The Persistence of Value-Added School Effects

Derek C. Briggs
Jonathan P. Weeks

University of Colorado at Boulder

March 2008

Paper presented at the annual meeting of the American Educational Research Association, March 27, 2008.

The research reported in this paper was supported by a grant from the Carnegie Corporation.

Abstract

This paper describes the results from empirical investigation as to the persistence of value-added estimates of school effects. We apply specify six different versions of the variable persistence model as implemented by Lockwood, McCaffrey, Mariano & Setodji (2007) to five years of longitudinal test score data. Our sample consists of 17,839 students in a single state who were enrolled in grades 4 through 8 during the years of 2001 through 2005. Our different specifications of the models varied as a function of (1) constraints placed upon parameters that represent the persistence of school effects over time, and (2) assumptions made about the independence of school effects over time. We find that subsequent inferences about school effectiveness are quite sensitive to choices made in the parameterization of persistence. The correlations of school effects across models is positive, but tends to be only moderate in magnitude. Subsequent inferences are also sensitive to assumptions made about the independence of school effects because there is an interaction between estimates of persistence parameters and the covariance matrix associated with school effects.

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 other value-added models used to estimate teacher effects, school effects, or both, could be expressed as restricted versions of the variable persistence model. The variable persistence model relaxes an implicit assumption made by most value-added models—that the effect of a teacher or school on a student's achievement persists undiminished over time. Intuitively, this assumption will often be implausible, and in a more recently published study, Lockwood, McCaffrey, Mariano & Setodji (2007) provide empirical evidence that the contribution of a teacher two or more years removed from a student's current level of achievement 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 are sensitive to the way that persistence is parameterized longitudinally. Because the precision of estimated effects appears to be much greater under the variable persistence model relative to models that assume complete persistence[1], highly effective or ineffective teachers are more likely to be distinguished from the "average" teacher.

[1] This assumption is made most prominently by what is known as the "layered model" (Sanders, Saxton & Horn, 1996).

To our knowledge, there have been no published applications of the variable persistence model to longitudinal data in which schools, rather than teachers, are the unit of analysis. The motivation for our present study was to fill this void, and we started by posing the following research question: To what extent do conclusions about school effectiveness change when a variable persistence value-added model is used to estimate longitudinal school effects relative to a value-added model that assumes complete persistence?

In the process of addressing this question, we have made two interesting discoveries. First, in the context of data where schools constitute the adjacent level in which students are nested, the variable persistence model must be specified and estimated in a constrained form because the fully parameterized version is not (in principal at least), identifiable. However, the best way to constrain the model is unclear, and this choice can have a big impact on subsequent inferences about estimated school effects. Second, whether the unit of analysis is teachers or schools, and whether effects are assumed to persist completely or vary with (or without) constraints, it is typically assumed that such effects are independent across time. We have found an apparent interaction between this assumption and the magnitude of the school persistence parameter, and this interaction raises some questions about how the latter parameter should be interpreted. The purpose of this presentation is to demonstrate how the answer to our motivating research question varies in the context of these two discoveries.

Methods

The Variable Persistence Model

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

$$Y_{it} = \mu_t + \sum_{t^* \leq t} \alpha_{tt^*} \boldsymbol{\theta}_{t^*} + \varepsilon_{it} . \tag{1}$$

In equation 1, Y_{it} represents the test score of student *i* in year *t*, *t* = 1, …, *T*, and the parameter μ_t denotes the test score mean for a given year. The vector $\boldsymbol{\theta}_t$ represents the collection of school effects[2] for each year, and the parameter α_{tt^*} captures the persistence of the school effects $\boldsymbol{\theta}_{t^*}$ in year *t* (given that $t^* \leq t$). Finally, ε_{it} represents the test score residual associated with student *i* in year *t*. Under the variable persistence model both $\boldsymbol{\theta}_t$ and ε_{it} are assumed to be independent latent random variables, where $\varepsilon_{it} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ and $\theta_t \sim N(\mathbf{0}, \tau)$. We note the following to motivate our subsequent analyses:

1. 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

[2] The term "residual" is actually more appropriate characterization of $\boldsymbol{\theta}_t$ than the term "effect," but we use the latter to be consistent with the literature.

here in order to focus attention on the relationship between the persistence parameters and schools effects.

