approximately one third of the students in S45.

We now turn to questions concerning features of growth at the school level. Consider again Models A (Equations 1, 2a and 2b) and B (Equations A1–A4). First, how similar are the resulting estimates of school mean initial status based on these models? (Note that we term these estimates $\hat{\beta}_{00 \text{ Lin}}$ and $\hat{\beta}_{00 \text{ Quad}}$.) Computing the difference between these estimates for each of the 8 schools ($\text{DIFF}_{\text{MN IS}} = \hat{\beta}_{00 \text{ Lin}} - \hat{\beta}_{00 \text{ Quad}}$), we find that the absolute differences range from 0.05 to 0.83 for 7 of the 8 schools (see Table A3). The largest absolute difference is 1.59 (S11).

Table A3

Comparisons of School Mean Initial Status and School Mean Rates Based on Linear and Quadratic Models for Individual Change

School	$\hat{\beta}_{00 \text{ Lin}}$	$\hat{\beta}_{00 \text{ Quad}}$	$\text{DIFF}_{\text{MN IS}}$	$\hat{\beta}_{10 \text{ Lin}}$	$\hat{\beta}_{10 \text{ Mid}}$	$\text{DIFF}_{\text{MN RATE}}$
S10	56.17	56.89	-0.72	3.89	4.27	-0.38
S11	55.31	56.90	-1.59	4.15	4.30	-0.15
S12	48.53	48.01	0.52	4.28	4.11	0.17
S19	46.02	45.97	0.05	3.32	3.30	0.02
S28	49.38	50.12	-0.74	4.13	4.24	-0.11
S37	47.39	47.26	0.13	3.22	3.22	0.00
S43	45.77	46.60	-0.83	4.55	4.58	-0.03
S45	51.15	51.00	0.15	3.78	3.74	0.04

We now turn to school mean rates of change. First, consider a quadratic model for individual growth in which GRADE_{it} is centered around a value of 8.5: $a_{it} = \text{GRADE}_{it} - 8.5$, where 8.5 represents the midpoint of the series of grades in which LSAY students were observed (i.e., 7, 8, 9, 10). It can be shown that for a student who has been observed at all four time points, the OLS estimate of that student's rate of change under a linear model for growth, which we term $\hat{\pi}_{1i \text{ Lin}}$, will be identical to the OLS estimate of the instantaneous rate of change at $\text{GRADE} = 8.5$, which we term $\hat{\pi}_{1i \text{ Mid}}$ (see Seigel, 1975). Note also that following Seigel (1975), estimates of the expected amount of change across Grades 7–10 would be equivalent based on the two models: $3 \times \hat{\pi}_{1i \text{ Lin}} = 3 \times \hat{\pi}_{1i \text{ Mid}}$. If each student in a school were observed at all four time points, then it follows that the estimate of the mean rate of change based on a linear model for growth ($\hat{\beta}_{10 \text{ Lin}}$) would be identical to the estimate of the mean instantaneous rate at $\text{GRADE} = 8.5$ ($\hat{\beta}_{10 \text{ Mid}}$), and hence the expected amount of

change for students in that school across Grades 7–10 would be equivalent based on both models.

Though not all students in our sample have complete data, it is still instructive to compare estimates of mean rate of change based on a linear model for change and a quadratic model with *GRADE* centered around a value of 8.5. Comparing the differences in the resulting estimates ($\text{DIFF}_{\text{MN RATE}} = \hat{\beta}_{10\text{Lin}} - \hat{\beta}_{10\text{Mid}}$), we see that for 7 of 8 schools, the absolute values of the differences range between 0.00 and 0.17 (see Table A3). The largest absolute difference is 0.38 (S10). Recall that nearly half of the students in this school do not have fall Grade 10 test scores. In sum, for these 8 schools we tend to see little difference in results regarding key concepts of change based on the use of linear and quadratic models for individual change.

References

- Alder, A., Adam, J., & Arenberg, D. (1990). Individual-difference assessment of the relationship between change in and initial level of adult cognitive functioning. *Psychology and Aging, 5*, 560-568.
- Arbuckle, J., & Wothke, W. (1999). *Amos users' guide*. Chicago: Smallwaters.
- Bentler, P. M. (2002). *EQS 6 structural equations program manual*. Encino, CA: Multivariate Software, Inc.
- Blomqvist, N. (1977). On the relation between change and initial value. *Journal of the American Statistical Association, 72*, 746-749.
- Bloom, B. S. (1964). *Stability and change in human characteristics*. New York: Wiley.
- Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. Reading, MA: Addison-Wesley.
- Browne, W. J., & Draper, D. (2000). Implementation and performance issues in the Bayesian and likelihood fitting of multilevel models. *Computational Statistics, 15*, 391-420.
- Browne, W. J., & Draper, D. (2003). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. University of London. Submitted.
- Burstein, L. (1980). The analysis of multi-level data in educational research and evaluation. *Review of Research in Education, 8*, 158-233.
- Carlin, B. (1992). Comment on Morris and Normand (1992). In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. S. Smith (Eds.), *Bayesian statistics 4. Proceedings of the Fourth Valencia International Meeting on Bayesian Statistics* (pp. 336-338). New York: Oxford University Press.
- Carlin, B., & Louis, T. (1996). *Bayes and empirical Bayes methods for data analysis*. London: Chapman and Hall.
- Choi, K. (2002). *Latent variable regression in a three-level hierarchical modeling framework: A fully Bayesian approach*. Unpublished doctoral dissertation, University of California, Los Angeles.
- 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, NJ: Lawrence Erlbaum Associates.

