Latent Variable Regression Four-Level Hierarchical Model Using Multisite Multiple-Cohort Longitudinal Data

Kilchan Choi
Jinok Kim
CRESST/UCLA

This article proposes a latent variable regression four-level hierarchical model (LVR-HM4) that uses a fully Bayesian approach. Using multisite multiple-cohort longitudinal data, for example, annual assessment scores over grades for students who are nested within cohorts within schools, the LVR-HM4 attempts to simultaneously model two types of change, arising from individual student over grades, and successive cohorts in the same grade over years. In addition, as an extension of Choi and Seltzer, the LVR coefficients, that is, gap-in-time parameter, capturing the relationships between initial status and rates of changes within each cohort and school, help bring to light the distribution of student growth and differences in the distribution over different cohorts within schools. Advantages associated with the LVR-HM4 can be highlighted in studies on monitoring school performance or evaluations of policies and practices that may target different aspects of student academic performance such as initial status, growth, or gap over time in schools.

Introduction

Data on academic performance have become more widely available and more systematic over the last decade. Since the No Child Left Behind Act (NCLB, 2002) that mandated states to administer end-of-year assessments to all students in Grades 3 through 8, statewide data on student achievement have been accumulating and can be linked across years. Such abundant data combined with innovative modeling techniques facilitate a more detailed examination of changes over time in academic performance.

For example, Choi and Seltzer (2010) examine equity in academic performance in addition to growth patterns in academic performance. They apply three-level hierarchical models (HMs) with latent variable regressions (LVRs) in

which the LVR coefficients—capturing the relationships between initial status and rates of change within each school—are posited to vary across schools. In the context of monitoring school performance, the technique can help bring to light the within-school distribution of academic performance (see also Seltzer, Choi, & Thum, 2003). In some schools, students with relatively high initial status may progress rapidly; in these schools, initial gaps in achievement are magnified over time. In other schools where students with lower initial status progress more rapidly than students with higher initial status, initial gaps diminish over time. The coefficients of LVR capture the extent to which initial gaps will magnify or shrink over time. That is a useful framework for studying the equity issues in the sense that the modeling framework puts equity in school performance on a time dimension; the technique quantifies how equitable performances within schools are over time (see also Choi, Seltzer, Herman, & Yamashiro, 2007, for application of this technique).

Drawing on these recent advances in growth modeling, this article presents four-level HMs with LVRs (LVR-HM4) that use a fully Bayesian approach. This technique underscores the state's annual assessment data cumulated over time that consist of four levels: annual assessment scores over grades nested within students, who are in turn nested within cohorts, which are in turn nested within schools. As compared to the three-level HMs previously discussed, the LVR-HM4 can be very useful in studies of student achievement by focusing on, and incorporating, the cohort level in the nesting structure. First, the four-level data structure makes it possible to jointly model two types of change. By introducing the cohort level in the model, this technique creates a time series for each school, which consists of successive cohorts in schools. Thus, the model simultaneously examines changes in the academic performance of individual students over grades and changes of successive cohorts over academic years.

As will be seen, monitoring systems or evaluation studies based on this approach can benefit from advantages associated with joint modeling. Monitoring school performance or evaluation of policies and practices may target multiple aspects such as status, growth, and equity over time. The multiple target dimensions would often be strongly correlated. Simultaneous estimation assuming the joint distribution of these outcomes of interest helps draw sound inferences concerning each of the outcomes representing multiple target dimensions. Although estimating parameters of such models can be extremely complex, a fully Bayesian approach using the Gibbs sampler would make such estimations possible.

In the following, we introduce two key areas in which this technique can provide novel approaches, school monitoring, and evaluation studies. We provide brief background information on research in these areas and potential contributions of our proposed framework to these areas.

School Monitoring

In studying changes over time in the academic performance of schools (Goldschmidt & Choi, 2007), one can track the same group of students as they progress through grade levels or alternatively one can track successive cohorts in the same grade over academic years. Tracking the same group of students across grades is the comparison of the current grade score to the prior grade score of individual students. Thus, it is concerned with changes of *individual students*, which are averaged at the school level for the purpose of monitoring school performance. Tracking successive cohorts in the same grade over years is the comparison of scores of the current year cohort to the scores of the previous year cohort. Haertel (2005; also see Thum, 2006) refers to tracking the same group of students over grades as "an individual growth design" (an IG design; p. 5) and tracking successive cohorts in the same grade over years as "a successive cohort design" (an SC design; p. 4).

Many research studies and monitoring systems in education are explicitly or implicitly based on the IG design. To evaluate or identify different types of instructional programs or educational practices, most often the key interest lies in how much or how rapidly students would progress during the implementation period. This is because, as Willet (1989) writes, "the very notion of learning implies growth and change" (p. 346). Value-added models that have been a prevailing measure of teacher and/or school effectiveness (McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004; Ponisciak & Bryk, 2005; Sanders, Saxton, & Horn, 1997) can also be viewed as being based on the IG design. By analyzing individual students' time-series data using regression-based models (e.g., multilevel or mixed models), value-added models are fundamentally concerned with individual students' progress over time, which are estimated at the teacher or school level to represent teachers' or schools' value-added scores.

The SC design is often used when a key interest lies in the change in schools (see also Leckie & Goldstein, 2009). Because the SC design often examines how well schools perform regardless of the entry and exit of their students, it is at the heart of the monitoring of school performance over time. As Haertel (2005) notes, if some low-performing schools do not improve or even worsen over years, states or districts may determine that those schools may be in great need of intervention or support.

Simultaneously modeling both types of change arising from both the IG and the SC designs may take advantage of HM4s that incorporate the cohort level. Standard HM4 allows us to estimate both changes of individual students corresponding to the IG design and changes of cohorts in successive years corresponding to the SC design. In addition to this, the LVR features of the LVR-HM4 offer additional performance indicators of equity in academic achievement for school monitoring systems.

In evaluation studies in which an event of interest (e.g., implementation of new educational policies or programs) occurs during a time series and is hypothesized to affect academic achievement upon or after the event in the time series, the interrupted time series (ITS) design has been increasingly used. It is particularly useful in assessing the influence of major educational policy initiatives that are implemented without manipulations such as random assignment. In such studies, the ITS designs are often based on the SC design that tracks successive cohorts. They examine whether there is a sudden change in the academic performance of cohorts that corresponds to the onset of a policy implementation relative to the baseline trends of the previous cohorts, while typically adjusting for various characteristics of cohorts.

In the framework of LVR-HM4, a time series at the cohort level (i.e., a time series that consists of successive cohorts) includes latent variables measuring various dimensions of performance such as status, learning over time, and equity over time. Thus, for example, instead of simple measures, LVR-HM4 estimates whether the onset of policies or practices of interest would be associated with the significant boost or interruption in trends in terms of schools' value added (rate of progression of students in each cohort over consecutive years) and schools' equity (distribution of progression of students in each cohort over consecutive years).

In the rest of this article, we aim to elucidate how the important features of this modeling technique, such as regressions among growth parameters (i.e., latent variables), enable us to measure various equity parameters of school performance, by presenting standard HM4s and their extensions embedded in an illustration of analysis of multisite, multiple-cohort longitudinal data. The illustration uses multisite multiple-cohort longitudinal data drawing from annual assessment systems in a large urban school district located in a northwestern state and assesses the influence of NCLB on student achievement in the district. Based on the ITS design, we use an LVR-HM4 in order to compare pre-NCLB cohorts and post-NCLB cohorts on three different dimensions of student growth: initial status, learning over time, and equity over time. This illustration uses the reading scale scores of the Iowa Tests of Basic Skills (ITBS), which was not a particularly high-stakes test but was administered annually. Note that our illustration focuses more on the application of LVR-HM4s in evaluation studies, instead of drawing conclusive inferences regarding the impact of NCLB.

Specifically, we first present the descriptive statistics of the data and thereby explain the structure of data that are consistent with the modeling approach of this article. Then, we introduce a useful preliminary analysis of an HM4, followed by two LVR-HM4s. Note that all analyses of the LVR-HM4 presented in this article are conducted using the software WinBUGS (Version 1.4.3; Windows

TABLE 1.
Data Structure: Cohort, Grade, and Year

Cohort/Year	1998	1999	2000	2001	2002	2003	2004
Cohort 1	Grade 3		Grade 5				
Cohort 2		Grade 3		Grade 5			
Cohort 3			Grade 3		Grade 5		
Cohort 4				Grade 3		Grade 5	
Cohort 5					Grade 3		Grade 5

version of Bayesian analysis using the Gibbs sampler: Spiegelhalter, Thomas, Best, & Lunn, 2003).

Model 1, unconditional HM4, estimates student initial status (performance status at Grade 3) and gain (growth rate from Grade 3 to Grade 5) and examines the extent of variability associated at each level (across students within cohorts within schools, across cohorts within schools, and across schools). Model 2, unconditional LVR-HM4, incorporates LVRs into Model 1 and estimates the gap-in-time indicator within cohorts within schools as well as the relationships both across cohorts and across schools in the district. The gap-in-time indicator is estimated also depending on the initial status of cohorts and on the initial status of schools. Building upon Model 2, Model 3 examines cohort-to-cohort changes in terms of initial status, gain, and the gap-in-time indicator. Model 3 also examines whether the installment of NCLB relates to any changes in cohort-to-cohort performances in these elementary schools by adding time-metric variables at the cohort level. We include an observed student characteristic (student free/reduced-price lunch [FRL] status) as well as an observed school characteristic (school's adequate yearly progress [AYP] status) to increase the precision of inferences concerning NCLB effects. We conclude with discussions of potential value, challenges, and extensions of the modeling strategies, as well as findings from an analysis of the illustrative example.