2. When a complete persistence model is being specified, all persistence parameters are set equal to 1 ($\alpha_{tt*} \equiv 1$ for all $t* \leq t$). When Lockwood et al. used the variable persistence model to estimate teacher effects, the only constraint placed on the persistence parameters was that there is no decay in the current effect of a teacher on student achievement ($\alpha_{tt} \equiv 1$). In the context of applying the model to estimate school effects, because cohorts of students only mix substantially in the transition from elementary to middle school, constraints in addition to $\alpha_{tt} \equiv 1$ must be imposed in order to identify the persistence parameters. Two reasonable possibilities, which we explore in our analysis, include (a) constraining $\alpha_{tt*} \equiv \alpha$ for all $t* < t$, or (b) constraining $\alpha_{t1} \equiv 1$ along with $\alpha_{tt*} \equiv \alpha$ for all $(1 < t* < t)$. The second of these constraints distinguishes between base year school differences ($\boldsymbol{\theta}_1$) and value-added school effects ($\boldsymbol{\theta}_2, \boldsymbol{\theta}_3, ..., \boldsymbol{\theta}_t$) because we should expect that the persistence of the former will be considerably larger than the latter.
3. In previous empirical applications of the variable persistence model, while ε_{it} is given a completely unstructured covariance matrix $(\boldsymbol{\Sigma})$, the variance components for $\boldsymbol{\theta}_t$ have been assumed to be independent across time, hence only the diagonal of the associated covariance matrix $(\boldsymbol{\tau})$ is estimated. This assumption, whether made in the context of estimating teacher or school effects, seems tenuous. For example, one might suspect that a school that has a positive effect on the reading

achievement of its students in grade 6 is also likely to have positive effect on the same students the next year when they are in grade 7.

4. The parameters of the variable persistence model have been estimated using maximum likelihood based methods by McCaffrey et al. (2004), and using Bayesian methods with MCMC estimation by Lockwood et al. (2007). In our analysis we take a Bayesian approach using MCMC estimation with the package "openbugs" in the R statistical environment.

Data Sources

The longitudinal data for this study come from a convenience sample of roughly 37 school districts in a mid-sized state. The students enrolled in the schools within these districts represent about 45% of the state's student population. The longitudinal cohort under analysis includes those students who were in the 4th grade as of 2001 and in the 8th grade as of 2005. We restricted this sample 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 17,839 students who attended 191 different elementary schools and 65 different middle schools from 2001 to 2005. Demographic variables were available to characterize each student, along with scale scores from the annual administration of the state's standardized reading assessment[3]. Summary statistics for demographic and test score information for the full population of students was taken from the relevant annual reports from this state's assessment program, data which is

[3] No school level variables were available, and all district identifiers were removed from the data we were provided, making it impossible to take into account the clustering of schools within common districts.

made publicly available each year. A comparison of these descriptive statistics with those for our longitudinal sample suggests that our sample is generally representative of students in the state as a whole over the same time period. For example, on average 28% of students in the longitudinal sample were black and Hispanic, and 28% were eligible for free or reduced lunches. The respective numbers for the state population were 31% and 32%.

The test scores that serve as the outcome measures in our analyses are derived from responses to a mixture of multiple-choice and constructed-response items. These scores are calibrated onto a vertical scale by the state's test developer with the assumption that the resulting scale has a consistent interpretation over time[4]. For ease of interpretation in the analysis that follows, test scores have been standardized to have a grand mean across all five years of 0 and a standard deviation of 1.

		Assumptions about School Effects	
		Independent	Correlated
Constraints on Persistence Parameters	Complete Persistence $\alpha_{tt*} \equiv 1$ for all $t* \le t$	Model 1: cp.i	Model 4: cp.c
	Constrained Variable Persistence 1 $\alpha_{tt*} \equiv 1$ for all $t* = t$ $\alpha_{tt*} \equiv \alpha$ for all $t* < t$	Model 2: cvp.i1	Model 5: cvp.c1
	Constrained Variable Persistence 2 $\alpha_{tt*} \equiv 1$ for all $t* = t$ $\alpha_{tt*} \equiv \alpha$ for all $(1 < t* < t)$ $\alpha_{t1} \equiv 1$	Model 3: cvp.i2	Model 6: cp.c2

Table 1. Model Specifications Applied to Longitudinal Data

[4] The assumption that this can be accomplished using typical vertical scaling approaches is itself somewhat open to question. We address the sensitivity of this assumption as part of a related study in Briggs, Weeks & Wiley, 2008. Also see Martineau & Reckase, 2006.

