Journal of Educational and Behavioral Statistics
October 2011, Vol. 36, No. 5, pp. 616–637
DOI: 10.3102/1076998610396887

The Persistence of School-Level Value-Added

Derek C. Briggs
Jonathan P. Weeks
University of Colorado at Boulder

Using longitudinal data for an entire state from 2004 to 2008, this article describes the results from an empirical investigation of the persistence of value-added school effects on student achievement in reading and math. It shows that when schools are the principal units of analysis rather than teachers, the persistence of estimated school effects across grades can only be reasonably identified by placing strong constraints on the variable persistence model implemented by Lockwood, McCaffrey, Mariano, and Setodji. In general, there are relatively strong correlations between the school effects estimated using these constrained models and a reference model that assumes full persistence. These correlations vary somewhat by grade and the underlying test subject. The results from this study indicate cautious support for previous findings that the assumption of full persistence for cumulative value-added effects may be untenable, and evidence is also presented, which indicates a strong interaction by test subject. However, the practical impact of violating the assumption of full persistence appears to be smaller in the context of schools than it is for teachers.

Keywords: *accountability; statistics; educational policy; high stakes testing; research methodology*

Introduction

In a special issue of the *Journal of Educational and Behavioral Statistics* devoted to the topic of value-added modeling of student achievement, McCaffrey, Lockwood, Koretz, Louis, and Hamilton (2004) introduced what is now known as the "variable persistence model" for longitudinal student outcomes. McCaffrey and colleagues demonstrated that many other value-added models used to estimate teacher effects,[1] school effects, or both could be expressed as restricted versions of their more general model. The key feature of this model

The research reported in this article was supported by a grant from the Carnegie Corporation. An earlier version of this article was presented at the annual meeting of the American Educational Research Association, March 27, 2008.

is that it relaxes an implicit assumption made in the value-added model developed for large-scale usage in educational accountability by William Sanders and colleagues known as the "layered model" (c.f. Ballou, Sanders, & Wright, 2004; Sanders, Saxton, & Horn, 1997). Namely, the layered model assumes that the contribution of a teacher to a student's future test score performance stays the same from year to year (i.e., persists[2]) even as a student is cumulatively exposed to instruction from new teachers in different classroom settings. Intuitively, this assumption seems implausible. In a later study, Lockwood et al., (2007) provided empirical evidence that the contribution of a teacher 2 or more years removed from a student's current level of test performance does not, in fact, persist with undiminished magnitude. The two practical upshots to this finding are that both the size and the precision of estimated teacher effects appear to be sensitive to the way that persistence is parameterized. Because the precision of estimated effects appears to be much greater under the variable persistence model relative to a model such as the layered model that assumes full persistence, highly effective or ineffective teachers are more likely to be distinguished from the "average" teacher.

McCaffrey et al. (2004) parameterized their model in a way that allows—at least in theory—for the estimation of teacher effects, school effects, or both. However, to our knowledge, there have been no applications of the variable persistence model to longitudinal data in which schools, rather than teachers, are the principal units of analysis. Conceptually, the same issues that have arisen in the estimation of teacher effects should apply to the estimation of school effects because the latter, one might argue, should subsume the former. Hence, if teacher effects do not fully persist over time, then neither should school effects. Our initial motivation for the present study was to evaluate this empirically by posing the following research question: To what extent do conclusions about school effectiveness change when the variable persistence model is used to estimate longitudinal school effects in comparison to a version of the model that assumes full persistence?

Yet while the concept of persistence has the same intuitive appeal with respect to schools as it does for teachers, in the school context, there are significant obstacles to the quantification of this concept using a statistical model. The principal obstacle is that of identifying the persistence parameters of interest. As we will show, given 5 years of longitudinal data, one would ideally estimate up to 10 unique persistence parameters. But when schools are the units of analysis, we argue that in most cases it will only be plausible to identify and estimate a single persistence parameter. Given this modeling constraint, some hard choices need to be made with regard to the parameterization of school effect persistence. In this article, we walk the reader through these choices in one specific empirical context and examine the sensitivity of conclusions one might reach about the magnitude and precision of school effects on the basis of these choices.

TABLE 1
Comparison of Grade 6 Cohort Samples to Population in State

	Reading Cohort		Math Cohort	
	Population	Sample	Population	Sample
Students	56,791	29,126	56,711	27,803
Schools	635	225	649	240
Female	49%	49%	49%	49%
White/Asian	65%	61%	66%	60%
Black/Hispanic	35%	39%	34%	40%
Free and reduced price lunch	37%	40%	36%	41%
Students with IEP	10%	10%	10%	10%
English language learner	12%	14%	16%	19%
Identified as "gifted"	10%	12%	11%	13%
Students with disability	10%	10%	10%	10%
<12 months in school district	16%	16%	16%	15%
Unstandardized test score mean (*SD*)	623 (67)	618 (70)	537 (77)	532 (79)
Mean score gain Grade 5 to Grade 6 (*SD*)	12.5 (37)	10.9 (38)	17.31 (38)	15.85 (38)

Note: The cohort of students taking the reading test was in Grade 6 as of 2006 while the cohort taking the math test were in Grade 6 as of 2007.

Data

The data for this study come from two longitudinal cohorts taken from the full population of students and schools in a midsized American state west of the Mississippi. The first cohort took the state's standardized assessment in reading over 5 years: in 2004 as fourth graders, and in 2008 as eighth graders. Because there were no Grade 4 tests in math administered until 2005, our second longitudinal cohort consists of students that took the state's standardized assessment in math over 4 years: as fourth graders in 2005 and as seventh graders in 2008. In other words, each cohort consists of a different set of students—the Grade 5 students taking the reading assessment were not the same as the Grade 5 students taking the math assessment. To hold constant one source of confounding in the analysis that follows (for reasons that we explain in the next section), we restricted both cohorts to those students who were enrolled in elementary schools with a Grade K–5 configuration and middle schools with a Grade 6–8 configuration. This left us with a sample of 29,126 students in our reading cohort, who attended roughly 547 different elementary schools and 225 different middle schools, respectively.[3] The numbers for the math cohort were 27,803 students attending 555 and 240 unique elementary and middle schools. Summary statistics that characterize our cohort samples and their comparability to the full population of students and schools in the state as of Grade 6 are presented in Table 1.

TABLE 2
Summary Statistics for Unconditional Growth Across Grades

	Reading		Math	
Grade	*M*	*SD*	*M*	*SD*
4	0	1	0	1
5	0.42	1.09	0.51	0.99
6	0.56	1.06	0.72	1.02
7	0.78	1.11	0.87	0.97
8	1.03	0.98	—	—

Note: Score *M*s and *SD*s were standardized relative to the scale score means in Grade 4.