Application of LVR-HM4s

Data

Table 1 shows the data structure that serves as an example of the structure of the multisite multiple-cohort longitudinal data required for the HM4s with LVRs. The columns of Table 1 show the academic years that correspond to the study data. Each row shows one cohort, while the cells show corresponding years of data for each cohort. The first cohort in the sample comprises students who entered Grade 3 in the 1998–1999 school year, and the remaining four successive cohorts are students who entered the same grade for the 4 following years. Thus, the last (the fifth) cohort consists of students who entered Grade 3 in 2002–2003. As for the longitudinal aspect of the data, for all five cohorts in all 74 schools, assessments were completed at two longitudinal time points, third grade and fifth

TABLE 2.

Cohort-by-Cohort Descriptive Statistics of Observed Initial Status, Gain, and Percentage of Students Eligible for Free/Reduced-Price Lunch

Cohort		Status at Grade 3		Gain (Grade 3 to Grade 5)		% Free/Reduced-Price Lunch	
	N	Mean	*SD*	Mean	*SD*	Mean	*SD*
1 (1998–2000)	2,501	188.4	23.5	30.5	16.1	41.6	4.9
2 (1999–2001)	2,324	191.9	21.6	29.7	14.1	37.9	4.9
3 (2000–2002)	2,054	191.7	23.6	28.7	14.4	40.2	4.9
4 (2001–2003)	2,548	191.2	21.8	29.8	13.8	36.1	4.8
5 (2002–2004)	2,103	191.8	23.3	29.0	14.0	40.1	4.9

grade. The outcomes of interest are the ITBS reading scale scores that are vertically equated developmental scores. Most of the 74 elementary schools have data available for five successive cohorts. However, for 5 of the 74 schools, data are available for four cohorts, and for 1 of the 74 schools, data were available for three cohorts. On average, about 13% of students transitioned to different schools in the district between Grades 3 and 5, and an additional 2% of students did not have a test score either in Grade 3 or 5. The analysis thus included only students who had test scores at both Grades 3 and 5 and remained in the same school.

In Table 1, the arrows that go to the right (by row) show changes of individual students across grades (the growth of individual students, IG design), while the arrows that go down or diagonally down show cohort-to-cohort changes over years (the change in successive cohorts, SC design). The vertical line in Table 1 displays the beginning of the implementation of the NCLB Act. We have multiple years of data for the pre-NCLB era (1998–2001 academic years) and for the post-NCLB era (2002–2004 academic years). Regarding the implementation of NCLB, Hanushek and Raymond (2003) identify states that had implemented strong accountability systems for schools prior to the initiation of NCLB (see also Carnoy & Loeb, 2002, for statewide implementation levels). The district in this study belonged to a state that was not an early adopter state; therefore, academic years 2002–2003 and after were used as the timing of the intervention in this article. Based on the 2002–2003 initiation, note that while Cohorts 1 and 2 of this study are entirely pre-NCLB cohorts with Cohort 5 being entirely a post-NCLB cohort, for Cohorts 3 and 4, NCLB went into effect while their growth/ gain is being studied. As will be seen in the model section, this feature will require multiple time-metric variables indicating the NCLB era in the cohort-to-cohort part of the analysis.

Table 2 presents the descriptive statistics for the sample. The total number of students in the sample is 11,530, and the average number of students per cohort is

2,306. For the five cohorts, the mean ITBS reading score at Grade 3 is around 191, and its standard deviation (*SD*) is approximately 22. The minimum and maximum scores of the ITBS reading scores from the study sample (not shown in Table 2) are 139 and 307. The observed mean change/gain between Grades 3 and 5 is around 29.5 points, of which the magnitude is about 1.3 *SD*s. In addition, the percentage of students eligible for FRL is approximately 39.5. Thus, based on the sample means and *SD*s, the five cohorts in the sample appear similar in all three measures.

Although the by-cohort statistics indicate that the sample means are similar across cohorts, it is important to note that cohort means may vary within schools. If cohort means vary in different directions across schools, the average values over all schools (as shown in Table 2) will cancel out and mask the variation across cohorts within schools. Similarly, the relationship between initial status and gain might vary across cohorts and schools although the *SD*s of status and gain of the cohorts in Table 2 are similar.

Under the assumption that NCLB might affect students' performance in different cohorts, we first examined whether and how much cohorts' mean scores and gain scores would vary across cohorts within a school. Although modeling with the sample means may be problematic due to the different sample sizes of cohort–school combinations, this preliminary analysis shows different variations in between cohorts within a school and between schools in the cohort's mean scores and gain scores. Thus, three quantities were examined: cohorts' mean scores at Grade 3, cohorts' mean gain scores (mean of the Grade 5 score minus mean of the Grade 3 score), and cohorts' percentages of students eligible for FRL.

To obtain variance components of cohorts and schools, we fitted an unconditional two-level HM separately to each of the sample means of those three quantities for each cohort within schools. Note that the relative ratios of the between-cohort variability to the between-school variability are of interest in this two-level approach since the unconditional four-level model presented in the following section would have correctly portioned off level 1 and level 2 variability. As for the mean scores at Grade 3, the intraclass correlation is .87, which means that the between-cohort, within school variability is 13%, whereas the between-school variability is 87%. That is, the between-school variability of ITBS mean scores greatly exceeds the between-cohort, within-school variability. In contrast, mean gain scores show greater variability between cohorts than between schools. The intraclass correlation based on a two-level HM is .31, which indicates that the between-cohort variability is more than twice the between-school variability (69% vs. 31%). As for the percentage of FRL students, the intraclass correlation is .83; that is, the between-cohort variability is approximately 17%, and the between-school variability is 83% (see Figure 1).
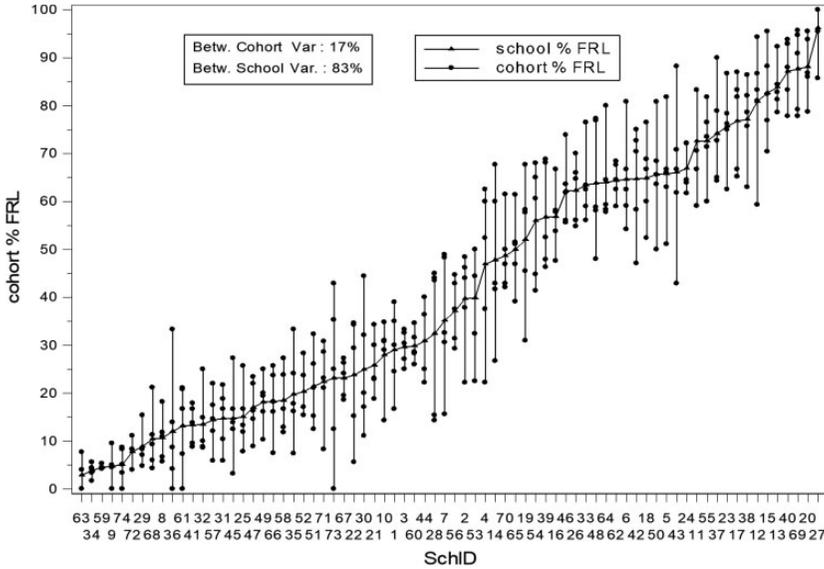
FIGURE 1. *Percentage of students eligible for free/reduced-price lunch by cohort and school.*

Figures 1 through 3 present a graphical display of the intraclass correlation results, which help translate the results to the actual scales for this particular study. In these figures, a circle represents a cohort's mean score, while a triangle represents a school's overall mean across cohorts. In Figure 2 (mean score at Grade 3), the difference between the lowest and highest mean score by school is 55 points (170–225). However, the largest difference between cohorts within a school did not exceed 25 points (School 40). In Figure 3, school mean gains of the 74 elementary schools in the sample range from approximately 17 to 35, while the variability between cohorts in some schools is very large, as much as 30 to 40 points in Schools 48 and 40.

In summary, for mean scores at Grade 3 and percentage of students eligible for FRL, there is far greater variability across schools than across cohorts within schools, whereas there is a great deal of variability across cohorts within schools in mean gain scores. The substantial variability across cohorts suggest that HM4s including a cohort as a unit of analysis is necessary to model cohorts' change over time instead of more conventional three-level HMs ignoring cohorts' variability.