We specified six different versions of the model characterized by equation 1 to estimate school effects from our longitudinal data. The different specifications of the model, summarized in Table 1, varied with respect to the way that the persistence of school effects was parameterized (i.e., rows), and assumptions made about the independence of school effects across time (i.e., columns). Non-informative prior distributions were specified for all model parameters, and initial values were either generated randomly, or chosen on the basis of our theoretical understanding of plausible parameter values. 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 "M3" procedure taken by Lockwood et al (p. 135-136) in the context of estimating teacher effects. All models in which school effects are assumed to be independent where estimated on the basis of a sample burn-in of 2,500 followed by 5,000 iterations of 3 different MCMC chains. For models in which school effects are allowed to be correlated, a burn-in of 5,000 iterations was needed to ensure convergence to a stationary distribution.

Results

The full results from specifying the different models described above are provided in the appendix in Tables A1 and A2. Below we focus on noteworthy aspects of these results.

Impact of Constraints on Persistence Parameters

Under the complete persistence model (model 1), it is assumed that $\alpha_{tt^*} \equiv 1$. In contrast, when a single value for α is estimated under the constrained variable persistence model (model 2), this value is considerably smaller that 1 at $\alpha = 0.53$. This result, at first glance, appears consistent with the findings presented by Lockwood et al. (2007). However, in the present context it appears that this is largely an artifact of the decision to constrain the persistence parameters for both base year school differences and value-added school effects to be equal. When the persistence parameter associated with $\boldsymbol{\theta}_1$ is constrained to equal 1 (model 3), the estimate for α drops to 0. This pattern for the relative change in the estimates for α is also observed when models 5 and 6 are compared to model 4 (conditions where the assumption of independent school effects has been relaxed). The impact of these different estimates for α on conclusions about school effectiveness can be seen in Table 2.

Independent School Effects			
Grade	cor(cp.i, cvp.i1)	cor(cp.i ,cvp.i2)	cor(cvp.i1, cvp.i2)
5	0.58	0.60	0.84
6	0.67	0.65	0.91
7	0.85	0.61	0.89
8	0.95	0.73	0.88

Correlated School Effects			
Grade	cor(cp.c, cvp.c1)	cor(cp.c ,cvp.c2)	cor(cvp.c1, cvp.c2)
5	0.45	0.51	0.91
6	0.47	0.47	0.92
7	0.62	0.35	0.89
8	0.85	0.41	0.77

Note: “cp” = complete persistence, “cvp” = constrained variable persistence

Table 2. Correlations of Estimated School Effects by Model

While in most cases there is a moderate to strong positive correlation of estimated school effects across models, it is clear that the choice of model will have a strong impact on conclusions about school effectiveness—some schools that appear effective under one model may not appear effective under another. This point is demonstrated visually in Figure 1. The y-axes of the three panels within Figure 1 represents the scale of the grade 5 value-added school effects estimated for models 1 to 3 respectively. (The full collection of these plots for each grade and model are provided in the appendix). Along each x-axis the 191 individual schools in the grade 5 sample are ranked from lowest to highest on the basis of their value-added effects. Each point on the plot represents the posterior mean of a specific school, and the vertical bars above and below these points represent the 95% credibility interval for these estimates. Schools for which these intervals do not cross the horizontal line referencing the sample average effect of 0 are ones that could be "safely" classified as above or below average in their effectiveness. Under the complete persistence model (model 1), 23% (44 out 191 schools) can be classified as above or below average. Our first specification of a constrained variable persistence model (model 2), leads to a dramatic increase in our ability to distinguish effective and ineffective schools: 50% of the schools can now be classified as above or below average. Yet this advantage largely disappears with our second specification, where only 29% are classified as above or below average. The differences in school classifications as a function of models 1 and 2 are summarized in Table 3.

