The Persistence of School-Level Value-Added

Derek C. Briggs
Jonathan P. Weeks

University of Colorado at Boulder

July 2009

Pre-print for Briggs, D. C. & Weeks, J. P. (2011) The persistence of value-added school effects. *Journal of Educational and Behavioral Statistics*, 36(5), 616-637.

Abstract

Using longitudinal data for an entire state from 2004 to 2008, this paper 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 & Setodji (2007). 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 that 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.

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 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. Sanders, Saxton & Horn, 1997, Ballou, Sanders & Wright, 2004). 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, McCaffrey, Mariano & Setodji (2007) provided empirical evidence that the contribution of a teacher two 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 precision of estimated teacher effects appear to be sensitive to the way that persistence is

[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.

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 even 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 can be, to a large extent, conceptualized as an aggregation of the former. Hence, if teacher effects do not fully persist over time, than 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 relative 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 five 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 unique

persistence parameter. Given this modeling constraint, some hard choices need to be made with regard to the parameterization of school effect persistence. In this paper 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.

Data

The data for this study come from two longitudinal cohorts taken from the full population of students and schools in a mid-sized state west of the Mississippi. The first cohort took the state's standardized assessment in reading over five years: in 2004 as 4th graders, and in 2008 as 8th 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 four years: as 4th graders in 2005 and as 7th 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 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[3]. The respective numbers for the math cohort were 27,803

[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

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.

Insert Table 1 about here

The students in our restricted samples were somewhat more likely to be nonwhite, 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 to 6 score gains relative to the full population. The test scores in the subjects of reading and math that 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 3-10 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.**

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.

Item parameters and ability estimates from subsequent tests are horizontally equated so that they can be linked back to the base vertical scale.[4].

Insert Table 2 about here

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

$$Y_{it} = \mu_t + \sum_{t^* \le t} \alpha_{it^*} \theta_{t^*} + \varepsilon_{it} \qquad (1)$$

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 test score 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 $\varepsilon_{it} \sim N(0, \Sigma)$ and $\theta_{t^*} \sim N(0, \tau)$. The two covariance matrices differ in that the former is unstructured

[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 & Brennan (2004). For details on vertical scaling and its relationship to value-added modeling, see Authors, 2009. Also see Martineau & Reckase, 2006.

while the latter is typically specified as a diagonal matrix. The vector $\boldsymbol{\theta}_{t^*}$ represents the collection of school effects for each year (given that $t^* \leq t$). (Since a unique parameter is associated with each year, 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 $\boldsymbol{\theta}_{t^*}$. The two different subscripts, *t* and t^* are used to distinguish the association of persistence parameters with school effects over time. 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 school effects to specific students in the event that students attend more than one school in a given year (c.f., Lockwood et al., 2007, p. 127-128). We have chosen this simpler specification here in order to focus attention on the relationship between the persistence parameters and schools effects and to 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 to 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

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

In the model above there are a total of 10 distinct persistence parameters (the $\alpha's$). This makes the model quite flexible in its ability to represent change in the persistence in 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 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 that do switch schools within the state are unlikely to be representative of those that 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 in (2) 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 $\alpha's$ 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 $\alpha's$ to be equal to a single constant. This would lead to the following "constrained persistence" (CP) model[5]:

$$
\begin{aligned}
Y_{i04} &= \mu_{04} + \boldsymbol{\theta}_{04} + \varepsilon_{i04} \\
Y_{i05} &= \mu_{05} + \alpha\boldsymbol{\theta}_{04} + \boldsymbol{\theta}_{05} + \varepsilon_{i05} \\
Y_{i06} &= \mu_{06} + \alpha\boldsymbol{\theta}_{04} + \alpha\boldsymbol{\theta}_{05} + \boldsymbol{\theta}_{06} + \varepsilon_{i06} \\
Y_{i07} &= \mu_{07} + \alpha\boldsymbol{\theta}_{04} + \alpha\boldsymbol{\theta}_{05} + \alpha\boldsymbol{\theta}_{06} + \boldsymbol{\theta}_{07} + \varepsilon_{i07} \\
Y_{i08} &= \mu_{08} + \alpha\boldsymbol{\theta}_{04} + \alpha\boldsymbol{\theta}_{05} + \alpha\boldsymbol{\theta}_{06} + \alpha\boldsymbol{\theta}_{07} + \boldsymbol{\theta}_{08} + \varepsilon_{i08}.
\end{aligned}
\qquad \text{(CP1)}
$$