The students in our restricted samples were somewhat more likely to be non-White, English language learners, and eligible for free and reduced lunch services than those students excluded from the analysis but not dramatically so. Our student and school sample also tended to have lower average test scores and Grade 5–6 score gains relative to the full population. The test scores in the subjects of reading and math, which serve as the outcome measures in our analyses come from responses to a mixture of multiple-choice and constructed response items. These scores were calibrated onto a vertical score scale by the state's test developer. The vertical scale is based on a common item nonequivalent groups linking design that was established by the state's test contractor in 2001. It was created by scaling each grade-specific test from 3 to 10 using an item response theory model and then linking the tests using the Stocking-Lord method (Stocking & Lord, 1983). Since the initial creation of this vertical scale, new test forms in math and reading have been administered at each grade. Item parameters and ability estimates from subsequent tests are horizontally equated so that they can be linked back to the base vertical scale.[4]

For ease of interpretation in the analysis that follows, test scores have been standardized relative to the mean and standard deviation of the Grade 4 tests for reading and math, respectively. Summary statistics for the resulting grade-specific growth trajectories are provided in Table 2 and serve as a useful frame of reference when interpreting the magnitude for estimates of school value-added.

The Variable Persistence Model

A variable persistence model for a single longitudinal test score outcome can be written as

$$Y_{it} = \mu_t + \sum_{t* \leq t} \alpha_{tt*}\theta_{t*} + \varepsilon_{it}. \tag{1}$$

In Equation 1, Y_{it} represents the test score for student i in year t, $t = \{1, \ldots, T\}$, and the parameter μ_t denotes the population mean for a given grade. The term ε_{it}

represents the test score residual associated with student i in year t. Under the variable persistence model, θ_{t*} and ε_{it} are assumed to be independent random variables, where $\varepsilon_{it} \sim N(\mathbf{0}, \mathbf{\Sigma})$ and $\theta_{t*} \sim N(\mathbf{0}, \tau)$. The two covariance matrices differ in that the former is unstructured while the latter is typically specified as a diagonal matrix, which implies the assumption that school effects are longitudinally uncorrelated. The vector θ_{t*} represents the collection of school effects for the current and preceding years (given that $t^* \leq t$). Note that for any given year and student, there will only be one associated school effect within the vector θ_{t*} (analogous to a scenario in which each school is parameterized as a dummy variable in a fixed effects regression model). Finally, since a unique parameter for θ_{t*} is associated with each year of the longitudinal cohort, these can also be conceptualized as "grade" effects.

The parameters α_{tt*}, which are of principal interest to us in this study, capture the persistence of the school effects θ_{t*}. The two different subscripts, t and $t*$, are used to distinguish the link between persistence parameters and school effects as the effects cumulate over time. As noted above, $t*$ is always $\leq t$. When $t* = t$, it will always be the case that $\alpha_{tt*} = 1$ because, by definition, there is no decay for the effect of a school on student achievement in a current year. However, when $t* < t$ and $\alpha_{tt*} = 1$, this reflects an assumption that the effect of a school on student test performance in a prior grade persists fully into a subsequent grade. Conversely, when $t* < t$ and $\alpha_{tt*} = 0$, only the current year test scores convey information about a school's effect on student achievement for a given grade. When $0 < \alpha_{tt*} < 1$, the effect of a school on student test performance in a prior grade diminishes in a subsequent grade. Finally, when $\alpha_{tt*} > 1$, the positive school effects in prior grades have positive effects on subsequent student gains, and negative school effects have negative effects on subsequent gains.

The model above can be extended to allow for multivariate test outcomes, background covariates, and a term that links more than one school effect to specific students in the event that students attend more than one school in a given year (c.f., Lockwood et al., 2007, pp. 127–128). We have chosen this simpler specification here in order to focus attention on the relationship between the persistence parameters and school effects and to more easily evaluate whether the relationship differs by test subject.

Model Specification When Schools Are Units of Analysis: The Identification Problem

Specifying a variable persistence model when schools are the units of analysis is not a straightforward task. To illustrate this, we start by imagining that links between students and teachers were available for our cohort of students taking the reading test from Grades 4 to 8 over the years 2004 through 2008. If the variable persistence model were to be specified and written out as a system of multiple (correlated) equations, it would take the form

$$
\begin{aligned}
Y_{i04} &= \mu_{04} + \theta_{04} + \varepsilon_{i04} \\
Y_{i05} &= \mu_{05} + \alpha_{21}\theta_{04} + \theta_{05} + \varepsilon_{i05} \\
Y_{i06} &= \mu_{06} + \alpha_{31}\theta_{04} + \alpha_{32}\theta_{05} + \theta_{06} + \varepsilon_{i06} \\
Y_{i07} &= \mu_{07} + \alpha_{41}\theta_{04} + \alpha_{42}\theta_{05} + \alpha_{43}\theta_{06} + \theta_{07} + \varepsilon_{i07} \\
Y_{i08} &= \mu_{08} + \alpha_{51}\theta_{04} + \alpha_{52}\theta_{05} + \alpha_{53}\theta_{06} + \alpha_{54}\theta_{07} + \theta_{08} + \varepsilon_{i08}.
\end{aligned} \tag{2}
$$

In the model above there are a total of 10 unique persistence parameters (the α's). This makes the model quite flexible in its ability to represent change in the persistence of teacher effects over time. For example, if on average the influence of Grade 4 teachers on subsequent student performance becomes weaker and weaker over time, we would expect to see that $\alpha_{51} < \alpha_{41} < \alpha_{31} < \alpha_{21} < 1$.

One complication in the model above is that most of the equations involve the product of two unknown parameters. Hence, to identify each parameter we must be convinced that an estimate of persistence (e.g., α_{32}) can be separated from estimates of teacher effects (e.g., θ_{05}). It can be shown that teacher effects are identified by classroom-level means and mean deviations across years. What identifies the persistence parameters? This depends upon whether, and the extent to which, students and teachers mix from grade to grade. If they do mix, then a grade-specific persistence parameter can be readily identified. To see this, imagine that we have estimated the average value added in mathematics by Ms. Shepard to her Grade 5 classroom. In Grade 6, half of these students move on to a class taught by Mr. Fisher and the other half take a class with Mr. Gasol. In this case, the relative difference in the mean Grade 5 to Grade 6 math score gains for these two groups of students in the classes taught by Mr. Fisher and Mr. Gasol, respectively, becomes a sufficient statistic for the persistence of Ms. Shepard's effect on Grade 6 achievement.