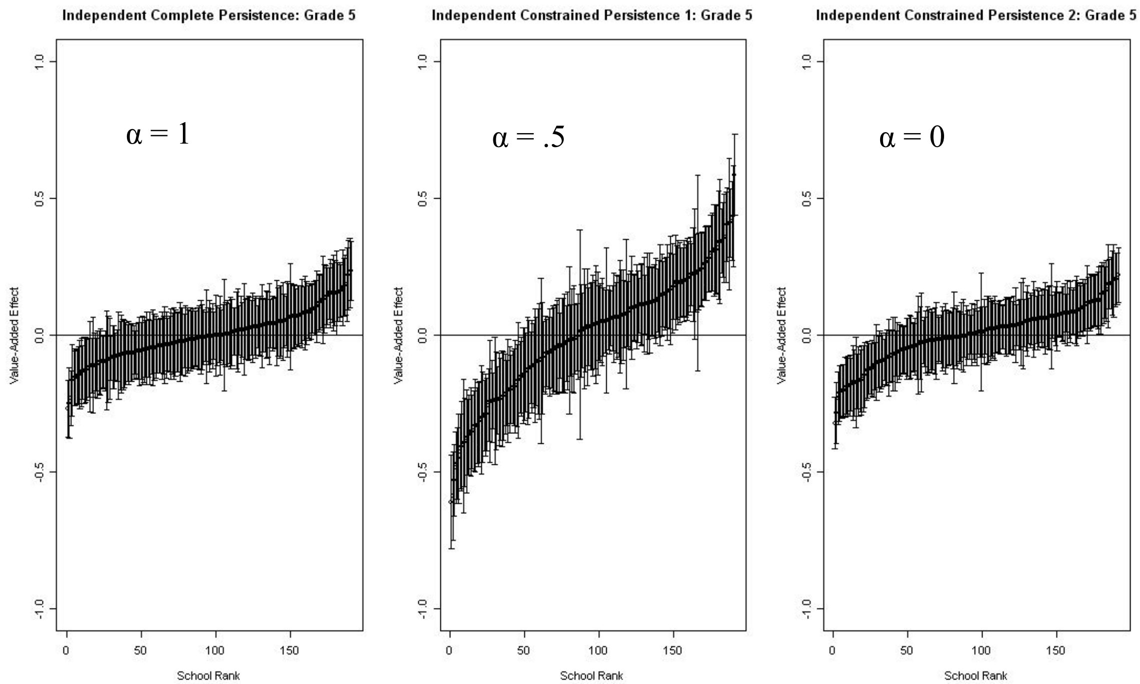


Figure 1. Caterpillar plots of Estimated School Effects for Complete Persistence and Constrained Variable Persistence Models, Grade 5.

		Complete Persistence Not Assumed		
	Effect	–	0	+
Complete Persistence Assumed	–	13	6	1
	0	34	82	31
	+	0	8	16
Note: – indicates that the estimated school effect would be flagged as likely to be negative, 0 indicates the estimated school effect could not be reasonably distinguished from zero, and + indicates that the estimated school effect would be flagged as likely to be positive.				

Table 3. Cross-Tabulation of School Effects on Students in Grade 5.

Impact of Relaxing the Assumption of Independent School Effects

The results from our analyses indicate that there is an apparent interaction between the parameterization of α_{tt*} and the covariance matrix of school effects. One symptom of this is the fact that estimates for α under models 5 and 6 are .11 and .27 lower than the estimates found for models 2 and 3 respectively, where the only difference between the models is the assumption of independence for $\boldsymbol{\theta}_t$ across time. A second symptom is the pattern of estimated values for the correlations of $\boldsymbol{\theta}_t$ across time. Under the constrained persistence model these values tend to be small and negative. Under the two specifications of the constrained persistence models these values tend to be moderately strong and positive.

One possible explanation for these results may be found through analogy to multidimensional item response theory (MIRT), where the persistence parameter α may be analogous to the multidimensional item discrimination parameter. The latter can only be estimated in MIRT applications after placing constraints on the covariance matrix for person ability (c.f., Yao & Schwartz, 2006).

Discussion

Value-added modeling is becoming increasingly popular as a tool to be used within educational accountability systems. In the context of modeling teacher effects, recent research has suggested that decisions about how to parameterize the persistence of effects over time can have a substantial impact on classification decisions. This study is the first to examine this issue within the context of modeling school effects, and suggests that subsequent classification results are particularly sensitive to key decisions made about the specification of the underlying statistical model.

References

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.

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.

Yao, L. & Schwartz, R. D. (2006) A multidimensional partial credit model with associated item and test statistics: an application to mixed-format tests. *Applied Psychological Measurement*. Vol 30(6), 469-492.