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 grade 7 and 8 equations. One aspect of this generalization that is probably most intuitively unpalatable is the

[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.**

constraint that the school effects contained in the vector θ_{04} persist at the same rate as those in θ_{05}, θ_{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 θ_{05}, θ_{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.

$$
\begin{aligned}
Y_{i04} &= \mu_{04} + \boldsymbol{\theta}_{04} + \varepsilon_{i04} \\
Y_{i05} &= \mu_{05} + \boldsymbol{\theta}_{04} + \boldsymbol{\theta}_{05} + \varepsilon_{i05} \\
Y_{i06} &= \mu_{06} + \boldsymbol{\theta}_{04} + \alpha\boldsymbol{\theta}_{05} + \boldsymbol{\theta}_{06} + \varepsilon_{i06} \\
Y_{i07} &= \mu_{07} + \boldsymbol{\theta}_{04} + \alpha\boldsymbol{\theta}_{05} + \alpha\boldsymbol{\theta}_{06} + \boldsymbol{\theta}_{07} + \varepsilon_{i07} \\
Y_{i08} &= \mu_{08} + \boldsymbol{\theta}_{04} + \alpha\boldsymbol{\theta}_{05} + \alpha\boldsymbol{\theta}_{06} + \alpha\boldsymbol{\theta}_{07} + \boldsymbol{\theta}_{08} + \varepsilon_{i08}.
\end{aligned}
\qquad \text{(CP2)}
$$

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 one the CP1 to CP2 model specification. In addition, the two

[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.

models above can be contrasted to a reference model that assumes full persistence of all school effects by constraining all of the $\alpha's$ above to 1.

$$
\begin{aligned}
Y_{i04} &= \mu_{04} + \boldsymbol{\theta}_{04} + \varepsilon_{i04} \\
Y_{i05} &= \mu_{05} + \boldsymbol{\theta}_{04} + \boldsymbol{\theta}_{05} + \varepsilon_{i05} \\
Y_{i06} &= \mu_{06} + \boldsymbol{\theta}_{04} + \boldsymbol{\theta}_{05} + \boldsymbol{\theta}_{06} + \varepsilon_{i06} \\
Y_{i07} &= \mu_{07} + \boldsymbol{\theta}_{04} + \boldsymbol{\theta}_{05} + \boldsymbol{\theta}_{06} + \boldsymbol{\theta}_{07} + \varepsilon_{i07} \\
Y_{i08} &= \mu_{08} + \boldsymbol{\theta}_{04} + \boldsymbol{\theta}_{05} + \boldsymbol{\theta}_{06} + \boldsymbol{\theta}_{07} + \boldsymbol{\theta}_{08} + \varepsilon_{i08}
\end{aligned}
\qquad \text{(LM)}
$$

This is the layered model (LM; Sanders, Saxton & Horn, 1997) applied to schools instead of teachers, and historically this is the value-added model that has and is being used by 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 this results 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[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-

[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.

Rubin convergence statistic $\hat{R}$ (Gelman, Carlin, Stern & Rubin, 2004). For each model we found evidence to suggest that (1) our MCMC chains were stationary following our burn-in period, and (2) 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

Insert Table 3 about here

Table 3 presents summary statistics from each of the three value-added models described above (CP1, CP2, LM) by test subject. The last row of the table shows the estimated posterior mean and SD of the persistence parameter, $\hat{\alpha}$. 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 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 the effects of base year school-level differences persist at a different rate than do 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 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 SDs 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).

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 $\{\hat{\boldsymbol{\theta}}_5, \hat{\boldsymbol{\theta}}_6, \hat{\boldsymbol{\theta}}_7, \hat{\boldsymbol{\theta}}_8\}$. Table 4 provides the correlations between the posterior means of school-level value-added across models by grade for each test subject.

