NATIONAL CENTER ON

Performance Incentives

Value-Added and Other Methods for Measuring School Performance

Robert H. Meyer
Michael S. Christian

Prepared for *Performance Incentives: Their Growing Impact on American K-12 Education* in Nashville, Tennessee on February 29, 2008

Working Paper 2008-17
February 2008

LED BY
VANDERBILT PEABODY COLLEGE

IN COOPERATION WITH:
RAND CORPORATION
Mizzou University of Missouri - Columbia

The National Center on Performance Incentives (NCPI) is charged by the federal government with exercising leadership on performance incentives in education. Established in 2006 through a major research and development grant from the United States Department of Education's Institute of Education Sciences (IES), NCPI conducts scientific, comprehensive, and independent studies on the individual and institutional effects of performance incentives in education. A signature activity of the center is the conduct of two randomized field trials offering student achievement-related bonuses to teachers. The Center is committed to air and rigorous research in an effort to provide the field of education with reliable knowledge to guide policy and practice.

The Center is housed in the Learning Sciences Institute on the campus of Vanderbilt University's Peabody College. The Center's management under the Learning Sciences Institute, along with the National Center on School Choice, makes Vanderbilt the only higher education institution to house two federal research and development centers supported by the Institute of Education Services.

This working paper was supported by the National Center on Performance Incentives, which is funded by the United States Department of Education's Institute of Education Sciences (R30SA06034). This is a draft version of a paper that will be presented at a national conference, *Performance Incentives; Their Growing Impact on American K-12 Education, in Nashville,* Tennessee on February 28-29, 2008. The authors would like to thank the Chicago Community Trust, the Institute for Education Sciences, the Joyce Foundation, and the Spencer Foundation for generously supporting their research on value-added methods and applications. They gratefully acknowledge support from the Center for Education Compensation Research. The authors would also like to thank the many districts and states that have worked with us in the development and application of value-added models as well as Lali Abril for very helpful assistance in preparing this paper. The views expressed in this paper do not necessarily reflect those of sponsoring agencies or individuals acknowledged. Any errors remain the sole responsibility of the author.

Please visit **www.performanceincentives.org** to learn more about our program of research and recent publications.

Value-Added and Other Methods for Measuring School Performance

ROBERT H. MEYER
MICHAEL S. CHRISTIAN

Value-Added and Other Methods for Measuring School Performance: An Analysis of Performance Measurement Strategies in Teacher Incentive Fund Proposals

Robert H. Meyer and Michael S. Christian

One of the central challenges of designing and implementing a performance pay program is developing an approach for determining which schools, teachers, and administrators have performed well enough to have earned a bonus. The U.S. Department of Education has specified that performance pay programs supported by the federal Teacher Incentive Fund (TIF) program "must include both gains in student achievement as well as classroom evaluations...and provide educators with incentives to take on additional responsibilities and leadership roles."[1] Within those broad guidelines, TIF grantees have substantial latitude to create incentive pay systems that fit their specific needs.

In this paper we review the methods proposed by TIF grantees for measuring the performance of schools, teachers, and administrators with respect to student achievement. We recognize (and expect) that many grantees are likely to improve and modify their performance measurement approaches over time, as they and their stakeholders develop a better understanding of the available options for measuring performance and as better data becomes available. One of the major objectives of this paper is to evaluate the different performance measurement approaches in terms of a specific statistical standard – a value-added model (VAM). To simplify our analysis, we focus primarily on value-added models of grade-level performance. Most, if not all, of our conclusions also apply to value-added models of classroom/teacher performance.

---

[1] U.S. Department of Education, *Teacher Incentive Fund Program At A Glance* <http://www.ed.gov/programs/teacherincentive/gtepteacherincentive.pdf>

A value-added model is a quasi-experimental statistical model that yields estimates of the contribution of schools (or other educational units) to growth in student achievement (or other student outcomes), controlling for other sources of student achievement growth, including prior student achievement and student and family characteristics. The model produces estimates of school performance – value-added indicators -- under the counterfactual assumption that all schools serve the same group of students. The objective is to facilitate valid and fair comparisons of student outcomes across schools given that the schools may serve very different student populations.

The conclusions of the paper are presented in the final section. Below we summarize the primary measurement methods proposed by TIF grantees.

I. Approaches to Measuring the Performance of Schools, Teachers, and Administrators

As mentioned above, Teacher Incentive Fund grantees have substantial latitude to create incentive pay systems that fit their specific needs. As a result, there is substantial variation across grantees in the approaches proposed to measure the performance of schools, teachers, and administrators. The distribution of approaches across the 34 TIF grantees is presented in Table 1.