Model 1: HM4s Estimating Initial Status and Gains of Individual Students

To examine the extent of the variability in both student initial status and student growth rate at each level (student, cohort, and school levels), an unconditional HM4 is posed.
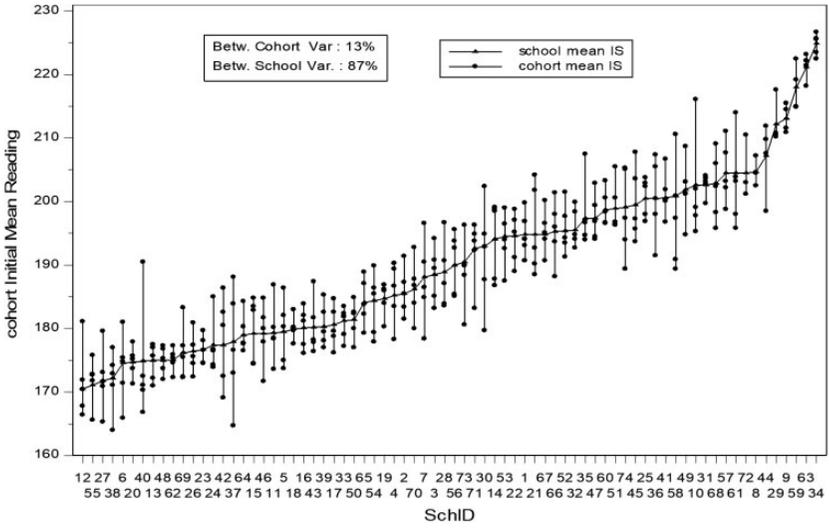
FIGURE 2. *Observed initial Iowa Tests of Basic Skills reading mean scores by cohort and school.*
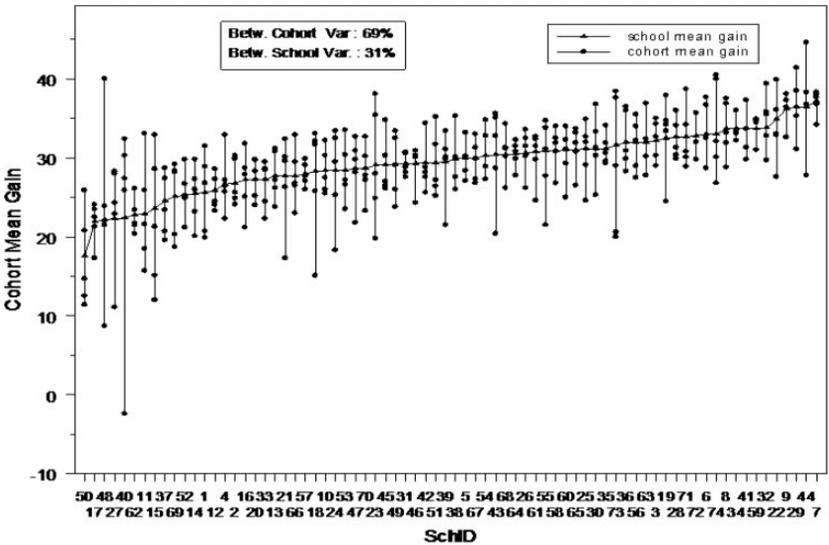
FIGURE 3. *Observed mean gain scores by cohort and school.*

$$Y_{tijk} = \pi_{0ijk} + \pi_{1ijk}\text{Time}_{tijk} + \varepsilon_{tijk} \qquad \varepsilon_{tijk} \sim N(0, \sigma^2). \qquad (1-1a)$$

Y_{tijk} is the outcome score at measurement occasion or grade t ($t = 3, 5$), for student i ($i = 1, 2, \ldots, N_{jk}$), in cohort j ($j = 1, 2, \ldots, J_k$), in school k ($k = 1, 2, \ldots, K$). The time-metric variable, Time_{tijk}, takes a value of 0 for the outcome test score at Grade 3 and 1 for the test score at Grade 5. Thus, the two growth parameters, π_{0ijk} and π_{1ijk}, represent initial status and gain, respectively, for student i in cohort j, in school k. Note that those two parameters are latent variables and treated as random variables at level 2.

Note that we use standard errors (SEs) of students' ITBS scores to estimate individual student growth rate/gain with ITBS reading scores measured at two time points. Specifically, the left and right sides of Equation 1 – 1a are scaled by the inverse of conditional SEs of measurement of ITBS reading score for time t, student i, cohort j in school k.

$$Y^*_{tijk} = \pi_{0ijk} + \pi_{1ijk}\text{Time}^*_{tijk} + \varepsilon^*_{tijk} \qquad \varepsilon^*_{tijk} \sim N(0, 1). \qquad (1-1b)$$

By this rescaling, the level 1 residuals become $\varepsilon^*_{3ijk} \sim (0,1)$ and $\varepsilon^*_{5ijk} \sim (0,1)$ (See more details in Bryk, Thum, Easton, & Luppescu, 1998, p. 135; Choi & Seltzer, 2010, p. 80; Choi et al., 2007, p. 23). Due to the rescaling, the level 1 model in Equation 1 – 1b becomes a just-identifiable model with two time-point measures and two estimands (π_{0ijk} and π_{1ijk}).

Equations 1 – 2a and 1 – 2b specify the level 2 (between-student) model. β_{00jk} and β_{10jk} represent cohort j in school k's mean initial status and mean gain, respectively. In Equations 1 – 3a and 1 – 3b, γ_{000k} and γ_{100k} are mean initial status and mean gain for school k. Lastly, in Equations 1 – 4a and 1 – 4b, θ_{0000} and $\theta_{1,000}$ are grand mean initial status and grand mean gain, respectively.

$$\pi_{0ijk} = \beta_{00jk} + r_{\pi 0ijk} \qquad r_{\pi_{0ijk}} \sim N(0, \tau_{\pi 0}) \qquad (1-2a)$$

$$\pi_{1ijk} = \beta_{10jk} + r_{\pi 1ijk} \qquad r_{\pi 1ijk} \sim N(0, \tau_{\pi 1}) \qquad (1-2b)$$

$$\beta_{00jk} = \gamma_{000k} + U_{\beta 00jk} \qquad U_{\beta 00jk} \sim N(0, \tau_{\beta 00}) \qquad (1-3a)$$

$$\beta_{10jk} = \gamma_{100k} + U_{\beta 10jk} \qquad U_{\beta 10jk} \sim N(0, \tau_{\beta 10}) \qquad (1-3b)$$

$$\gamma_{000k} = \theta_{0000} + V_{\gamma 000k} \qquad V_{\gamma 000k} \sim N(0, \tau_{\gamma 000}) \qquad (1-4a)$$

$$\gamma_{100k} = \theta_{1000} + V_{\gamma 100k} \qquad V_{\gamma 100k} \sim N(0, \tau_{\gamma 100}) \qquad (1-4b)$$

Random effects at each level are assumed normally distributed with mean 0 and a level-specific variance. Covariances at levels 2, 3, and 4 are, respectively, $\tau_{\pi 0, \pi 1}$, $\tau_{\beta 00, \beta 10}$, and $\tau_{\gamma 000, \gamma 100}$.

Table 3 presents the results for Model 1 posterior means, medians, SDs, and the 95% intervals of the marginal posterior distributions of each parameter. Note that

TABLE 3.
Model 1: Unconditional Four-Level Hierarchical Model

	Estimate	SE	95% Interval	Median
Fixed effects				
Grand mean initial status (θ_{0000})	189.1	1.46	[186.2, 192.0]	189.1
Grand mean gain (θ_{1000})	30.32	0.50	[29.34, 31.29]	30.32
Variance components				
1. Level 2 variance: Between-student				
Initial status ($\tau_{\pi00}$)[a] 61.5%	264.1	4.40	[255.6, 272.9]	264.1
Gain ($\tau_{\pi10}$)[b] 75.4%	81.28	2.46	[76.51, 86.16]	81.26
Cov ($r_{\pi0}, r_{\pi1}$);[c] Corr ($r_{\pi0}, r_{\pi1}$) = 0.17	25.23	2.40	[20.46, 29.86]	25.26
2. Level 3 variance: Between-cohort				
Initial status ($\tau_{\beta00}$)[a] 2.8%	11.75	1.77	[8.55, 15.49]	11.65
Gain ($\tau_{\beta10}$)[b] 11.6%	12.56	1.60	[9.68, 15.95]	12.47
Cov ($u_{\beta00}, u_{\beta10}$);[c] Corr ($u_{\beta00}, u_{\beta10}$) = −0.31	−3.82	1.24	[−6.37, −1.48]	−3.79
3. Level 4 variance: Between-school				
Initial status ($\tau_{\gamma000}$)[a] 35.7%	153.2	26.4	[109.6, 212.8]	150.4
Gain ($\tau_{\gamma100}$)[b] 13.0%	14.01	3.00	[9.10, 20.8]	13.68
Cov ($V_{\gamma000}, V_{\gamma100}$);[c] Corr ($V_{\gamma000}, V_{\gamma100}$) = 0.66	30.57	7.21	[18.43, 46.78]	29.89

[a]Initial status variance percentage for each of the three levels (between-student, between-cohort, between-school). [b]Gain variance percentage for each of the three levels (between-student, between-cohort, between-school). [c]Correlations between random effects for each of the three levels (between-student, between-cohort, between-school).

the 95% interval is constructed based on .025 and .975 quantiles of marginal posterior distribution.[1] The grand mean initial status is approximately 189.1 and the grand mean gain is 30.3. As for the variability in initial status, the percentages of between-student, between-cohort within-school, and between-school variances over the total variance are, respectively, 61.5, 2.8, and 35.7. In the gain parameter, the percentages of those three variances over the total variance are, respectively, 75.4, 11.6, and 13.0. As such, the initial status is rather homogeneous across cohorts but fairly heterogeneous across schools. However, the gains from Grades 3 through 5 appear to be substantially heterogeneous across cohorts within schools. These results are consistent with the findings in the preliminary analysis.

Model 2: LVR-HM4 Estimating the Gap-in-Time Indicator

Model 2 is motivated to address the following substantive questions: (a) To what extent is the initial gap at Grade 3 magnified or diminished at Grade 5? (b) To what extent is the cohort's initial status associated with the cohort's gain? In other words, are cohorts with higher initial status gaining more than cohorts with lower initial status? (c) To what extent is the cohort's initial status

associated with the gap-in-time indicator? In other words, is the gap diminished or magnified more in cohorts with higher initial status than in cohorts with lower initial status? (d) To what extent is the school's initial status associated with the school's gain and the gap-in-time indicator? In other words, does a school with a higher mean initial status gain more and is the gap reduced more in comparison to a school with a lower mean initial status?