APPENDIX

	Independent School Effects						Correlated School Effects					
	cpi.i		cvp.i1		cvp.i2		cp.c		cvp.c1		cvp.c2	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
Sigma.4	0.98	0.11	0.97	0.11	0.98	0.11	0.98	0.11	0.97	0.11	0.98	0.11
Sigma.5	0.98	0.11	0.97	0.11	0.98	0.11	0.98	0.11	0.97	0.11	0.97	0.11
Sigma.6	0.98	0.11	0.97	0.11	0.98	0.11	0.98	0.11	0.97	0.11	0.97	0.11
Sigma.7	0.99	0.11	0.98	0.11	0.99	0.11	0.99	0.11	0.98	0.11	0.98	0.11
Sigma.8	0.93	0.10	0.93	0.10	0.93	0.10	0.93	0.10	0.92	0.10	0.92	0.10
tau.4	0.52	0.03	0.63	0.04	0.51	0.03	0.49	0.17	0.58	0.21	0.48	0.17
tau.5	0.11	0.01	0.24	0.02	0.11	0.01	0.12	0.04	0.37	0.14	0.13	0.05
tau.6	0.09	0.01	0.12	0.01	0.09	0.01	0.12	0.05	0.17	0.08	0.13	0.06
tau.7	0.07	0.01	0.09	0.01	0.09	0.01	0.10	0.05	0.13	0.06	0.15	0.07
tau.8	0.09	0.01	0.09	0.01	0.12	0.01	0.12	0.05	0.12	0.05	0.20	0.09
cor(tau.4,tau.5)	NA	NA	NA	NA	NA	NA	0.10	NA	0.93	NA	0.65	NA
cor(tau.6,tau.7)	NA	NA	NA	NA	NA	NA	-0.22	NA	0.45	NA	0.60	NA
cor(tau.6,tau.8)	NA	NA	NA	NA	NA	NA	0.07	NA	0.26	NA	0.61	NA
cor(tau.7,tau.8)	NA	NA	NA	NA	NA	NA	-0.15	NA	0.25	NA	0.73	NA
alpha	1.00	NA	0.53	0.02	-0.06	0.07	1.00	NA	0.42	0.02	-0.33	0.05

Table A-1. Variance Component Estimates by Model

Grade 4	cp.i	cp.c	cvp.i1	cvp.c1	cvp.i2	cvp.c2
cp.i	1	1	0.96	0.95	0.99	0.99
cp.c	1	1	0.97	0.95	0.99	0.99
cvp.i1	0.96	0.97	1	0.99	0.94	0.94
cvp.c1	0.95	0.95	0.99	1	0.93	0.94
cvp.i2	0.99	0.99	0.94	0.93	1	1
cvp.c2	0.99	0.99	0.94	0.94	1	1

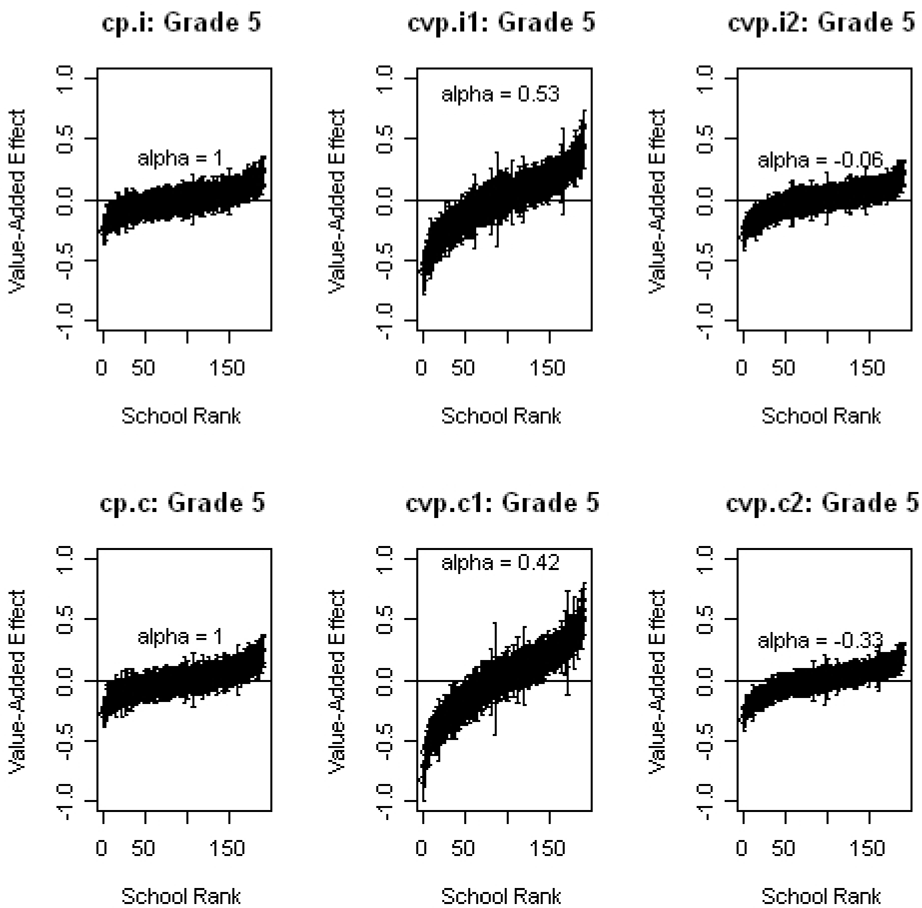
Grade 5	cp.i	cp.c	cvp.i1	cvp.c1	cvp.i2	cvp.c2
cp.i	1	1	0.58	0.42	0.6	0.48
cp.c	1	1	0.6	0.45	0.61	0.51
cvp.i1	0.58	0.6	1	0.98	0.84	0.91
cvp.c1	0.42	0.45	0.98	1	0.8	0.91
cvp.i2	0.6	0.61	0.84	0.8	1	0.96
cvp.c2	0.48	0.51	0.91	0.91	0.96	1