The identification of persistence is plausible when teachers are the units of analysis for a value-added model because (as in the hypothetical scenario above) there is structural mixing between students and teachers from grade to grade. At the other extreme, if all of Ms. Shepard's students moved together as a cohort to learn math in Mr. Gasol's class, the persistence of Shepard's effect would be unidentifiable. Unfortunately, this is the most likely scenario when schools are the units of analysis for a value-added model. Within any given school, students move from grade to grade as a cohort. While technically it would be possible to identify school-level persistence on the basis of students that transfer between schools, this would be a very weak and suspect source of identification. Students who do switch schools within the state are unlikely to be representative of those who do not in terms of their demographic characteristics or academic achievement, factors that are typically associated with the likelihood of a student switching schools. In contrast to the situation where teachers are the units of analysis, the only juncture at which we can expect to see structural mixing of students and schools in the present data context is in the transition from elementary school (i.e., Grade 5) to middle school (i.e., Grade 6).

This has important implications for the general form of the variable persistence model represented by the equations (two) above when schools are the units of analysis. Namely, instead of estimating 10 unique persistence parameters, there is only enough information in our longitudinal data structure to plausibly estimate 1. This means that strong constraints will need to be imposed on the α's if one wishes to test the sensitivity of the assumption of complete persistence made implicitly in the layered model. How these constraints should be imposed will depend upon the extent to which the persistence of school effects estimated from the structural transition from Grade 5 to 6 can be generalized to other grade-specific equations that precede or follow this transition.

The strongest generalization would be to set all α's to be equal to a single constant. This would lead to the following "constrained persistence" (CP) model:[5]

$$
\begin{aligned}
Y_{i04} &= \mu_{04} + \theta_{04} + \varepsilon_{i04} \\
Y_{i05} &= \mu_{05} + \alpha\theta_{04} + \theta_{05} + \varepsilon_{i05} \\
Y_{i06} &= \mu_{06} + \alpha\theta_{04} + \alpha\theta_{05} + \theta_{06} + \varepsilon_{i06} \\
Y_{i07} &= \mu_{07} + \alpha\theta_{04} + \alpha\theta_{05} + \alpha\theta_{06} + \theta_{07} + \varepsilon_{i07} \\
Y_{i08} &= \mu_{08} + \alpha\theta_{04} + \alpha\theta_{05} + \alpha\theta_{06} + \alpha\theta_{07} + \theta_{08} + \varepsilon_{i08}.
\end{aligned}
\tag{CP1}
$$

In this model although the estimate for persistence is being driven by the structural mixing that occurs as of Grade 6, this estimate is being interpolated to inform the Grade 5 equation and extrapolated to inform the Grades 7 and 8 equations. One aspect of this generalization that is probably most intuitively unpalatable is the constraint that the school effects contained in the vector θ_{04} persist at the same rate as those in $\theta_{05}, \theta_{06},$ and θ_{07}. The former capture information about base year school differences in levels of achievement, while the latter are intended to capture information about the subsequent value added to student achievement by the school. Differences among the quantities in θ_{04} can be plausibly explained by variables that are correlated with levels of student achievement (e.g., family education and income, school and district resources, etc.), factors that are, in theory, controlled when estimating school value added for subsequent grades (so long as they do not vary over time). One might hypothesize that the subsequent influence of θ_{04} in future years decays much less rapidly than $\theta_{05}, \theta_{06},$ and θ_{07} if it decays at all.[6]

To better capture this hypothesis, one could specify an alternate version of the constrained persistence model in which base year school differences are assumed to persist undiminished over time while school-level value-added decays by a constant amount.

$$
\begin{aligned}
Y_{i04} &= \mu_{04} + \theta_{04} + \varepsilon_{i04} \\
Y_{i05} &= \mu_{05} + \theta_{04} + \theta_{05} + \varepsilon_{i05} \\
Y_{i06} &= \mu_{06} + \theta_{04} + \alpha\theta_{05} + \theta_{06} + \varepsilon_{i06} \\
Y_{i07} &= \mu_{07} + \theta_{04} + \alpha\theta_{05} + \alpha\theta_{06} + \theta_{07} + \varepsilon_{i07} \\
Y_{i08} &= \mu_{08} + \theta_{04} + \alpha\theta_{05} + \alpha\theta_{06} + \alpha\theta_{07} + \theta_{08} + \varepsilon_{i08}.
\end{aligned} \tag{CP2}
$$

If in fact base year school effects persist at a different rate than value-added effects, one would expect to find significant differences in the estimated persistence parameter from the CP1 to CP2 model specification. In addition, the two models above can be contrasted to a reference model that assumes full persistence of all school effects by constraining all of the α's above to 1.

$$
\begin{aligned}
Y_{i04} &= \mu_{04} + \theta_{04} + \varepsilon_{i04} \\
Y_{i05} &= \mu_{05} + \theta_{04} + \theta_{05} + \varepsilon_{i05} \\
Y_{i06} &= \mu_{06} + \theta_{04} + \theta_{05} + \theta_{06} + \varepsilon_{i06} \\
Y_{i07} &= \mu_{07} + \theta_{04} + \theta_{05} + \theta_{06} + \theta_{07} + \varepsilon_{i07} \\
Y_{i08} &= \mu_{08} + \theta_{04} + \theta_{05} + \theta_{06} + \theta_{07} + \theta_{08} + \varepsilon_{i08}.
\end{aligned} \tag{LM}
$$

This is the layered model (LM; Sanders et al., 1997) applied to schools instead of teachers, and historically this is the value-added model that has and is being used by many American states and school districts to evaluate teacher performance.

To recap, our principal objective in this study was to test the assumption made implicitly in the LM that school effects fully persist as they cumulate over time. However, we found that due to identification obstacles inherent when schools are the unit of analysis instead of teachers, it is not possible to test this assumption by specifying a saturated model in parallel to the approach taken by Lockwood et al. (2007). Instead, we use the structural mixing of students and schools between Grades 5 and 6 as the basis for estimating a single persistence parameter and then generalizing this parameter to other grade equations (i.e., Grades 5, 7, and 8). This resulted in the specification of two candidate "constrained" persistence models. Of interest to us in what follows is the extent to which the estimates for the persistence parameter in models CP1 and CP2 differ from one another and differ from the value of 1 implied by the LM. We then ask whether either specification of constrained persistence would lead to substantively different inferences about school effectiveness.[7]

Parameter Estimation

The parameters of the variable persistence model have been estimated in previous studies using maximum likelihood-based methods as described in McCaffrey et al. (2004) and using Bayesian methods with MCMC estimation as described in Lockwood et al. (2007). In our analysis, we take a Bayesian approach using MCMC estimation with the package "R2WinBUGS" in the R statistical environment to invoke the software WinBUGS (Spiegelhalter, Thomas, & Best, 1999).[8] Our approach to the specification of prior distributions generally mirrors that of Lockwood et al.: Non-informative prior distributions were specified for all model parameters, and initial values were generated randomly. In each model, students with missing test score values in any given year were assumed to be missing at random and linked to a "pseudo-school" for that grade, an approach consistent with the "M2" procedure described by Lockwood et al. in the context of estimating teacher effects. All models were estimated on the basis of a sample burn-in of 2,500 followed by 5,000 iterations. This was done using three different MCMC chains, each generated using different starting values. These chains were then thinned by a factor of 5 before evaluating convergence and reporting summary statistics from the resulting posterior distributions of interest. Convergence was assessed first by visual examination of the chain history and then by computing the Gelman-Rubin convergence statistic $\hat{R}$ (Gelman, Carlin, Stern, & Rubin, 2004). For each model, we found evidence to suggest that (a) our MCMC chains were stationary following our burn-in period and (b) our three chains converged to the same region of the posterior distribution for each parameter.