Insert Table 4 about here

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. This is because 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 stronger correlation between grade 4 effects for the former model pairings relative to the latter model pairings. Even more generally, since the CP2 model equations more closely resemble the LM model equations, one would expect the school effects by grade for each model to be more closely correlated than when the comparison to the LM is made with the CP1 model.

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: 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. That is, 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 while it stays roughly the same from CP1 to CP2 in math. 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 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 poverty than that found when full persistence is not assumed. 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. Here we see that the decision to parameterize persistence can have a substantial impact on a key feature of value-added estimates—their correlation with preexisting measures of status. Only for the CP2 model using math outcomes do we find the same low correlation with FRL (about –.2) that was found using the LM.

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 not (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 SDs. Next, we create crosstabulations of these classification by grade and model. If the two models agree in their classifications of schools, we would expect to see

[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.

values falling along the main diagonal of the crosstab. To the extent that they disagree, we will see schools that fall along the off-diagonals of the crosstab. 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 whom value-added effects were estimated. For example, in grade 5 reading, 10% of the schools (.10*547 ≈ 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 whom 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).

Insert Tables 6-8 about here

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 whom 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

Value-added modeling is becoming increasingly popular as a tool used within educational accountability systems. In the context of modeling teacher effects, research by McCaffrey, Lockwood and their colleagues has suggested that decisions about how to parameterize the persistence of effects over time can have a substantial impact upon the classification of teachers as effective or ineffective. This study is the first to examine this issue within the context of modeling value-added at the school level. The identification and estimation of persistence parameters is more complicated in the school context because there are few occasions where there is the kind of structural mixing between schools and students that occurs from grade to grade between teachers and students. Given this, we proposed and estimated two constrained versions of the variable persistence model as a method for testing the tenability of the assumption of full persistence. In each version only a single persistence parameter could be estimated, but the models differed in terms of the way persistence was differentiated for parameters that represent base year school differences and parameters intended to represent a school's value-added contribution to student achievement. Both these models were contrasted to a reference model (the layered model) in which all persistence parameters are fixed to equal 1. This operationalizes the implicit assumption of full persistence in school effects across grades made in the LM.

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 is not very tenable. The values we found when estimating a persistence parameter were .64 and .10 for reading outcomes and .51 and .48 for math outcomes. Interestingly, however, 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. These sorts of subject specific differences in persistence were not discussed in the study by Lockwood et al because the authors specified a joint model for both reading and math test score outcomes.

With regard to the practical impact of specifying models with and without full persistence, our findings represent something of a mixed bag. 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 some cases the specification of models with a persistence parameter can result in a significant increase in the correlations between school effects and measures of school-level socioeconomic status. On the other hand, 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 and math. 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 aforementioned 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 through 8 in the present study, grades 1 through 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 paper 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 though the specification of prior distribution in using our Bayesian estimation approach, we would argue that the theoretical defense for such identification would be suspect. Because of this, our contention 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 either equal a single constant, or to assume full persistence (i.e., fixing the value at 1).

Beyond the issue of identification, which is primarily technical in nature, there are also important conceptual difference 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. What is 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[10].

Limitations and Caveats

Unfortunately, because students and schools do not mix randomly there is no guarantee that any of the model specifications discussed in this paper 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 & 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 $\hat{\alpha}$. To assess the sensitivity of $\hat{\alpha}$ to student mixing, we conducted a "back of the envelope" experiment by attempting to vary the quantity of mixing in an underlying sample of students and schools. We split the sample that constituted our reading cohort into two subsets, "stable" and "mixed". The stable subset (N=2,735) consisted of students who attended middle schools in which 60% or more of grade 6 students came from the same

[10] 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 ms, but see Harris (2009).

elementary school. The mixed subset (N=10,967) consisted of students attending middle schools in which no more than one quarter of grade 6 students came from the same source elementary school. Our hypothesis was that if we specified the same CP1 and CP2 models with each subsample, we would arrive at very different estimates for the associated persistence parameters. This is precisely what we found. When the underlying sample of students tends to stay together in the same cohort from grade 5 to 6, the estimates for $\hat{\alpha}$ were .53 and .41 for the CP1 and CP2 models. When the underlying sample of students is indicative of greater mixing across schools from grade 5 to 6, the corresponding estimates for $\hat{\alpha}$ were .83 and .06. It is, of course, possible that both sets of estimates are in some sense accurate if each of our samples represents a distinct population of students and schools, but clearly, the estimates are not invariant to the choice of sample.