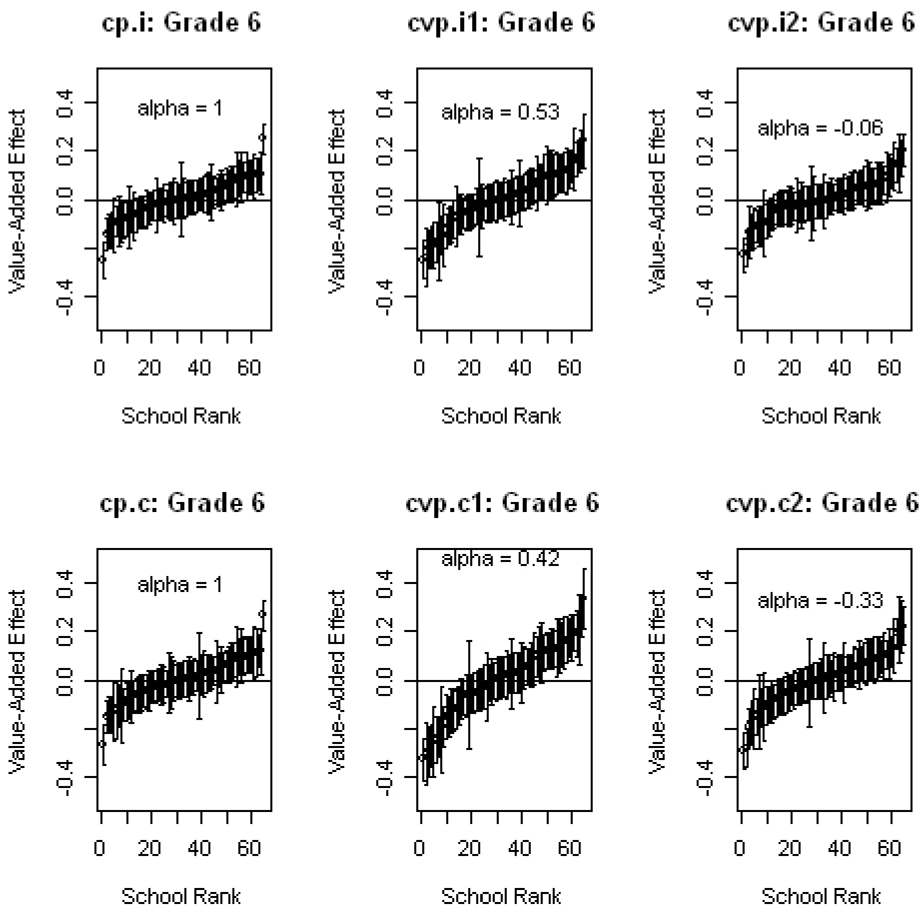
Grade 6	cp.i	cp.c	cvp.i1	cvp.c1	cvp.i2	cvp.c2
cp.i	1	1	0.67	0.46	0.65	0.46
cp.c	1	1	0.68	0.47	0.66	0.47
cvp.i1	0.67	0.68	1	0.96	0.91	0.91
cvp.c1	0.46	0.47	0.96	1	0.85	0.92
cvp.i2	0.65	0.66	0.91	0.85	1	0.95
cvp.c2	0.46	0.47	0.91	0.92	0.95	1