- Cronbach, L. J. (1976). *Research on classrooms and schools: Formulations of questions design and analysis* (Occasional Paper). Stanford, CA: Stanford Evaluation Consortium.
- Gelfand, A., Hills, S., Racine-Poon, A., & Smith, A. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association*, *85*, 972-985.
- Gelman, A., Carlin, J., Stern, H., & Rubin, D. (1995). *Bayesian data analysis*. London: Chapman & Hall.
- Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in practice*. London: Chapman & Hall.
- Hox, J. J. (1993). Factor analysis of multilevel data: Gauging the Muthén model. In J. H. J. Oud & R.A.W. van Blokland-Vogeleesang (Eds.), *Advances in longitudinal and multilevel analysis in the behavioral sciences* (pp. 141-156). Nijmegen, The Netherlands: 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*, 157-174.
- Joreskog, K., & Sorbom, D. (2000). *LISREL 8.30*. Chicago: Scientific Software International.
- Khoo, S. T. (2001). Assessing program effects in the presence of treatment-baseline interactions: A latent curve approach. *Psychological Methods*, *6*, 234-257.
- Miller, J. D., Kimmel, L., Hoffer, T. B., & Nelson, C. (1999). *Longitudinal Study of American Youth: User's manual*. Chicago: International Center for the Advancement of Scientific Literacy, Academy of Sciences.
- Morris, C. N. (1987). Comment on Tanner and Wong. *Journal of the American Statistical Association*, *82*, 542-543.
- Muthén, O. B. (1994). Multilevel covariance structural analysis. *Sociological Methods and Research*, *22*, 376-398.
- Muthén, O. B. (1997). Latent variable modeling with longitudinal and multilevel data. In A. Raftery (Ed.), *Sociological Methodology 1997* (pp. 453-480). Boston: Blackwell Publishers.
- Muthén, 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*, 371-402.

- Muthén, L., & Muthén, O. B. (2002). *Mplus user's guide (Version 2.1)*. Los Angeles: Author.
- Muthén, 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.
- Natarajan, R., & Kass, R. (2000). Reference Bayesian methods for generalized linear mixed models, *Journal of American Statistical Association*, 95, 227-237.
- Racine-Poon, A. (1992). SAGA: Sample assisted graphical analysis. In J. Bernardo, J. Berger, A. Dawid, & A. Smith (Eds.), *Bayesian statistics 4. Proceedings of the Fourth Valencia International Meeting on Bayesian Statistics* (pp. 389-404). New York: Oxford University Press.
- Rasbash, J., Browne, W. J., Goldstein, H., Yang, M., Plewis, I., Healy, M., et al. (2002). *A user's guide to MlwiN, Version 2.1*. London: University of London, Institute of Education.
- Raudenbush, S. W., & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, 59, 1-17.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.
- Raudenbush, S., Bryk, A., Cheong, Y., & Congdon, R. (2000). *HLM5: Hierarchical linear and non-linear modeling*. Homewood, IL: Scientific Software International.
- Raudenbush, S. W., & Sampson, R. (1999). Assessing direct and indirect associations in multilevel designs with latent variables. *Sociological Methods and Research*, 28, 123-153. Newbury Park, CA: Sage.
- Rubin, D. B. (1981). Estimation in parallel randomized experiments. *Journal of Educational Statistics*, 6, 377-400.
- Seigel, D. G. (1975). Several approaches for measuring average rates of change for a second degree polynomial. *The American Statistician*, 29, 36-37.
- 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. (1995). Furthering our understanding of the effects of educational programs via a slopes-as-outcomes framework. *Educational Evaluation and Policy Analysis*, 17, 295-304.

- Seltzer, M., Choi, K., & Thum, Y. (2003). Examining relationships between where students start and how rapidly they progress: Using new developments in growth modeling to gain insight into the distribution of achievement within schools. *Education Evaluation and Policy Analysis, 25*, 263-286.
- 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.
- Spiegelhalter, D., Best, N., Gilk, W., & Inskip, H. (1996). Hepatitis B: A case study in MCMC methods. In W. Gilks, S. Richardson, & D. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in practice* (pp. 21-43). London: Chapman & Hall.
- Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2003). *WinBUGS: windows version of Bayesian inference using Gibbs sampling, version 1.4, User Manual*. Cambridge: Cambridge University, MRC Biostatistics Unit.
- Svardssudd, K., & Blomqvist, N. (1978). A new method for investigating the relation between change and initial value in longitudinal blood pressure data: Description and application of the method. *Scandinavian Journal of Social Medicine, 6*, 85-95
- Wasserman, L. A. (2000). Asymptotic inference for mixture models using data dependent priors. *Journal of the Royal Statistical Society, Series B, 62*, 159-180.
- Werts, C. F., & Hilton, T. L. (1977). Intellectual status and intellectual growth, again. *American Educational Research Journal, 14*, 137-146.
- Wu, M., Ware, J., & Feinleib, M. (1980). On the relation between blood pressure change and initial value. *Journal of Chronic Disease, 33*, 637-644.
- Zellner, A. (1971). *An introduction to Bayesian inference in econometrics*. New York: Wiley and Sons.