Some Recommendations

Taken together, it might be easy to interpret the studies by McCaffrey et al (2004), Lockwood et al (2007), and the present study as making the case that value-added models which formally parameterize persistence are in some sense "better" than the layered model, which does not. But this is not necessarily a fair assessment, at least based on the empirical results reported here. In general, whether or not one parameterizes persistence, both teacher and school estimates of value-added tend to be strongly correlated. Hence to a large extent, any policy decision about which to prefer might depend upon which of two other criteria is considered more important:

(1) value-added estimates with weak to nonexistent correlation with measures of socioeconomic status (this might be viewed as a proxy indicator for bias), or (2) value-added estimates that make finer grained distinctions between schools that will be classified as more or less effective. If only the first criterion were considered, then the LM would be favored for reading outcomes, and either the LM or CP2 model would be considered acceptable for math outcomes. If only the second criterion were considered, then for both reading and math outcomes either the CP1 or CP2 models would be favored.

In our view, it would be a mistake to make model specifications solely on the basis of increased precision (the second criterion) at the cost of a potential increase in the bias in the estimated value-added effect (the first criterion). If a constrained persistence model is to be used, 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.

There is another reason to be cautious in claiming a preference for models that parameterize persistence. As implemented in the Tennessee Value-Added Assessment System (TVAAS) described by Sanders, Saxton & Horn (1997), the LM uses as inputs five years of panel data. That is, rather than using a single longitudinal cohort as has been done here, the TVAAS would use the results from

ten unique longitudinal cohorts across a five year time span. This would result in an increase in the precision of estimated teacher (or school) effects, and an increase in the numbers of teachers/schools classified as above or below average in their effectiveness, thus potentially removing the principal advantage that seems to be associated with the parameterization of variable persistence. To our knowledge, no one has compared the estimates of teacher or school effects with and without the assumption of full persistence using panel data. (Indeed, the latter may not be feasible given the computational burden involved.) It would be interesting to address the following question: Given the choice of estimating teacher or school effects using a single longitudinal cohort that relaxes the assumption of full persistence (i.e., the variable persistence model) or panel data that imposes the assumption of complete persistence (i.e., the layered model), which should be preferred and why? This would seem to be a fruitful direction for future research[11].

Finally, we note that there is something to be gained by specifying and estimating even the very constrained persistence models illustrated in this paper 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

[11] A big reason this question has been difficult to address empirically to date is that the code used to estimate the parameters of the layered model in the context of panel data remains proprietary. In contrast, the code for the variable persistence model in the context of a single longitudinal cohort is publicly available.

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.

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*(1), 37-66.

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

Breiman, L. (2001). Statistical modeling: the two cultures. *Statistical Science*, 16(3), 199-231.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2004). *Bayesian Data Analysis.* CRC Press, 2nd Edition.

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*, Vol 4(4), 319-350.

Kolen, M. J. and Brennan, R. L. (2004). *Test Equating, Scaling and Linking*. 2nd Edition. New York: 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*. Vol 32(2), 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, and Hamilton, L. (2004) Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, Vol 29:1, 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*(1), 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.

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

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 & 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%
Test Score Mean (SD)	623 (67)	618 (70)	537 (77)	532 (79)
Mean Score Gain Gr 5 to Gr 6 (SD)	12.5 (37)	10.9 (38)	17.31 (38)	15.85 (38)

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

Table 2. Summary Statistics for Unconditional Growth Across Grades

	Reading		Math	
Grade	Mean	SD	Mean	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 means and SDs were standardized relative to the scale score means in grade 4.

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 (.02)	0.10 (.03)		0.51 (.01)	0.48 (.02)

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

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.

Table 5. Correlations Between Grade 6 School Effects and %FRL by Model and Test Subject.

	LM	CP1	CP2
Reading	-.26	-.52	-.43
Math	-.20	-.45	-.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.

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 means and SDs of school effects. The category "+" represents a school with an estimated value-added effect that remains above 0 after two posterior SDs 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 SDs 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 have been added to its posterior mean.