Results

Comparing Variance Component and Persistence Parameter Estimates Across Models

Table 3 presents summary statistics from each of the three value-added models described above (CP1, CP2, and LM) by test subject. The last row of the table shows the estimated posterior mean and *SD* of the persistence parameter, α. For the baseline LM, this value is not an estimate but is fixed at a value of 1. For the CP1 and CP2 models, $\hat{\alpha}$ is .64 and .10 for reading and .51 and .48 for math. On the whole, these findings support the conclusions by Lockwood et al. that the LM assumption of full persistence ($\alpha = 1$) is not supported by the data, whether teacher or schools are the units of analysis. In reading, the difference in $\hat{\alpha}$ values from CP1 to CP2 suggests that base year school-level differences persist at a different rate than value-added effects. The small value of .10 under CP2 indicates that a very small proportion of a student's academic achievement in subsequent grades is attributable to the influence of school effects in previous grades. In contrast, for math, the similarity in $\hat{\alpha}$ values from CP1 to CP2 suggests that the effects of base year school-level differences persist at about the same rate as value-added effects. One possible explanation for this interaction when going from reading as the test outcome to math is that student performance in the latter is more malleable than the former. That is, reading achievement might be more heavily influenced than math by what parents do

TABLE 3
Parameter Estimates for Variance Components and Persistence by Test Subject and Model Specification

	Reading			Math		
	LM	CP1	CP2	LM	CP1	CP2
Student-level *SD*						
Grade 4	0.96	0.96	0.96	0.90	0.89	0.90
Grade 5	1.05	1.04	1.04	0.91	0.90	0.90
Grade 6	1.00	1.00	1.00	0.95	0.95	0.95
Grade 7	1.05	1.04	1.04	0.90	0.90	0.90
Grade 8	0.93	0.93	0.93	—	—	—
School-level *SD*						
Grade 4	0.71	0.73	0.72	0.69	0.70	0.69
Grade 5	0.33	0.48	0.37	0.41	0.49	0.44
Grade 6	0.30	0.35	0.32	0.40	0.43	0.41
Grade 7	0.28	0.31	0.34	0.33	0.37	0.36
Grade 8	0.31	0.31	0.33	—	—	—
Persistence of value-added		0.64	0.10		0.51	0.48
		(.02)	(.03)		(.01)	(.02)

Note: Parameters estimated by the model were student and school-level variance terms. The table above expresses these in *SD* units to facilitate comparisons with Table 2.

with children outside of schools, an unobserved factor picked up by baseline differences in school means.

In addition to persistence parameter estimates, the posterior means of student and school-level variance component estimates are provided for the main diagonal of $\boldsymbol{\Sigma}$ and the main diagonal of τ. To make these values more interpretable relative to the scale of the summary statistics provided in Table 2, we show the square root of the estimated variance component. (The associated posterior *SD*s are not included in the table to conserve space since they all are very small, ranging between .01 and .02.) The most noticeable difference in these estimates across models can be seen in the school-level variability in Grade 5. Under the LM, these amount to .33 for reading and .41 for math. In contrast, under the CP1 and CP2 models these estimates are .48 and .37, respectively, for reading, and .49 and .44, respectively, for math. Note that the estimated school-level variance components for models with constrained persistence parameters are never smaller than those from the layered model. Finally, though they are not included in Table 3, we find strong intercorrelations between the grade-specific equations of each model at the student level (i.e., the off-diagonals of $\boldsymbol{\Sigma}$). The magnitudes range from a low of .77 (between Grades 4 and 8 for reading outcomes) and a high of .89 (between Grades 6 and 7 for math outcomes).

TABLE 4
Correlations Between School Effects Across Model Specification

	Grade	Schools (*N*)	LM, CP1	LM, CP2	CP1, CP2
Reading	5	547	0.68	0.58	0.78
	6	225	0.90	0.79	0.92
	7	230	0.93	0.74	0.92
	8	231	0.98	0.76	0.87
Math	5	555	0.47	0.91	0.65
	6	240	0.87	0.93	0.91
	7	245	0.86	0.86	0.96

Note: LM = layered model; CP1, CP2 = specifications of constrained persistence models that do not and do assume full persistence of base year school effects, respectively.

Comparisons of School-Level Value-Added Across Models

We now examine whether a violation of the assumption of full persistence is practically significant. The purpose of a value-added model is to draw inferences about teacher or school effects on student achievement. From this standpoint, the key parameter estimates of interest are summary statistics from the posterior distribution of the value-added terms $\{\hat{\theta}_5, \hat{\theta}_6, \hat{\theta}_7, \hat{\theta}_8\}$. Table 4 provides the correlations between the posterior means of school-level value-added across models by grade for each test subject.

We find that estimated school effects across models are moderately to strongly correlated irrespective of the specific test subject or pair of models considered. However, there is considerable variability in these correlations, variability that is larger than that found by Lockwood et al. in the context of their teacher effect estimates for Grades 2 through 5. In the latter case, the four correlations of teacher effects across joint models for reading and math that did and did not assume complete persistence were .82, .81, .77, and .84. In the present context, we find instances where the correlations between models that do and do not assume complete persistence are both much stronger (up to .98 between the effects under LM vs. CP1 for Grade 8 reading) and much weaker (down to .47 between the effects under LM vs. CP1 for Grade 5 math).