Table 1. Distribution of Approaches to Measuring School Performance

Performance Measurement Approach	Count
Value added	17
Gain	2
Movement across proficiency categories	3
Proficiency rates or attainment	5
Gain/movement/proficiency hybrid	6
Individual learning plans	1
Total	34

The distribution in Table 1 is based on a survey of the 34 TIF proposals. The classification of the proposals is broad, and there are subtle but important differences between individual proposals within each category. The breadth of the categories makes it possible to classify proposals in which the approach to using student assessments is vaguely described. Each grantee's approach was classified as being in one and only one category. In cases where multiple approaches to using student assessment data were described, effort was made to identify the approach likely to be the most important; however, some approaches were sufficiently mixed that they could only be categorized as hybrid. When it was known that the approach described in the proposal was technically different from the approach actually used, the approach actually used was recorded in our survey.

Value-added

The most commonly used measure of school performance used among the TIF grantees is value added. Sixteen of the 34 grantees included the use of value added in their proposals, and a seventeenth, Houston Independent School District, has included value added in its program, ASPIRE, since the proposal.[2] Most of the grantees that use value added employ an outside research partner to help measure value added. The most common research partner is SAS, especially in states that have already partnered with SAS to measure value added or have collaborated with SAS in the Teacher Advancement Program (TAP). Other TIF grantees are partnered with Mathematica, RAND, and Wisconsin Center for Education Research. Dallas is unusual among school districts for measuring value added in-house, which it has done since the 1980's.

[2] Houston Independent School District, *Houston Independent School District ASPIRE Awards Model: 2006-2007 School Year,* at <http://www.houstonisd.org/ResearchAccountability/Home/Teacher%20Performance%20Pay/Teacher%20Performance%20Pay/Board%20Items/ASPIRE_AWARDS_OVERVIEW.pdf>

Grantees that do not use value added use a wide variety of measures. These include gains in student test scores, movement by students across proficiency categories, overall levels of proficiency and attainment, and meeting goals outlined in individual learning plans.

Gain

Gain in student test scores is the difference between average student performance on an achievement test in one year minus average performance on the test by the same students in the previous year. Calculating gain requires matching individual students from one year to the next so that improvement on the assessment is measured over the same group of students (the matched sample). Gain is similar to value added in its emphasis on growth in student achievement from one year to the next for the same group of students. In that sense, gain is a value-added-like indicator. It is different from a value-added indicator because it is simpler and because its approach to controlling for previous student performance is determined not by statistical evidence but rather by using an *a priori* assumption that gain is a well-defined and appropriate measure of school performance. We address this assumption in later in the paper.

Gain is used in only a few districts, usually in tandem with other approaches and often with some adjustments. The proposed program at Brooke Charter School in Boston includes incentives for making high average grade levels of progress in reading at the elementary level. The proposed incentive programs in Charlotte-Mecklenburg Schools and Cumberland County Schools in North Carolina are partially based on gain in normalized test scores with an adjustment for regression to the mean.[3] Houston's ASPIRE program includes Comparable

[3] State of North Carolina, Department of Public Instruction, *2006-07 Determining School Status for ABCs*, <http://www.ncpublicschools.org/docs/accountability/reporting/abc/2006-07/determiningschoolstatus0607.pdf>

Improvement, a metric measured by the Texas Education Agency that compares an adjusted version of gain for a school to the adjusted gains of demographically similar schools.[4]

Gain in Proficiency Status: *Movement Across Proficiency Categories*

A measure that is similar to gain that is used in several TIF districts is movement across proficiency categories. Districts who use this measure base their awards on the changes in the proficiency levels of students ("Below Proficient", "Proficient", "Advanced", etc.) from one year to the next, with successful schools and classrooms being the ones in which students move up to higher proficiency levels or at least maintain their previous proficiency level from one year to the next. In some cases, school performance is measured on a point scale, with different amounts of points given to different movements. Points can be based on the magnitude of the movement, with more points given for students who move up rather than maintain; they can also be based on the perceived need for the movement, with more points given for students who move, for example, from below proficient to proficient than for students who move from proficient to advanced.

Programs based on movement across proficiency categories were described in proposals from Chugach School District in Alaska, Hillsborough County Public Schools in Florida, and Harrison County District 2 in Colorado. The proposed Chugach and Hillsborough programs scored different movements differently, with the scores based on value tables developed by state education agencies; the tables used in Hillsborough were developed for Florida's Special Teachers are Rewarded program, while those used in Chugach were developed by the Alaska Department of Education and Early Development. The Harrison proposal, in contrast, treated

[4] Texas Education Agency, 2006 Accountability Manual, <http://www.tea.state.tx.us/perfreport/account/2006/manual/index.html>

positive and negative movements of equal magnitude across twelve proficiency categories as equally good or bad.

Proficiency Rates, Attainment, and Hybrids

Another metric used in several proposals to identify schools and classrooms that have earned their incentives is the proficiency rate: the percentage of students who scored at or above a proficiency threshold. Related to proficiency rates is average attainment, which is equal to the average score of students in the school or classroom on the assessment. These measures are often measured as part of state accountability systems, and do not take past performance into account when attributing school performance to student performance.