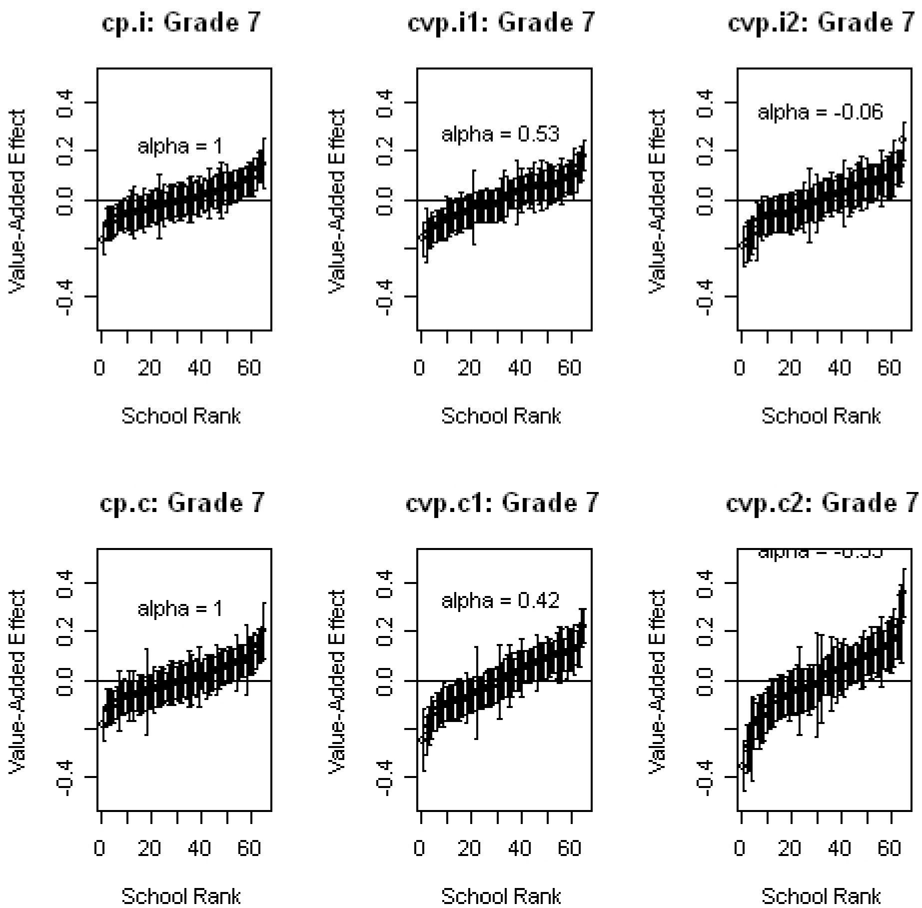
Grade 7	cp.i	cp.c	cvp.i1	cvp.c1	cvp.i2	cvp.c2
cp.i	1	0.99	0.85	0.64	0.61	0.37
cp.c	0.99	1	0.83	0.62	0.59	0.35
cvp.i1	0.85	0.83	1	0.94	0.89	0.77
cvp.c1	0.64	0.62	0.94	1	0.89	0.89
cvp.i2	0.61	0.59	0.89	0.89	1	0.92
cvp.c2	0.37	0.35	0.77	0.89	0.92	1