At first glance, the pattern of correlations across models on display in Table 4 might appear confusing or even counterintuitive. Our explanation for this is that the correlation of any subject-specific grade effect across models is driven by two different factors. The first factor is the similarity between the grade-specific equations in each model. Ceteris paribus, the closer the match between the grade-level equations, the stronger the correlation between school-level effect estimates. For example, consider the Grade 5 equations found in the LM and the CP1 and CP2 models. The equations in LM and CP2 are identical, while the equations in LM and CP1 differ as a function of α. Hence one might expect a

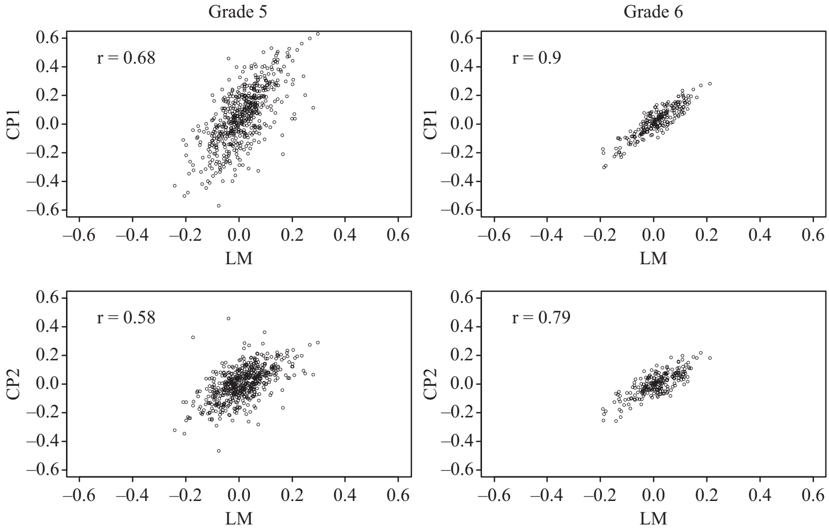

FIGURE 1. *Scatterplots of estimated school effects across models with and without assumption of full persistence: reading tests.*

stronger correlation between Grade 5 effects for the LM, CP2 model pairings relative to the LM, CP1 pairings.

The problem with this interpretation is that it ignores the multivariate structure of each model. Because the equations are strongly intercorrelated (generally .8 or higher), changes to the parameterization of persistence in any single equation can have an impact on estimates of school effects in subsequent or even prior grades. This implies a second factor that will affect the correlation between grade-level school effects across models: The magnitude of the difference in implicit or parameterized values of persistence in any of the equations that specify the full model. Understanding this helps to explain a seemingly contradictory pattern in Table 4: For reading outcomes, the correlation between school effects between the LM and the CP2 model is always smaller for each grade relative to the correlations between the LM and CP1 model; for math outcomes we generally see the opposite. This is because the estimated persistence parameter drops precipitously from CP1 to CP2 in reading (from .64 to .10) while it stays roughly the same from CP1 to CP2 in math ($\sim$.50). When compared to a model in which the implicit value of persistence is always 1 as in the LM, the impact of a drop in $\hat{\alpha}$ from .64 to .10 in reading has much bigger impact on school effect estimates then a drop from .51 to .48, even when the grade-specific equation (e.g., Grade 5) is identical in the CP2 model.

Figures 1 and 2 present these comparisons visually by test subject using scatterplots of the value-added estimates for Grades 5 and 6 (the last year of elementary school and the first year of middle school, respectively, for our restricted

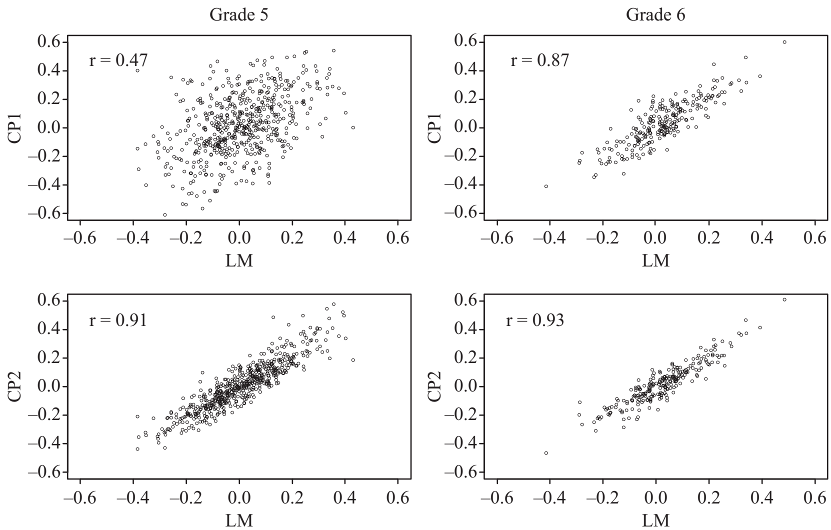

FIGURE 2. *Scatterplots of estimated school effects across models with and without assumption of full persistence: math tests.*

sample) for models that do and do not assume full persistence. From these plots, it is also readily apparent that there is more variability in the distributions of school effects as a function of math outcomes relative to reading outcomes.

Correlations With School-Level Indicator of Poverty

The desirability of value-added measures hinges in large part upon the extent to which they "level the playing field" such that schools are being evaluated on the basis of what students have learned and not on the basis of socioeconomic factors that are outside of a school's control. To this end, it is of interest to examine the extent to which school effects estimated under the assumption of full persistence exhibit a different correlation with an indicator of school-level poverty than that found when full persistence is not assumed. The second column of Table 5 presents the correlations of Grade 6 school effects in reading and math with the school-level percentage of students eligible to receive free or reduced lunch (%FRL) services. The results in Table 5 show that the decision to parameterize persistence appears to have a substantial impact on a key "quality control" indicator of value-added estimators. In most cases, the inclusion of an estimated persistence parameter < 1 for the CP1 and CP2 models leads to a higher correlation between school value-added and %FRL.

The explanation for this pattern of findings is similar in nature to the one provided for the correlations shown in Table 4. Consider the Grade 6 equation

TABLE 5

Correlations Between Grade 6 School Effects and %FRL by Model and Test Subject

Reading	$r(\hat{\boldsymbol{\theta}}_{06},\%\text{FRL})$
LM : $\hat{Y}_{i06} = \hat{\mu}_{06} + \hat{\boldsymbol{\theta}}_{04} + \hat{\boldsymbol{\theta}}_{05} + \hat{\boldsymbol{\theta}}_{06}$	−.26
CP1 : $\hat{Y}_{i06} = \hat{\mu}_{06} + .64\hat{\boldsymbol{\theta}}_{04} + .64\hat{\boldsymbol{\theta}}_{05} + \hat{\boldsymbol{\theta}}_{06}$	−.52
CP2 : $\hat{Y}_{i06} = \hat{\mu}_{06} + \hat{\boldsymbol{\theta}}_{04} + .10\hat{\boldsymbol{\theta}}_{05} + \hat{\boldsymbol{\theta}}_{06}$	−.43
Math	$r(\hat{\boldsymbol{\theta}}_{06},\%\text{FRL})$
LM : $\hat{Y}_{i06} = \hat{\mu}_{06} + \hat{\boldsymbol{\theta}}_{04} + \hat{\boldsymbol{\theta}}_{05} + \hat{\boldsymbol{\theta}}_{06}$	−.20
CP1 : $\hat{Y}_{i06} = \hat{\mu}_{06} + .51\hat{\boldsymbol{\theta}}_{04} + .51\hat{\boldsymbol{\theta}}_{05} + \hat{\boldsymbol{\theta}}_{06}$	−.45
CP2 : $\hat{Y}_{i06} = \hat{\mu}_{06} + \hat{\boldsymbol{\theta}}_{04} + .48\hat{\boldsymbol{\theta}}_{05} + \hat{\boldsymbol{\theta}}_{06}$	−.19