By employing LVRs at levels 2, 3, and 4 in Model 2, we examine the extent to which initial status is consequential to the amount of gain in three different levels: student, cohort, and school. In addition, we examine whether initial status at the beginning grade is consequential to the amount of gap-in-time in cohort and school levels. Based on the specification in Model 1, Model 2 will add LVRs at levels 2, 3, and 4, while the level 1 specification will remain identical. What follows in this Model 2 section will show the LVR extension in equation forms. Before presenting specifications of LVRs, there is an important consideration of whether LVR coefficients are treated as fixed or random variables. For example, the school-level relationship is viewed as a between-cluster relationship and a fixed effect. In contrast, because the student-level relationship can be assumed to vary across cluster levels (i.e., cohorts and schools), it can be modeled with random effects. In other words, the student-level relationship can be estimated by each cohort within a school and by each school. In doing that, the variances of two latent variables (i.e., initial status and gain) are allowed to vary across cohorts and schools; otherwise, the LVR coefficients would be the same across cohorts and schools. Likewise, the cohort-level relationship can also be treated as random variables assuming to vary across schools. However, treating the cohort-level relationship as a fixed effect is a reasonable choice in this data set since only a limited number of cohorts ($J = 5$) is available to estimate separate variance–covariance matrix within each school.

The level 1 (within-student) model in Model 2 is the same as the one in Model 1:

$$Y^*_{tijk} = \pi_{0ijk} + \pi_{1ijk} \text{Time}^*_{tijk} + \varepsilon^*_{tijk} \qquad \varepsilon^*_{tijk} \sim N(0,\ 1) \qquad (2-1)$$

At the level 2 (between-student; within-cohort) model, we estimate three key latent variables (i.e., growth parameters) for cohort j in school k: initial status, growth rate, and the relationships of initial status to growth rate.

$$\pi_{0ijk} \ = \ \beta_{00jk} + r_{\pi 0ijk} \qquad r_{\pi 0ijk} \sim N(0, \tau_{\pi 0jk}) \qquad (2-2a)$$

$$\pi_{1ijk} = \beta_{10jk} + B_{w\_jk}(\pi_{0ijk} - \beta_{00jk}) + r_{\pi 1ijk}, \text{Cov}(r_{\pi 0ijk}, r_{\pi 1ijk}) \ = \ 0 \qquad r_{\pi 1ijk} \sim N(0, \tau_{\pi 1jk})$$
$$(2-2b)$$

Equation $2-2b$ presents LVR at level 2, where the student growth rate is modeled as a function of student initial status. Note that the student's initial status is centered on his or her cohort's mean initial status in school k. By virtue of this centering, β_{10jk} represents the mean growth rate for cohort j in school k.

The LVR coefficient, B_{w_jk}, captures the relationship between student's initial status and growth rate in cohort j in school k. Choi and Seltzer (2010) show that by treating this within-group LVR coefficient as random (i.e., varying across schools), it addresses an important question concerning how the student growth is distributed within each school. Similar to Choi and Seltzer (2010), we refer to this LVR coefficient (B_{w_jk}) as a within-cohort, within-school initial status/rate of change slope, or gap-in-time indicator. The gap-in-time indicator tells us the extent to which the initial gap (i.e., achievement gap at Grade 3) between students, say 30 points initially, becomes magnified or diminished over grades (at Grade 5) for each cohort of students in each school. The gap-in-time indicator is a practical interpretation of this LVR coefficient. For the LVR coefficients of this type in this article, B followed by w, b, or c denotes within-cohort, between-school, or between-cohort relationships, respectively (B followed by b and c will be presented in equations below). The within-cohort LVR coefficient, B_w, is assumed to vary across cohorts nested within schools. Random effects ($r_{\pi 0ijk}$ and $r_{\pi 1ijk}$) are assumed to be normally distributed with mean 0 and their respective variances $\tau_{\pi 0jk}$ and $\tau_{\pi 1jk}$. As noted earlier, these variances differ across cohorts and schools, which is a necessary assumption for within-cohort LVR coefficient to be treated as a random variable (see also Leckie, French, Charlton, & Browne, 2014, for heterogeneous variances in a cluster level in two-level models). Since the growth rate is conditional on the initial status, the covariance between these two random effects is equal to 0, which is required to be an identifiable model in the setting of LVRs.

In the level 3 (between-cohort, within-school) model, the three growth parameters—cohort mean initial status (β_{00jk}), cohort mean growth rate (β_{10jk}), and the within-cohort initial status/rate of change slope (B_{w_jk})—are treated as outcomes.

$$\beta_{00jk} = \gamma_{000k} + U_{\beta 00jk} \qquad\qquad U_{\beta 00jk} \sim N(0, \tau_{\beta 00}) \quad (2-3a)$$

$$\beta_{10jk} = \gamma_{100k} + B_{c1}(\beta_{00jk} - \gamma_{000k}) + U_{\beta 10jk} \qquad U_{\beta 10jk} \sim N(0, \tau_{\beta 10}) \quad (2-3b)$$

$$B_{w\_jk} = B_{w\_0k} + B_{c2}(\beta_{00jk} - \gamma_{000k}) + U_{Bw\_jk} \qquad U_{Bw\_jk} \sim N(0, \tau_{Bw}) \quad (2-3c)$$

$$T_U = \begin{pmatrix} \tau_{\beta_{00}} & 0 & 0 \\ 0 & \tau_{\beta_{10}} & \tau_{\beta_{10}, B_w} \\ 0 & \tau_{B_w, \beta_{10}} & \tau_{B_w} \end{pmatrix}. \qquad (2-3d)$$

In Equations $2-3a$, $2-3b$, and $2-3c$, the coefficients, γ_{000k} and γ_{100k}, represent the mean initial status and the mean growth rate, respectively, for school k. Likewise, B_{w_0k} is the mean within-cohort initial status/rate of change slope across cohorts for school k. T_U is the variance–covariance matrix of the three random effects ($U_{\beta 00jk}$, $U_{\beta 10jk}$, U_{Bw_jk}), and we also define the submatrix,

$T_{U_{(2,3)}}$, as the variance–covariance matrix of the latter two random effects (i.e., $U_{\beta 10jk}$, U_{Bw_jk}).[2]

There are two LVR coefficients in the level 3 model. First, the LVR coefficient, B_{c1}, represents the expected relationship between cohort mean initial status and cohort mean growth rate. In contrast to the within-group LVR coefficient (B_{w_jk}), this coefficient is termed as the between-cohort initial status/rate of change slope and addresses a question concerning the extent to which cohort initial status is associated with cohort growth. Another LVR coefficient, B_{c2}, captures the relationship between cohort initial status and within-cohort, within-school initial status/rate of change slopes, and it allows us to examine whether the relationship between students' initial status and their rate of change slopes are associated with their cohort initial status. The random effects, $U_{\beta 00jk}$, $U_{\beta 10jk}$, and U_{Bw_jk}, are assumed to be multivariate normally distributed with mean 0 and variance–covariance matrix of T_U as shown in Equation 2 – 3d.

It is important to note that each of the three growth parameters for cohort j in school k can comprise a time series of cohorts within schools over subsequent academic years. Thus, the between-cohort model presented above in Equations 2 – 3a through 2 – 3c can be readily extended by including any time-varying cohort characteristics in the model. A time-metric variable can be included as a predictor for examining how each cohort's growth parameters change over successive cohorts in subsequent academic years. To illustrate this point, Equations 2 – 3a through 2 – 3c are shown again with a time-metric variable included in the following. By presenting this specification, we aim to reinforce the idea that one can simultaneously estimate changes of subsequent cohorts over academic years as well as changes/growth of individual students over grades.

$$\beta_{00jk} = \gamma_{000k} + f(\text{Year}_{jk}) \; + \; U_{\beta 00jk} \qquad\qquad U_{\beta 00jk} \sim N(0, \tau_{\beta 00})$$
$$(2 - 3a')$$

$$\beta_{10jk} = \gamma_{100k} + B_{c1}(\beta_{00jk} - \gamma_{000k}) + f(\text{Year}_{jk}) + U_{\beta 10jk} \qquad U_{\beta 10jk} \sim N(0, \tau_{\beta 10})$$
$$(2 - 3b')$$

$$B_{w\_jk} = B_{w\_0k} + B_{c2}(\beta_{00jk} - \gamma_{000k}) + f(\text{Year}_{jk}) + U_{Bw\_jk} \qquad U_{Bw\_jk} \sim N(0, \tau_{B_w})$$
$$(2 - 3c')$$

In Equations 2 – 3a′, 2 – 3b′, and 2 – 3c′, the time-metric variable, Year_{jk}, can take various forms such as linear, quadratic, a piecewise (i.e., discontinuous), or a general saturated form.

This reflects our assumption that the three growth parameters that capture individual growth within cohorts in schools would not necessarily show a similar pattern from cohort to cohort. For example, in some schools or in schools under certain interventions, the initial status of successive cohorts may increase, while the cohort mean rate of growth tends to remain at the same level, yet the

within-cohort initial status/growth rate relationships decrease over cohorts. This means that, though the schools might enroll a student body that consists of increasingly high-achieving students over years, how much or how rapidly they learn does not change much in these schools. However, by adding the last dimension (i.e., the within-cohort initial status/growth rate relationships), one can find that the successive cohorts that initially consist of increasingly high-achieving students exhibit a greater tendency of narrowing initial gaps in achievement over grades in these schools.