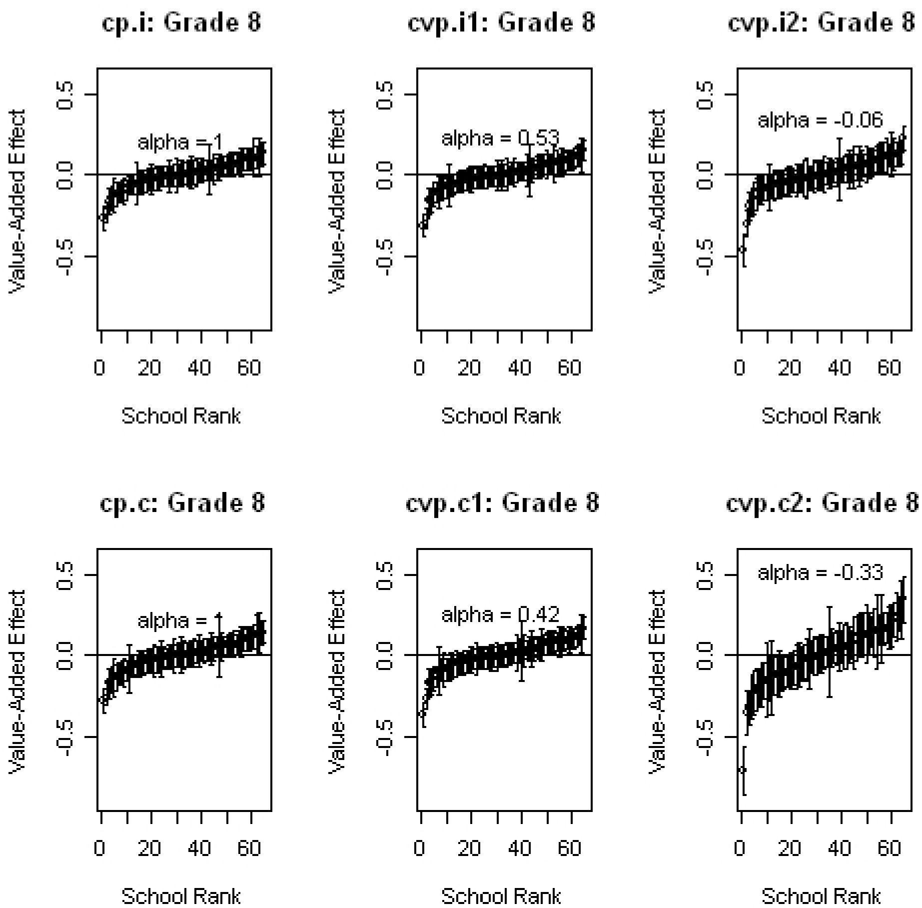
Grade 8	cp.i	cp.c	cvp.i1	cvp.c1	cvp.i2	cvp.c2
cp.i	1	1	0.95	0.86	0.73	0.43
cp.c	1	1	0.94	0.85	0.71	0.41
cvp.i1	0.95	0.94	1	0.97	0.88	0.63
cvp.c1	0.86	0.85	0.97	1	0.92	0.77
cvp.i2	0.73	0.71	0.88	0.92	1	0.89
cvp.c2	0.43	0.41	0.63	0.77	0.89	1

Table A-2. Correlations of School Effects by Grade Across Models


cp.i: Grade 5
alpha = 1
cvp.i1: Grade 5
alpha = 0.53
cvp.i2: Grade 5
alpha = -0.06
cp.c: Grade 5
alpha = 1
cvp.c1: Grade 5
alpha = 0.42
cvp.c2: Grade 5
alpha = -0.33
Value-Added Effect
School Rank
-1.0
-0.5
0.0
0.5
1.0
0
50
150


cp.i: Grade 6
alpha = 1
cvp.i1: Grade 6
alpha = 0.53
cvp.i2: Grade 6
alpha = -0.06
cp.c: Grade 6
alpha = 1
cvp.c1: Grade 6
alpha = 0.42
cvp.c2: Grade 6
alpha = -0.33
Value-Added Effect
-0.4
0.0
0.2
0.4
School Rank
0
20
40
60


cp.i: Grade 7
alpha = 1
cvp.i1: Grade 7
alpha = 0.53
cvp.i2: Grade 7
alpha = -0.06
cp.c: Grade 7
alpha = 1
cvp.c1: Grade 7
alpha = 0.42
cvp.c2: Grade 7
Value-Added Effect
School Rank
-0.4
0.0
0.2
0.4
0
20
40
60


cp.i: Grade 8
alpha = 1
cvp.i1: Grade 8
alpha = 0.53
cvp.i2: Grade 8
alpha = -0.06
cp.c: Grade 8
alpha = 1
cvp.c1: Grade 8
alpha = 0.42
cvp.c2: Grade 8
alpha = -0.33
Value-Added Effect
School Rank
0.5
0.0
-0.5
0
20
40
60