Note: LM = layered model; CP1, CP2 = specifications of constrained persistence models that do not and do assume full persistence of base year school effects, respectively.

for reading outcomes extracted from the LM, CP1, and CP2 models, respectively. These three equations are written out in the first column of Table 5, and for each equation the estimated values of the persistence parameters have been inserted. Conceptually, for each model the objective is to attribute variability in the reading test scores of Grade 6 students to the schools these students were attending in Grades 4, 5, and 6. Now examine just the Grade 6 equation as we move from LM to CP1. By specifying a persistence parameter in CP1, we necessarily reduce the amount of variability being attributed to schools in Grades 4 and 5 of and increase the amount being attributed to them in grade 6 (i.e., only 64% $\hat{\theta}_{04}$ and $\hat{\theta}_{05}$ are retained relative to LM). So as a general rule, the correlation of a current year/grade value-added estimate and a variable such as %FRL should always increase in absolute magnitude when all prior school effects do not fully persist—that is, when $\hat{\alpha} < 1$.

However, in the CP2 model, we have a mixed situation: the base year school effect ($\hat{\theta}_{04}$) has an implied persistence parameter of 1, but the Grade 5 value-added school effect has an estimated persistence parameter that is 0.1 (contrast this to the .64 estimate from the CP1 model). Relative to the CP1 model, the variability of Grade 6 reading achievement attributed to $\hat{\theta}_{04}$ has increased, but the variability attributed to $\hat{\theta}_{05}$ has decreased. These two changes tend to cancel one another, with the result that the variability attributed to $\hat{\theta}_{06}$ increases relative to LM, but decreases slightly relative to CP1. All this is reflected in the correlational patterns between the three different sets of Grade 6 reading value-added estimates with %FRL.

In contrast, for math outcomes note that the correlation of value added with %FRL increases in absolute magnitude much more dramatically from CP1 (−.45) to CP2 (−.19). This is because the common persistence parameter estimate associated with $\hat{\theta}_{05}$ in both CP1 and CP2 is about the same ($\sim$.5), while the persistence of the base year school effect goes from .51 in CP1 to 1 in CP2.

Comparison of Schools Classified as Effective or Ineffective Across Models

If value-added models were to be used as a basis for school accountability decisions, it might be likely that a classification rule would be established on the basis of the perceived precision of estimated school effects.[9] To evaluate the potential policy impact of specifying a model in which complete persistence is not assumed relative to one in which it is (i.e., CP1 or CP2 instead of LM), we place schools into three categories of "effectiveness" by grade: above average (+), average (0), or below average (−). A school is classified as above or below average in effectiveness when there is a 95% probability that it has a value-added effect greater or less than the mean effect over all schools in the sample. More specifically, at a given grade for each school, we create a credibility interval around that school's posterior mean by adding and subtracting two posterior *SD*s. Next, we create cross-tabulations of these classification by grade and model. If the two models agree in their classifications of schools, we would expect to see values falling along the main diagonal of the crosstab. To the extent that they disagree, we will see schools along the off-diagonal. Tables 6 and 7 present these crosstabs by grade for reading and math outcomes, respectively. The numbers in each cell represent the percentage of schools for which value-added effects were estimated. For example, in Grade 5 reading, 10% of the schools ($.10 \times 547 \approx 55$) that were classified as average in effectiveness under the LM would be classified as below average under either the CP1 or CP2 models. Note that while there are many schools for which classifications would shift from average to above or below average (or vice versa), there are almost no cases where a school would shift by two categories (i.e., from below average to above average).

The cumulative percentages of schools in the off-diagonals of the crosstabs in Tables 6 and 7 are summarized in Table 8. Inspection of these results indicates that there are a substantial number of schools for which classifications would change as a function of model specification. For reading and math, this ranges from highs of 33% and 38% to lows of 10% and 17%, respectively. In general, more schools are classified as above or below average when the CP1 or CP2 models are specified relative to the LM (this is denoted by the columns labeled "0 to (+/−)" in Table 8). For reading, the shift in classifications tends to be larger for the CP2 model (with the exception of Grade 5); for math, the shift is largest for the CP1 model.

Discussion

Summary of Findings

On the whole, our results support the conclusion reached by McCaffrey et al. (2004) and Lockwood et al. (2007) that the assumption of full persistence in a

TABLE 6
Comparisons of School Classifications by Value-Added Model for Reading

			CP1					CP2		
Grade			−	0	+			−	0	+
5	LM	−	3%	2%	0%	LM	−	3%	1%	0%
		0	10%	55%	20%		0	10%	69%	7%
		+	0%	1%	8%		+	0%	5%	5%
6	LM	−	8%	1%	0%	LM	−	8%	2%	0%
		0	6%	57%	11%		0	6%	62%	6%
		+	0%	2%	15%		+	0%	7%	10%
7	LM	−	6%	0%	0%	LM	−	5%	1%	0%
		0	6%	70%	8%		0	10%	64%	10%
		+	0%	1%	9%		+	0%	4%	6%
8	LM	−	10%	3%	0%	LM	−	7%	4%	0%
		0	3%	69%	3%		0	6%	61%	8%
		+	0%	1%	12%		+	0%	6%	6%

Note: LM = layered model; CP1, CP2 = specifications of constrained persistence models that do not and do assume full persistence of base year school effects, respectively. Percentages are expressed in terms of total number of schools in each grade (see Table 5). School classifications are based upon estimated posterior *M*s and *SD*s of school effects. The category "+" represents a school with an estimated value-added effect that remains above 0 after two posterior *SD*s have been subtracted from its posterior mean. The category "0" represents a school with an estimated value-added effect that crosses 0 after two posterior *SD*s have been subtracted from or added to its posterior mean. The category "−" represents a school with an estimated value-added effect that remains below 0 after two posterior *SD*s have been added to its posterior mean.

multivariate random effects model is not very tenable. The values we found when estimating a persistence parameter for school effects were .64 and .10 for reading outcomes and .51 and .48 for math outcomes. It appears that to the extent that there is a decay in school effects from grade to grade, this decay differs by test subject. In our findings we see some empirical evidence that student achievement in math is more malleable than achievement in reading.