As noted earlier, the time series that is examined at the cohort level can be used for monitoring schools; in evaluation studies, this provides a basis for an ITS design. For example, this approach can be readily adapted to examine whether the onset of an intervention changes achievement patterns of schools in terms of the three parameters. A time-metric variable can be coded to indicate the onset of an intervention of interest and included as shown in the above three equations. At this juncture, we would like to point out that, both in school monitoring and in ITS evaluation studies, this simultaneous estimation framework is very advantageous in that the estimation takes into account correlations among three growth parameters, that inferences are based on latent variables, and that accurate *SE* associated with parameters are naturally incorporated in the analysis.

Lastly, the following level 4 (between-school) model estimates grand means of all three growth parameters and specifies two LVRs at the school level.

$$\gamma_{000k} = \theta_{0000} + V_{\gamma000k} \qquad\qquad V_{\gamma000k} \sim N(0, \tau_{\gamma000}) \quad (2-4a)$$

$$\gamma_{100k} = \theta_{1000} + B_b(\gamma_{000k} - \theta_{0000}) + V_{\gamma100k} \qquad V_{\gamma100k} \sim N(0, \tau_{\gamma100}) \quad (2-4b)$$

$$B_{w\_0k} = B_{w\_00} + B_{w\_01}(\gamma_{000k} - \theta_{0000}) + V_{Bw\_0k} \qquad V_{Bw\_0k} \sim N(0, \tau_{B_{w\_0}}) \quad (2-4c)$$

θ_{0000} is the grand mean initial status, and θ_{1000} is the grand mean rate of change across cohorts and schools. In Equation $2-4b$, the LVR coefficient, B_b, captures the relationship between school mean initial status and school mean growth rate. In Equation $2-4c$, B_{w_01} is an LVR coefficient capturing the extent to which differences in school mean initial status relates to differences in the within-school relationship between a student's initial status and his or her rate of growth.

The random effects, $V_{\gamma000k}, V_{\gamma100k}$, and V_{Bw_0k}, are assumed to be multivariate and normally distributed with mean 0 and the following variance–covariance matrix:

$$T_V = \begin{pmatrix} \tau_{\gamma_{000}} & 0 & 0 \\ 0 & \tau_{\gamma_{100}} & \tau_{\gamma_{100}, B_{w\_0}} \\ 0 & \tau_{B_{w\_0}, \gamma_{100}} & \tau_{B_{w\_0}} \end{pmatrix}. \qquad (2-4d)$$

TABLE 4.

Model 2: LVR-HM4—Estimating Within-School Initial Status/Gain Slopes (B_{w_jk}), Cohort Initial Status/Gain Slope (B_{c1}), Cohort Initial Status/B_{w_jk} Slope, and Between-School Initial Status/Gain Slope (B_b)

	Estimate	SE	95% Interval	Median
Fixed effects				
Grand mean initial status (θ_{0000})	189.3	1.48	[186.4, 192.3]	189.3
Grand mean gain (θ_{1000})	30.2	0.47	[29.3, 31.2]	30.2
Between-school initial status/gain slope (B_b)	0.194	0.033	[0.13, 0.26]	0.194
Mean within-school initial status/gain slope (B_{w_00})	0.022	0.017	[−0.011, 0.055]	0.022
School mean initial status/B_{w_jk} slope(B_{w_01})	−0.005	0.001	[−0.008, −0.003]	−0.005
Cohort initial status/gain slope (B_{c1})	−0.321	0.131	[−0.59, −0.07]	−0.319
Cohort initial status /B_{w_jk} slope (B_{c2})	−0.001	0.006	[−0.012, 0.010]	−0.001
Variance components				
1. Level 3 variance: Between-cohort (T_u)				
Initial status ($\tau_{\beta00}$)	8.19	1.70	[5.17, 11.82]	8.09
Gain ($\tau_{\beta10}$)	6.35	1.24	[4.21, 9.01]	6.26
Initial/gain slope (τ_{B_w})	0.003	0.001	[0.002, 0.006]	0.003
Cov (τ_{β_{10},B_w})	−0.008	0.031	[−0.070, 0.053]	−0.008
2. Level 4 variance: Between-school (T_v)				
Initial status($\tau_{\gamma000}$)	158.4	28.19	[112.3, 222.1]	155.4
Gain ($\tau_{\gamma100}$)	7.01	1.789	[4.076, 11.05]	6.82
Initial/gain slope ($\tau_{B_{w\_0}}$)	0.007	0.002	[0.000, 0.012]	0.007
Cov ($\tau_{\gamma_{100},\,B_{w\_0}}$)	0.058	0.046	[−0.026, 0.155]	0.056

Note. LVR-HM4 = latent variable regression four-level hierarchical model.

With respect to the off-diagonal elements of T_v, we assume that $Cov(V_{\gamma000k}, V_{\gamma100k}) = 0$ and $Cov(V_{\gamma000k}, V_{Bw\_0k}) = 0$ since the initial status for school k (γ_{000k}) is employed as a predictor in Equations 2 – 4b and 2 – 4c. As in Equation 2 – 3d, we also define the submatrix, $T_{V(2,3)}$ as the variance–covariance matrix of the latter two random effects (i.e., $V_{\gamma100k}$, V_{Bw_0k}).

We present the results for Model 2 in Table 4. Figure 4 is a graphical representation of the results shown in Table 4, displaying the expected growth parameters for each cohort in three different schools with mean initial status values that are, respectively, approximately 26 points below the mean (i.e., approximately 2 *SD*s), close to the mean, and 26 points above the mean. The district average values denoted in triangles are also shown for reference points. Note that all the growth parameters for these three schools are estimated within a Gibbs sampler and that the point estimate (posterior mean
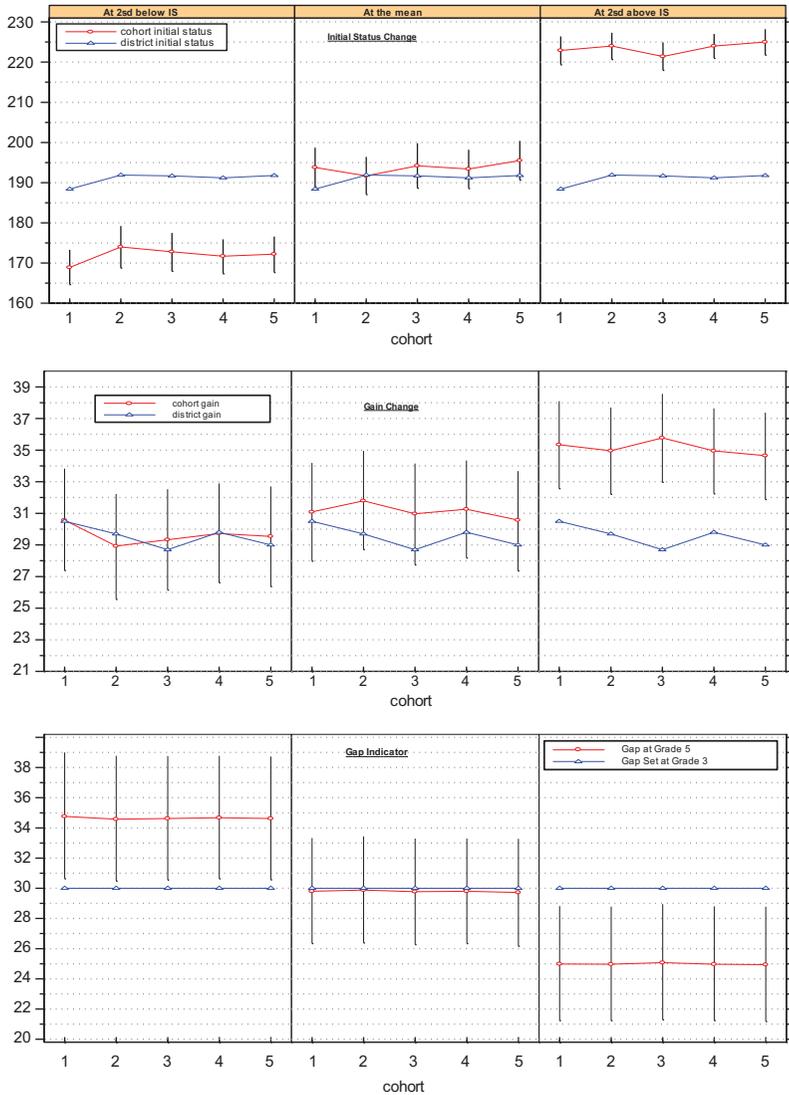
FIGURE 4. *Estimated school's three growth performance indicators: Cohort initial status* (β_{00jk}), *cohort gain* (β_{10jk}), *and cohort gap indicator* (B_{w_jk}) *using the latent variable regression four-level hierarchical model.*

shown with circles) and its 95% interval for each growth parameter are plotted.

The grand mean initial status is 189.3 and the grand mean gain is 30.2, which is similar to the results from Model 1, as would be expected. As to the LVR

coefficients, the estimated between-school initial status/gain slope (B_b) is .194, and its 95% interval is between .13 and .26. This indicates that on average, schools with higher initial status gain more than schools with lower initial status. In the middle panel of Figure 4, as you move from left to right (from schools with low initial status to middle initial status and to high initial status), the school average gain increases regardless of cohorts (the lines indicating school esti-mates—connecting circles—move up from left to right).

The between-cohort initial status/gain slope (B_{c1}) is $-.321$, and its 95% inter-val does not contain 0, which indicates the higher the initial status of cohorts, the lower the gain. Comparison of the top panel (initial status) and the middle panel (gain) of Figure 4 shows that high initial status cohorts tend to go with lower gain, whereas low initial status cohorts tend to go with higher gain. If we flip the cohort gain plot upside down, then the patterns over successive cohorts become extremely similar to those of the cohort initial status. The aforementioned sig-nificant negative between-cohort initial status/gain slope (i.e., B_{c1}) captures this strong relationship.