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

			CP1					CP2		
Grade			–	0	+			–	0	+
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 means and SDs of school effects. The category "+" represents a school with an estimated value-added effect that remains above 0 after two posterior SDs 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 SDs 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 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		LM to CP2	
			(+/−)to0	0to(+/−)	(+/−)to0	0to(+/−)
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%

Figure 1. Scatterplots of Estimated School Effects Across Models with and without Assumption of Full Persistence: Reading Tests

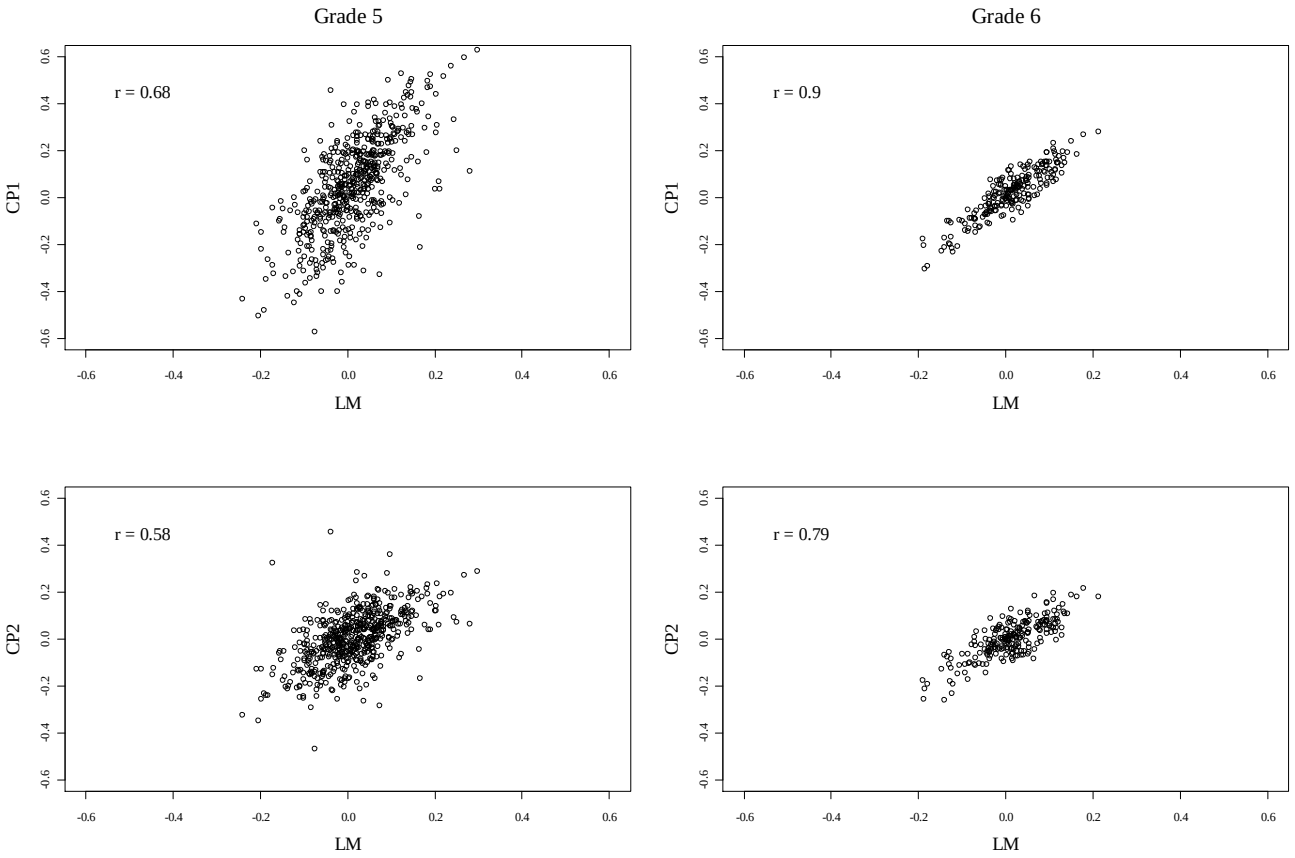

Figure 2. Scatterplots of Estimated School Effects Across Models with and without Assumption of Full Persistence: Math Tests