With regard to the practical impact of specifying models with and without full persistence, we found that while our correlation of school effects across models tended to be strong (with the notable exception of the Grade 5 effects), in most cases the specification of models with a persistence parameter resulted in a significant increase in the correlations between school effects and measures of school-level socioeconomic status. The value-added models we specified without the assumption of full persistence classified a larger proportion of schools as significantly above or below average in their effectiveness relative to the layered model. The mean increase in the proportions of schools that switch categories from 0 to + or − in our data were 8 and 10 percentage points in reading

TABLE 7
Comparisons of School Classifications by Value-Added Model for Math

Grade			CP1 −	CP1 0	CP1 +			CP2 −	CP2 0	CP2 +
5	LM	−	7%	6%	3%	LM	−	14%	2%	0%
		0	10%	39%	13%		0	5%	52%	4%
		+	1%	9%	12%		+	0%	5%	16%
6	LM	−	11%	3%	0%	LM	−	12%	2%	0%
		0	7%	42%	11%		0	6%	48%	5%
		+	0%	3%	24%		+	0%	4%	23%
7	LM	−	11%	4%	0%	LM	−	13%	3%	0%
		0	10%	45%	10%		0	9%	48%	8%
		+	0%	5%	13%		+	0%	6%	13%

Note: LM = layered model; CP1, CP2 = specifications of constrained persistence models that do not and do assume full persistence of base year school effects, respectively. Percentages are expressed in terms of total number of schools in each grade (see Table 5). School classifications are based upon estimated posterior *M*s and *SD*s of school effects. The category "+" represents a school with an estimated value-added effect that remains above 0 after two posterior *SD*s have been subtracted from its posterior mean. The category "0" represents a school with an estimated value-added effect that crosses 0 after two posterior *SD*s have been subtracted from or added to its posterior mean. The category "−" represents a school with an estimated value-added effect that remains below 0 after two posterior *SD*s have been added to its posterior mean.

TABLE 8
Proportion of Schools that Switch Classifications When Persistence Parameter is Estimated

		Schools (*N*)	LM to CP1: (+/−) to 0	LM to CP1: 0 to (+/−)	LM to CP2: (+/−) to 0	LM to CP2: 0 to (+/−)
Reading	Grade 5	547	3%	30%	6%	16%
	Grade 6	225	3%	16%	9%	12%
	Grade 7	230	1%	13%	5%	20%
	Grade 8	231	4%	6%	10%	14%
Math	Grade 5	555	15%	23%	8%	10%
	Grade 6	240	6%	18%	6%	11%
	Grade 7	245	10%	20%	9%	17%

and math, respectively. These increases, while significant, are considerably smaller than the corresponding increases observed in the proportions of teacher that switch in these categories in the RAND study (a mean increase across grades of 24 and 21 percentage points in reading and math). This may indicate that the impact of relaxing the assumption of full persistence is stronger in the context of estimating teacher effects than it is in the context of estimating school effects, perhaps because there is greater flexibility to specify multiple persistence parameters in the former context.

One possible explanation for the differences between our findings and those reported by Lockwood et al. is that the parameter estimates reported in the RAND study were based on the joint modeling of reading and math scores, while we have focused attention on parameter estimates from marginal models. Lockwood et al. noted that the joint modeling of test subjects tends to reduce estimates of persistence parameters and the variability of value-added effects relative to marginal modeling. Another possibility is that we are using different sources of data that encompass different grade spans (Grades 4–8 in the present study and Grades 1–5 in Lockwood et al.). Furthermore, our data include student test scores across an entire state, while the RAND study used test scores across a single large urban school district.

Differences Between Teacher and School "Effects"

A point of emphasis in this article has been that the identification and estimation of persistence parameters is considerably more difficult when schools are the units of analysis relative to teachers. We have argued that in most empirical contexts, it will only be reasonable to specify a single persistence parameter, the identification of which must be based upon the structural mixing of students when they move from elementary school to middle school. That is, even though weak forms of identification are technically possible on the basis of small numbers of students who transfer between schools from grade to grade, or even through the specification of prior distribution in using a Bayesian estimation approach, we would argue that the theoretical defense for such identification would be suspect. Because of this, our recommendation is that the best way to test the sensitivity of the assumption of full persistence made in value-added models such as the LM is to impose constraints on Lockwood et al.'s variable persistence model by setting persistence parameters to a single unknown constant. Whether this constant can be generalized to both base year school effects and value-added school effects is an open question.

There are also important conceptual differences between the specification of value-added models for schools instead of teachers. When only school effects are included without teacher effects, the school effects are likely to represent an aggregation of teacher effects of student achievement, but they are also likely to capture the influence of administrative leadership and policies that might fall under the heading of school "climate." It seems reasonable to assume that the effect of a school on student achievement should be larger than the effect of a teacher, but it is unclear whether we would also expect one to persist at a different rate than the other. When school effects are omitted from a value-added model (whether or not persistence has been parameterized), it seems likely that any estimated teacher effects will be biased to the extent that better teachers systematically attend better schools.[10] The inclusion of school effects might serve to reduce this bias, but it also changes the normative frame of reference for teacher effects to be within school. What is even less clear is the extent to which estimated school effects are biased when teacher effects have been omitted. We know of no examples in which a single value-added model has been implemented with the intent of drawing inferences about the effectiveness of *both* teachers and schools.[11]

Unfortunately, because students and schools do not mix randomly, it seems unlikely that any of the model specifications discussed in this article produce value-added estimates that can be plausibly interpreted as unbiased estimates of the causal effect of schools on student achievement. This issue has been well documented by Raudenbush (2004) and Rubin, Stuart, and Zanutto (2004). In this sense, relaxing the assumption of full persistence adds another wrinkle to this problem since there is similarly no guarantee that α can be estimated without bias. Indeed, the same problems of self-selection that would lead to bias in value-added estimates will lead to bias in $\hat{\alpha}$.

Some Conclusions

When a multivariate random effects model is used to estimate school-level value-added, there is no easy answer to the question of how to parameterize persistence. Our findings in this study indicate that the assumption of full persistence for school effects is hard to justify. However, the assumption that school effects decay at the same constant rate is also intuitively implausible. The dilemma for the pragmatic researcher in this context is whether to choose between two (seemingly) suboptimal model parameterizations given observational data constraints or to reject the premise that value-added models should be used to draw causal inferences about school quality. Depending upon the stakes attached to any subsequent classification of schools, the choice could have significant policy ramifications. When full persistence is assumed (LM vs. CP1 or CP2), grade-specific school effects will tend to be less correlated with indicators of school-level poverty, but a smaller proportion of schools will be classified as "effective" or "ineffective" for any given test subject.