As for the gap-in-time parameter, although the overall student's initial sta-tus/gain slope (B_{w_00}) is not appreciably different from 0, the significant neg-ative estimate of B_{w_01}, capturing the relationship between school mean initial status and the within-school initial status/gain slope, indicates that low initial status students tend to gain more than high initial status students when they are in a high mean initial status school. In other words, the initial gap in achieve-ment tends to decrease more in a high mean initial status school than in a low mean initial status school.

The bottom panel of Figure 4 represents this result by addressing a question: What is the achievement gap at Grade 5 between two students initially 30 points apart at Grade 3 (i.e., 15 points above and 15 points below the school mean initial status; the 30-point gaps at Grade 3 appear in the horizontal reference line and are labeled as "gap set at Grade 3") in these three different schools? Note that these three schools are hypothetical schools simulated from the fitted model. To estimate the expected gap at Grade 5, the gains for two students in cohort j in school k who are, respectively, 15 points above and 15 points below the school mean initial status are estimated based on estimated parameters, and we also estimate the final status to each cohort. The results are plotted in the circle line.

As can be seen in the resulting circle line, the 30-point initial gap becomes larger by as much as approximately 5 points in the low mean initial status school (left panel). In the school with its mean initial status similar to the district average (middle panel), the initial gap remains similar between Grades 3 and 5. Finally, in the high mean initial status school, the initial 30-point gap is reduced by up to 5 points (right panel). The bottom panel also shows that the gap indicators fluctuate minimally across cohorts as one can see from almost straight lines connecting different cohorts. These results are also shown by the LVR coefficient, capturing

the relationship between cohort mean initial status and within-cohort and within-school initial status/gain slope ($B_{c2} = -0.001$) that is negligible. In addition, the variance of the gap-in-time parameter, that is, initial status/gain slope, in cohort level (τ_{B_w}) is 0.003, while the variance of initial status/gain slope in school level 0.007. Thus, the intraclass correlation of the gap-in-time parameter is .7, which indicates that the gap-in-time parameter varies far more between schools than between cohorts within schools.

Model 3: LVR-HM4 Estimating Differences Between NCLB and Non-NCLB Cohorts

With the level 1 model being the same as that of Model 2, a student characteristic variable flagging eligibility of FRL status (students with FRL status coded 1; otherwise coded 0) is included in the following level 2 model (Equations 3 – 2a and 3 – 2b). This variable is centered on its grand mean, so that the growth parameters (student initial status and student gain) retain the same meanings as previous models.

$$\pi_{0ijk} = \beta_{00jk} + \beta_{01}(FRL_{ijk} - \overline{FRL..}) + r_{\pi0ijk} \qquad r_{\pi0ijk} \sim N(0, \tau_{\pi0jk})$$

$$(3 - 2a)$$

$$\pi_{1ijk} = \beta_{10jk} + B_{w\_jk}(\pi_{0ijk} - \beta_{00jk}) + \beta_{11}(FRL_{ijk} - \overline{FRL..}) + r_{\pi1ijk} \quad r_{\pi1ijk} \sim N(0, \tau_{\pi1jk})$$

$$(3 - 2b)$$

The level 3 model (i.e., the between-cohort, within-school model) is a key model in that cohort-to-cohort changes are compared between pre-NCLB and post-NCLB cohorts. The indicators for NCLB cohorts, $NCLB_{1jk}$, and $NCLB_{2jk}$ are included in Equations 3 – 3a, 3 – 3b, and 3 – 3c. $NCLB_{1jk}$ takes a value of 0 for the first four cohorts and 1 for the last cohort in Equation 3 – 3a where we can contrast initial status of third-grade students who entered Grade 3 in the pre- or post-NCLB era. However, we code the $NCLB_{2jk}$ variable 0 for the first two cohorts and 1 for the last three cohorts in Equations 3 – 3b and 3 – 3c. In using this variable, we consider cohorts that were Grade 4 or Grade 5 in the beginning year of NCLB as the cohort in the post-NCLB era, as well as the last cohort that were was in Grade 3 in the beginning year of NCLB. For the cohorts where NCLB was implemented in the middle of the study period (students who were Grade 4 or Grade 5), if there is an effect of NCLB, we would expect a negligible difference between pre-NCLB and post-NCLB cohorts in the initial status but significant differences in growth rates or the gap-in-time parameters.

$$\beta_{00jk} = \gamma_{000k} + \gamma_{001k} NCLB_{1jk} + U_{\beta00jk} \qquad U_{\beta00jk} \sim N(0, \tau_{\beta00})$$

$$(3 - 3a)$$

$$\beta_{10jk} = \gamma_{100k} + \gamma_{101k}\,\text{NCLB}_{2jk} + B_{c1}(\beta_{00jk} - \gamma_{000k}) + U_{\beta 10jk} \qquad U_{\beta 10jk} \sim N(0, \tau_{\beta 10})$$

$$(3-3b)$$

$$B_{w\_jk} = B_{w\_0k} + B_{w\_1k}\,\text{NCLB}_{2jk} + B_{c2}(\beta_{00jk} - \gamma_{000k}) + U_{Bw\_jk} \qquad U_{Bw\_jk} \sim N(0, \tau_{B_w})$$

$$(3-3c)$$

Note, however, that both NCLB_{1jk} and NCLB_{2jk} are centered on their respective group means. Thus, the meaning of intercepts remains the same as in Model 2. For example, γ_{000k} represents mean initial status for school k, while γ_{100k} represents mean growth rate for school k. The other three coefficients, γ_{001k}, γ_{101k}, and B_{w_1k}, capture the differences between pre- and post-NCLB cohorts, respectively, in the mean initial status, mean gain, and mean within-cohort initial status/gain slope for school k. These three coefficients are assumed to be random, so the differences can vary across schools.

The school-level model below (Equations $3-4a$ through $3-4f$) includes a variable that flags whether a school met AYP criteria or not. Note that AYP decision for an individual school was based on the percentage of fourth graders scoring proficient on the state's standards-based assessment not based on the outcome, ITBS score. AYP status is potentially a time-varying covariate; however, in our data, we used the AYP variable as a time-invariant variable because the AYP status was based on the 2002–2003 school year, when the sampled Cohort 5 students were in the fourth grade. Only six schools in the sample did not meet AYP. Thus, we coded NonAYP_k 0 for AYP schools and 1 for non-AYP schools, so that intercepts estimated in this level would represent a majority of schools (i.e., schools meeting AYP) in this study sample. In addition, this variable is centered on its mean which results in intercepts in the equations below representing the overall mean of the sample.

$$\gamma_{000k} = \theta_{0000} + \theta_{0001}\,\text{NonAYP}_k + V_{\gamma 000k} \qquad V_{\gamma 000k} \sim N(0, \tau_{\gamma 000})$$

$$(3-4a)$$

$$\gamma_{001k} = \theta_{0010} + B_{b1}(\gamma_{000k} - \theta_{0000}) + \theta_{0011}\,\text{NonAYP}_k + V_{\gamma 001k} \qquad V_{\gamma 001k} \sim N(0, \tau_{\gamma 001})$$

$$(3-4b)$$

$$\gamma_{100k} = \theta_{1000} + B_{b2}(\gamma_{000k} - \theta_{0000}) + \theta_{1001}\,\text{NonAYP}_k + V_{\gamma 100k} \qquad V_{\gamma 100k} \sim N(0, \tau_{\gamma 100})$$

$$(3-4c)$$

$$\gamma_{101k} = \theta_{1010} + B_{b3}(\gamma_{000k} - \theta_{0000}) + \theta_{1011}\,\text{NonAYP}_k + V_{\gamma 101k} \qquad V_{\gamma 101k} \sim N(0, \tau_{\gamma 101})$$

$$(3-4d)$$

$$B_{w\_0k} = B_{w\_00} + B_{w\_01}(\gamma_{000k} - \theta_{0000}) + B_{w\_02}\,\text{NonAYP}_k + V_{Bw\_0k} \qquad V_{Bw\_0k} \sim N(0, \tau_{B_{w\_0}})$$

$$(3-4e)$$

$$B_{w\_1k} = B_{w\_10} + B_{w\_11}(\gamma_{000k} - \theta_{0000}) + B_{w\_12}\text{NonAYP}_k + V_{Bw\_1k} \qquad V_{Bw\_1k} \sim N(0, \tau_{B_{w\_1}})$$

$$(3 - 4f)$$

As can be seen in Table 5, the post-NCLB cohort has significantly higher mean initial status than the pre-NCLB cohorts by approximately 1.96. However, the estimate of the difference in cohort mean gain between pre- and post-NCLB is 0.27, and its 95% interval includes the value of 0, which indicates that there is no statistical difference between the two groups (pre-NCLB cohorts vs. post-NCLB cohorts). Likewise, there is no statistical difference ($B_{w\_10} = 0.024$) in the gap-in-time indicator between pre- and post-NCLB cohorts.

Between AYP and non-AYP schools, although the coefficients indicating the differences are in the expected direction with AYP schools on average outperforming non-AYP schools, they are not significant partly due to the low statistical precision associated with the small sample size (as noted earlier there were only six non-AYP schools). For example, AYP schools have higher school mean initial status than non-AYP schools by 6.66, but the 95% interval for θ_{0001} does include the value of 0. Likewise, the difference in the school mean gain between AYP and non-AYP schools is negligible.

In addition, the estimates and 95% intervals of all the LVR coefficients are very similar to those from Model 2. The LVR coefficient of school mean initial status on school mean gain ($B_{b2} = 0.19$) indicates that schools with higher mean initial status tend to gain more than those with lower mean initial status. Furthermore, the LVR coefficient of within-school, cohort mean initial status is negatively associated with within-school, cohort gain with the magnitude greater in the same direction ($B_{c1} = -0.50$). This suggests that cohorts with lower mean initial status tend to gain more that cohorts with higher mean initial status, resulting in smaller gaps in reading proficiency between cohorts within schools over years.

Discussion

Longitudinal tracking of educational effectiveness or accountability indicators and comparison of such trends have become a significant part of educational research. In schools' academic performance, tracking individual students over time is always of great interest, in that how students learn as they progress across grades is one of the key outcomes in schooling. At the same time, tracking cohorts over time can be of special interest as well, as how, for example, successive cohorts of third graders perform can carry important implications: If a certain intervention was in place for third graders, one would not only be interested in seeing whether the first cohort of third graders were doing better as they progress to upper grades but also whether the effects would persist over successive cohorts (and whether the intervention had sustained effects on student achievement regardless of student entry).

TABLE 5.

Model 3: LVR-HM4—Comparing Performance Between the Pre- and Post-NCLB Era Cohorts and AYP School Versus Non-AYP School

	Estimate	SE	95% Interval	Median
Fixed effects				
Grand mean initial status (IS) (θ_{0000})	189.3	1.08	[187.1, 191]	189.3
IS difference: AYP vs. non-AYP school (θ_{0001})	−6.66	3.98	[−14.45, 1.22]	−6.66
IS difference: pre- vs. post-NCLB (θ_{0010})	1.96	0.658	[0.646, 3.24]	1.95
IS difference in pre- vs. post-NCLB: AYP vs. non-AYP schools in (θ_{0011})	−2.42	2.51	[−7.36, 2.51]	−2.42
Effect of school mean IS on IS diff in pre- vs. post-NCLB (B_{b1})	0.17	0.08	[0.01, 0.33]	0.17
Grand mean gain (θ_{1000})	30.4	0.43	[29.5, 31.2]	30.4
Gain difference: AYP vs. non-AYP school (θ_{1001})	−1.48	1.50	[−4.39, 1.47]	−1.48
Effect of school IS on school mean gain (B_{b2})	0.193	0.047	[0.102, 0.286]	0.192
Gain difference: pre- vs. post-NCLB (θ_{1010})	0.270	0.75	[−1.16, 1.79]	0.255
Gain difference in pre- vs. post-NCLB: AYP vs. non-AYP schools (θ_{1011})	0.100	2.64	[−5.17, 5.17]	0.151
Effect of school mean IS on gain diff. in pre- vs. post-NCLB (B_{b3})	0.03	0.09	[−0.15, 0.21]	0.03
Grand mean within-school initial status/gain slope (Bw_{00})	0.000	0.017	[−0.033, 0.031]	0.000
Bw_{jk} slope difference: AYP vs. non-AYP school (Bw_{02})	−0.046	0.058	[−0.160, 0.068]	−0.046
Effect of school mean IS on Bw_{jk} slope (Bw_{01})	−0.007	0.002	[−0.010, −0.004]	−0.007
Bw_{jk} slope difference: pre- vs. post-NCLB (Bw_{10})	0.024	0.033	[−0.04, 0.088]	0.024
Bw_{jk} slope difference in pre- vs. post-NCLB: AYP vs. non-AYP schools (Bw_{12})	−0.09	0.13	[−0.34, 0.16]	−0.09
Effect of school mean IS on Bw_{jk} slope difference in pre- vs. post-NCLB (Bw_{11})	0.01	0.004	[−0.001, 0.02]	0.01
Cohort initial status/gain slope (Bc_1)	−0.50	0.176	[−0.853, −0.161]	−0.498
Cohort initial status/Bw_{jk} slope (Bc_2)	−0.004	.007	[−0.018, 0.009]	−0.004
Effect of Free lunch status on IS (β_{01})	−13.88	0.41	[−14.7, −13.1]	−13.9

(continued)

TABLE 5. *(continued)*

	Estimate	*SE*	95% Interval	Median
Effect of Free lunch status on gain (β_{11})	−2.35	0.37	[−3.08, −1.61]	−2.35
Variance components				
1. Level 3 variance: between-cohort				
Initial status ($\tau_{\beta 00}$)	5.10	1.4	[2.6, 8.09]	5.02
Gain ($\tau_{\beta 10}$)	5.23	1.17	[3.21, 7.77]	5.14
Initial/gain slope (τ_{B_w})	0.003	0.001	[0.002, 0.006]	0.003
Cov ($u_{\beta 10}, u_{Bw}$)	−0.019	0.027	[−0.075, 0.033]	−0.019
2. Level 4 variance: between-school				
Initial status ($\tau_{\gamma 000}$)	81.9	15.1	[57.1, 115.9]	80.2
Initial status diff. ($\tau_{\gamma 001}$)	4.91	2.57	[1.64, 11.4]	4.34
Gain ($\tau_{\gamma 100}$)	7.07	1.82	[4.13, 11.2]	6.87
Gain difference ($\tau_{\gamma 101}$)	7.67	4.03	[2.51, 17.8]	6.80
Initial/gain slope ($\tau_{B_{w\_0}}$)	0.006	0.002	[0.003, 0.010]	0.006
Initial/gain slope diff. ($\tau_{B_{w\_1}}$)	0.004	0.002	[0.001, 0.010]	0.003
Cov ($V_{\gamma 001}, V_{\gamma 100}$)	0.746	1.61	[−2.37, 4.11]	0.70
Cov ($V_{\gamma 001}, V_{\gamma 101}$)	2.70	2.64	[−1.06, 9.18]	2.21
Cov ($V_{\gamma 001}, V_{Bw\_0}$)	0.005	0.040	[−0.074, 0.088]	0.005
Cov ($V_{\gamma 001}, V_{Bw\_1}$)	−0.013	0.056	[−0.138, .096]	−0.009
Cov ($V_{\gamma 100}, V_{\gamma 101}$)	1.74	1.85	[−1.70, 5.70]	1.64
Cov ($V_{\gamma 100}, V_{Bw\_0}$)	0.030	0.042	[−0.050, 0.118]	0.029
Cov ($V_{\gamma 100}, V_{Bw\_1}$)	−0.021	0.055	[−0.136, 0.086]	−0.018
Cov ($V_{\gamma 101}, V_{Bw\_0}$)	−0.011	0.049	[−0.114, 0.084]	−0.010
Cov ($V_{\gamma 101}, V_{Bw\_1}$)	−0.017	0.072	[−0.175, 0.122]	−0.012
Cov (V_{Bw_0}, V_{Bw_1})	0.000	0.001	[−0.003, 0.002]	0.000

Note. LVR-HM4 = latent variable regression four-level hierarchical model; NCLB = No Child Left Behind; AYP = adequate yearly progress.

In evaluation contexts, tracking of trends is readily associated with the study design of ITS, which can be used in settings where the intervention of interest has been rolled out without intentional assignment, and there were extant data that can be linked from before and after the initiation of the intervention. The ITS design often tracks successive cohorts over years in terms of the outcomes of interest. In this article, this approach was applied to an ITS evaluation of the influence of NCLB in elementary schools in a large urban school district. As NCLB put a special emphasis on all students gaining proficiency as well as on reducing achievement gaps among students, the dimensions of student growth that we wish to examine are highly relevant to and aligned with the purposes of NCLB (see Choi et al., 2007). Whether students tend to gain academic proficiency is represented by student initial status and gain parameters, and whether initial gaps among students tend to decrease over time is represented by gap-in-time parameters.

Many studies examining the impact of NCLB have assessed outcomes by percentage proficient or change in percentage proficient and by the same parameters by subgroups such as race or English language learners. This is partly because available data (e.g., National Assessment of Educational Progress [NAEP]) were based on aggregate data (e.g., state mean percentage proficient by subgroup of interest) and also because percentage proficient is often better communicated to the public. In contrast, this article fully takes into account the extant data structure drawing on the district's annual assessment systems over multiple academic years and estimates latent variables that address the same issues but from a slightly different perspective. The proficiency issue is addressed by where students are at Grade 3 and how much students gain for upper elementary grades (from Grade 3 to Grade 5) in a low-stakes reading assessment instead of percentage proficient in end-of-year state assessments.

The findings of this study indicate that for reading achievement, there was no statistical difference between pre-NCLB and post-NCLB cohorts in terms of gain from Grade 3 to Grade 5. There is evidence of significant increase in the initial status at Grade 3, but this is a relatively small amount of gain (approximately 1.3–1.8 points), which is less than 0.1 *SD*s of the outcome. Also, there is no evidence of significant narrowing of achievement gaps among students as they progress from Grades 3 to 5. Speaking of the gap-in-time parameter, there is evidence of magnifying achievement gaps over grades between students and between schools alike. In contrast, there is evidence that exhibit the direction of narrowing achievement gaps over years between cohorts, within schools. In summary, even as early as the upper elementary school grades, there is the tendency of students who already perform well and learn increasingly more rapidly. This is shown both at the student and at the school levels. Thus, similarly elementary schools who already perform well on average show their students learning increasingly more rapidly. This in a sense indicates that the proficiency distributions of students or schools become more inequitable over time but also may suggest that this may be a natural course of learning across students and schools in a number of core subject areas including the reading subject in this article.