If the use of a constrained persistence model is being contemplated, we think it is important to first specify and estimate more than one version as we have done here to test the sensitivity of the persistence parameter estimate and thereby the degree to which the estimate can be generalized. In particular, it is important to consider the plausibility that base year school effects decay at the same rate as value-added school effects. We found that while this might be plausible for math outcomes, it does not seem plausible for reading outcomes.

Finally, we note that there is something to be gained by specifying and estimating the sorts of constrained persistence models illustrated in this article because they can provide some insights into important differences in the ways that schools appear to influence learning by test subject. If it can be generalized that base year differences in math achievement decay at a much faster rate than base year differences in reading, then nationally we should expect to see the achievement gap in math narrowing much faster than the achievement gap in reading. In fact, there seems to be evidence to support this hypothesis in recent examination of trends in NAEP data since the implementation of No Child Left Behind (Wong, Cook, & Steiner, 2009). This is an instance when the specification and estimation of a value-added model would be undertaken not to draw inference about schools (or teachers) but to make and test hypotheses about student learning.

Notes

1. It can be argued that it is a mistake to use the term "effect" to characterize estimates of teacher or school value added because it implies a causal inference that seems at best very equivocal. However, we decided to use the "effect" terminology because (a) it is consistent with the extant literature on value-added modeling and (b) there is little question that the intent behind the application of these models in high-stakes settings is to draw inferences about teacher or school quality, whether the estimates are unbiased or not.
2. Economists tend to use the terms "decay" or "fade-out" rather than persistence. We use the terms "persistence" and "decay" somewhat interchangeably, though obviously the terms are inversely related.
3. We say "roughly" because from year to year, a small number of new schools were added either because they were newly formed or because their data had not been previously available. For example, the total number of middle schools in the reading cohort increased from 225 to 230 to 231 from 2006 to 2008. Likewise for the math cohort, the total number of middle schools increased from 240 to 245 from 2007 to 2008.
4. A full explanation of the approach used to create the vertical scale is outside the scope of this study. For a general background on vertical scaling, see Kolen and Brennan (2004). For details on vertical scaling and its relationship to value-added modeling, see Briggs and Weeks, 2009. Also see Martineau and Reckase (2006).
5. When applied to reading outcomes the model consists of all five of the equations above; when applied to math outcomes only the first four equations would apply.

6. Interestingly (and surprisingly), in the context of estimating unique persistence parameters with teachers as the units of analysis, the empirical results found by Lockwood et al. (2007) showed no such pattern. The decay of base year teacher effects was just as strong (i.e., small value of α) as that of value-added teacher effects.

7. We should note in passing that a number of other constrained specifications of the variable persistence model would be both possible and defensible. We do not claim that these two models are inherently valid, but they are also not implausible a priori.

8. The code used for this analysis is available upon request.

9. We remain agnostic as to whether taking such an approach is actually a wise idea. When applied to the full population of schools in a state, the chance process at work is little more than a thought experiment. We may wish to capture the hypothetical uncertainty associated with a school effect had a different cohort of students been available but actually doing so requires what Berk (2004) refers to as "model-based" inferences. This touches upon a philosophical argument about the appropriate uses of statistical models (c.f., Breiman, 2001) that is outside the scope of the present article, but it may well be a debate worth having in the context of value-added modeling applications.

10. By the same token, school effects are probably biased to the extent that better schools are clustered within better school districts.

11. The use of value-added models to make decisions related to educational accountability is likely to create different sets of incentives when the focus is schools rather than teacher. It may well be the case that if schools are the units of analysis, this creates an incentive for teachers to cooperate and work collaboratively, while when teachers are the units of analysis, they may be more likely to see themselves as being in competition with their colleagues. A full discussion of these issues is outside the scope of the present article, but see Harris (2009).

References

Ballou, D., Sanders, W. L., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics*, *29*, 37–66.

Berk, R. (2004). *Regression analysis: A constructive critique.* Thousand Oaks, CA: Sage.

Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science*, *16*, 199–231.

Briggs, D. C., & Weeks, J. P. (2009). The sensitivity of value-added modeling to the creation of a vertical scale. *Education Finance & Policy, 4,* 384–414.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2004). *Bayesian data analysis* (2nd ed.). Boca Raton, FL: CRC Press.

Harris, D. (2009). Would accountability based on teacher value added be smart policy? An examination of the statistical properties and policy alternatives. *Education Finance and Policy*, *4*, 319–350.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling and linking* (2nd ed.). New York, NY: Springer-Verlag.

Lockwood, J. R., McCaffrey, D. F., Mariano, L. T., & Setodji, C. (2007). Bayesian methods for scalable value-added assessment. *Journal of Educational and Behavioral Statistics*, *32*, 125–150.

Martineau, J. A., & Reckase, M. (2006). *Dimensionality and vertical equating*. Paper presented at the Council of Chief State School Officers National Conference on Large-Scale Assessment.

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

Raudenbush, S. W. (2004). *Schooling, statistics, and poverty: Can we measure school improvement?* Paper presented at the William H. Angoff Memorial Lecture Series, Princeton, NJ. Retrieved from January 25, 2005, from http://www.ets.org/Media/Education_Topics/pdf/angoff9.pdf

Rubin, D., Stuart, A., & Zannato, E. (2004). A potential outcome view of value-added assessment in education. *Journal of Educational and Behavioral Statistics*, *29*, 103–116.

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 measure?* (pp. 137–162). Thousand Oaks, CA: Corwin Press.

Spiegelhalter, D. J., Thomas, A., & Best, N. G. (1999). *WinBUGS Version 1.2 User Manual.* MRC Biostatistics Unit.

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement, 7,* 201–210.

Wong, M., Cook, T., & Steiner, P. (2009). *No Child Left Behind: An interim evaluation of its effects on learning using two interrupted time series each with its own non-equivalent comparison series*. IPR Working Papers. Retrieved January 28, 2010, from http://www.northwestern.edu/ipr/publications/workingpapers/wpabstracts09/wp0911.html

Authors

DEREK C. BRIGGS is Associate Professor of Education at the University of Colorado at Boulder <derek.briggs@colorado.edu>. He is Program Chair of the Research and Evaluation Methodology Program. His research agenda focuses upon building sound methodological approaches for the valid measurement and evaluation of growth in student learning.

JONATHAN P. WEEKS is a recent graduate from the Research and Evaluation Methodology doctoral program at the University of Colorado at Boulder <jonathan.weeks@colorado.edu>. His research interests include unidimensional and multidimensional vertical scaling and growth modeling.

Manuscript received July 3, 2009
Revision received August 24, 2010
Accepted November 6, 2010