However, we consistently find evidence of narrowing proficiency gaps between cohorts within schools. This, we believe, is promising news because this indicates that there are schools that bring their students to more similar proficiency levels despite initially disparate proficiency levels across cohorts. Our approach with LVR-HM4 pays attention to not only the gain but also the gap-in-time parameter. By attending to this additional dimension, we were able to suggest that there exist a number of effective schools in our sample of 74 schools. Some schools that begin with higher initial status also produce more equitable distributions of student proficiency levels regardless of student entry (to be exact, different proficiency levels at Grade 3 in this study) and also

facilitate above-average rates of student learning. In these schools, students not only perform higher; even when incoming cohorts have heterogeneous reading proficiency, by the end of Grade 5, the distribution of their proficiency becomes more equitable.

We hope to have demonstrated the value and the motivation of the LVR-HM4 with the ITS evaluation of NCLB in a large urban district. It entails treating the coefficients of time-metric variables in the cohort level to be random (i.e., to vary across schools) instead of treating them to be fixed. As shown in the cohort level of the final model in this article (shown in Equations 3 – 3a through 3 – 3c), the coefficients of NCLB indicators were treated as random. Such models can examine whether the effect of NCLB or, in general, the effect of an intervention in ITS evaluations would vary across schools. Also, in the presence of detailed data measuring school-level implementation and/or characteristics, such models can address questions as to the extent to which the implementation of key intervention components is associated with the ITS treatment effect on various dimensions of student growth and/or as to the settings where the ITS treatment effect is magnified or diminished.

In school effect research based on students' academic progress, some students change schools during the period in which we measure student progress. While a detailed examination of student mobility is beyond the scope of this article, we would like to lay out building blocks for extending LVM-HMs with multiple-cohort data to account for student mobility existing in the data. For students who change schools, alternative data structure arises because those students are no longer nested within one school; rather, they have multiple memberships in different schools. Multiple membership multilevel models (Browne, Goldstein, & Rasbash, 2001; Cafri, Hedeker, & Aarons, 2015; Grady & Beretvas, 2010; Leckie, 2009; Luo & Kwok, 2012) can appropriately model a cross-classified structure of data, that is, students in multiple schools, by using weights that represent the degree of membership of a student in different schools. For example, in a 4-year study of school effects, if a student stayed in one school for 2 years and then transferred to another school in which he or she stayed for the subsequent 2 years, then the weights of the student would be 0.5 for the former school and 0.5 for the latter school. In applying this approach to LVR-HMs with multiple-cohort data, LVR-HMs use a weight matrix that expands to each student in each cohort, with fractions corresponding to the length of stay for those who have membership in different schools, and with 1s for those who have not changed schools. Although a large weight matrix with multiple-cohort data may become computationally cumbersome, a fully Bayesian approach using Gibbs sampler is a viable option for estimating such models.

Notes

1. To assess the convergence of the Gibbs sampler, we examined trace plots, autocorrelation function plots, and Gelman–Rubin statistics (Gelman & Rubin, 1992). For all models, we ran two chains using different starting values with a burn-in period of 2,000 iterations, with an additional 30,000 iterations saved for posterior calculation. The results based on the two chains (e.g., posterior means and 95% intervals) are extremely similar for all models. All results presented in Tables 3 through 5 were based on the pooled two chains, sample of 60,000 deviates. Code for estimating the models is available from the authors upon request.

2. In specifying priors, we used diffuse priors for all the fixed effects and variance components, such that data dominate our inferences. Specifically, for the fixed effects, we specified normal priors with a mean 0 and extremely low precision, $N \sim (0, 1.0\text{E-}5)$, which are functionally equivalent to uniform priors. We placed uniform priors on all scalar variances (presented a few, e.g., $\tau_{\pi 0jk}$ and $\tau_{\pi 1jk}$ in Equations 2 – 2a and 2 – 2b; $\tau_{\beta 00}$ in Equation 2 – 3a; $\tau_{\beta 10}$ in Equation 2 – 4b). As placing uniform priors on scalar variances translates to placing particular Pareto priors on scalar precisions in models in WinBUGS (Spiegelhalter et al., 2003), thus we used Pareto priors, Pareto(1, .0001). Finally, in specifying priors for a random effects variance–covariance matrix (see, e.g., $T_{U(2,3)}$ in Equation 2 – 3d and $T_{V(2,3)}$ in Equation 2 – 4d), we used inverse Wishart priors with small degrees of freedom (ν) and scale matrix S and followed the method to specify the degrees of freedom and scale matrix, as described in Choi and Seltzer (2010, pp. 85–86).

References

Browne, W. J., Goldstein, H., & Rasbash, J. (2001). Multiple membership multiple classification (MMMC) models. *Statistical Modelling*, *1*, 103–124.

Bryk, A., Thum, Y. M., Easton, J. Q., & Luppescu, S. (1998). Assessing school academic productivity: The case of Chicago school reform. *Social Psychology of Education*, *2*, 103–142.

Cafri, G., Hedeker, D., & Aarons, G. A. (2015). An introduction and integration of cross-classified, multiple membership, and dynamic group random-effects models. *Psychological Methods*, *20*, 407.

Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*, *24*, 305–331.

Choi, K., & Seltzer, M. (2010). Modeling heterogeneity in relationships between initial status and rates of change: Treating latent variable regression coefficients as random coefficients in a three-level hierarchical model. *Journal of Educational and Behavioral Statistics*, *35*, 54–91.

Choi, K., Seltzer, M., Herman, J., & Yamashiro, K. (2007). Children left behind in AYP and non-AYP schools: Using student progress and the distribution of student gains to validate AYP. *Educational Measurement: Issues and Practice*, *26*, 21–32.

Gelman, A., & Rubin, D. B. (1992). Inferences from iterative simulation using multiple sequences. *Statistical Science*, *7*, 457–472.

Grady, M. W., & Beretvas, S. N. (2010). Incorporating student mobility in achievement growth modeling: A cross-classified multiple membership growth curve model. *Multivariate Behavioral Research*, *45*, 393–419.

Goldschmidt, P., & Choi, K. (2007). *The practical benefits of growth models for accountability and the limitations under NCLB* (Policy Brief 9). Los Angeles: University of California, Los Angeles, Center for Research on Evaluation, Standards, and Student Testing.

Haertel, E. H. (2005). *Using a longitudinal student tracking system to improve the design for public school accountability in California*. Palo Alto, CA: Stanford University. Retrieved from http://web.stanford.edu/∼haertel/Haertel-Value-Added.pdf

Hanushek, E. A., & Raymond, M. E. (2003). Lessons about the design of state accountability systems. In P. E. Peterson & M. R. West (Eds.), *No child left behind? The politics and practice of accountability* (pp. 127–151). Washington, DC: Brookings.

Leckie, G. (2009). The complexity of school and neighbourhood effects and movements of pupils on school differences in models of educational achievement. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *172*, 537–554.

Leckie, G., French, R., Charlton, C., & Browne, W. (2014). Modeling heterogeneous variance-covariance components in two-level models. *Journal of Educational and Behavioral Statistics*, *39*, 307–332.

Leckie, G., & Goldstein, H. (2009). The limitations of using school league tables to inform school choice. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *172*, 835–851.

Luo, W., & Kwok, O. M. (2012). The consequences of ignoring individuals' mobility in multilevel growth models: A Monte Carlo Study. *Journal of Educational and Behavioral Statistics*, *37*, 31–56.

McCaffrey, D., Lockwood, J. R., Koretz, D., Louis, T., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, *29*, 67–101.

No Child Left Behind Act of 2001, Pub. L. No. 107–110, 115 Stat. 1425 (2002).

Ponisciak, S. M., & Bryk, A. (2005). Value-added analysis of the Chicago public schools: An application of hierarchical models. In R. Lissitz (Ed.), *Value-added modeling: Issues with theory and applications* (pp. 40–81). Maple Grove, MN: JAM Press.

Sanders, W. L., Saxton, A. M., & Horn, S. P. (1997). The Tennessee value-added assessment system: A quantitative outcomes-based approach to educational assessment. In J. Millman (Ed.), *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* (pp. 137–162). Thousand Oaks, CA: Corwin Press.

Seltzer, M., Choi, K., & Thum, Y. M. (2003). Examining relationships between where students start and how rapidly they progress: Implications for conducting analyses that help illuminate the distribution of achievement within schools. *Educational Evaluation and Policy Analysis*, *25*, 263–286.

Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2003). WinBUGS: Windows version of Bayesian inference using Gibbs sampling (version 1.4) [Computer software]. Cambridge, England: MRC Biostatistics Unit, Cambridge University.

Thum, Y. M. (2006). Designing gross productivity indicators: A proposal for connecting accountability goals, data and analysis. In R. Lissitz (Ed.), *Longitudinal and value-added models of student performance* (pp. 436–479). Maple Grove, MN: JAM Press.

Willett, J. B. (1989). Questions and answers in the measurement of change. In E. Z. Rothkopf (Ed.), *Review of research in education* (Vol. 15, pp. 345–422). Washington, DC: American Educational Research Association.

Authors

KILCHAN CHOI is an associate director at the Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California, Los Angeles, 300 Charles E. Young Drive, GSE& IS Bldg. 3rd floor, Los Angeles, CA 90095; email: kchoi@cresst.org. His research interests are latent variable regression modeling in hierarchical models, Bayesian analysis with Markov chain Monte Carlo estimation strategies, school effectiveness, and longitudinal data analysis.

JINOK KIM is an independent consultant; email: jinokkimcho@gmail.com. Her areas of interest include applications of hierarchical models in longitudinal analysis and multi-site studies of programs and practices, causal inferences, and policy evaluations based on large-scale observational data.