Many proposed performance pay systems use a combination of proficiency rates, gain, and/or movement across proficiency categories to determine the recipients of incentives. The Denver ProComp program, which is expanded in its TIF proposal, is based on individual teachers meeting individually prescribed goals, which may be based on gain, movement, or proficiency.[5] In Charlotte-Mecklenburg's proposed program, teachers can only receive incentives if their students make sufficient gains and if a percentage of their students score at proficient or higher, while in the South Dakota's proposed program, school-level incentives are received if schools make *either* sufficient growth in student achievement or adequate yearly progress (AYP) proficiency targets.[6]

Individual Achievement Plans

Finally, two TIF proposals include the development of achievement plans for individual students, with incentive pay given to schools and teachers whose students meet the goals

---

[5] Denver Public Schools, *Checklist for Student Growth Objectives*, *Elementary Classroom Sample Objectives*, *Middle School Sample Objectives*, and *High School Sample Objectives* at <http://my.dpsk12.org/objectives/>

[6] South Dakota Department of Education, South Dakota IncentivesPlus Incentive Award Matrix, <http://www.doe.sd.gov/oatq/incentives_plus/docs/TIF_Matirix11282007.pdf>

outlined in the plans. Individual achievement plans are at the center of the proposal by the Center for Educational Innovation, which describes an incentive program for ten New York City charter schools. A predictive model determines the goals in the individual achievement plans. Individual plans are also included in Alaska's proposal.

Size and Approach

The TIF grantees are a very diverse group, including education agencies of all sizes – see Table 2. As indicated in the table, nine grantees have student enrollments equal to ten thousand students or less.

Table 2. Total Student Enrollment By TIF Grantees, 2005-06

Small grantees (Fewer than 10,000)		**Medium-sized grantees (10,000 to 100,000)**		**Large grantees (More than 100,000)**	
Brooke Charter	276	Harrison County	11,218	Memphis	120,275
Mare Island	723	NLNS Charters	16,601	Charlotte	124,005
Alaska	1,075	Amphitheater	16,768	Prince George's	133,325
Beggs	1,161	Lynwood	18,211	Dallas	161,244
School for Excl.	1,768	South Dakota	20,250	Orange County	175,609
Weld County	2,531	Florence County	22,876	Philadelphia	184,560
NYC Charters	2,718	Pittsburgh	32,506	Ohio	185,044
Eagle County	5,365	Lake County	38,060	Hillsborough	193,757
New Mexico	5,608	South Carolina	41,900	Houston	210,292
		Cumberland	53,201	Miami	362,070
		D.C.	59,616	Chicago	420,982
		Guilford County	68,951		
		Denver	72,312		
		U. of Texas	87,002		

Source: Common Core of Data[7]

[7] Enrollment data for the New Leaders for the New Schools national charter schools project was taken from the proposal. When a grant was made for a program proposed to cover multiple districts, the sum of enrollments across the districts was used; for example, the Ohio enrollment figure is combined enrollment for Cleveland, Cincinnati, Columbus, and Toledo.

Not surprisingly, the proposed performance measurement systems differ systematically with grantee enrollment – see Table 3. Small districts are less likely to use value added and more likely to use simpler approaches in their proposed performance pay programs.

Table 3. Proposed Performance Measurement Systems by Grantee Size

Performance Measurement Approach	Number of grantees using approach			
	Small grantees	Medium grantees	Large grantees	Total
Value added	2	9	6	17
Gain	1	0	1	2
Movement across proficiency categories	1	1	1	3
Proficiency rates or attainment	3	1	1	5
Gain/movement/proficiency hybrid	1	3	2	6
Individual learning plans	1	0	0	1
Total	9	14	11	34

II. A Value-Added Model for Measuring Grade-Level School Performance

In this section we present a specific value-added model and consider design features that enhance the ability of the model to accurately measure school performance at a given grade level.[8] Our purpose is twofold: to stimulate districts and states to think critically about value-added models and to posit a standard that can be used to evaluate other methods for measuring school performance. In selecting a model for this purpose, we considered two different classes of value-added models: (1) models based on two longitudinal achievement outcomes for each student (posttest and pretest) ("T2" models) and (2) models based on three (or more) achievement outcomes for each student ("T3+" models). Although we generally prefer the latter class of models, since they can (in some circumstances) better control for differences across

[8] To simplify the analysis we focus on grade-level value-added models rather than classroom/teacher value-added models. The design challenges for these two models are very similar.

schools in the student-level determinants of achievement growth, in this paper we decided to focus on models based on two achievement outcomes for two reasons. First, it is simpler to explain much of the logic of the value-added method using these models. Second, models based on two achievement outcomes are somewhat better suited to exploring the strengths and limitations of alternative methods of measuring school performance since these methods also are based on two (or fewer) student achievement outcomes. Appendix B provides further discussion of T2 and T3+ value-added models.[9,10]

A "T2" Value-Added Model

We begin by considering a "core" value-added model based on two longitudinal achievement outcomes for each student (posttest and pretest) that, for simplicity, assumes the following: [11]

- Students are tested at the beginning or end of the school year so that growth in student achievement does not cut across two school years.
- Students do not change schools during the school year.
- No students are retained in grade or skip a grade from one school year to the next.

---

[9] In the T2 model, differences across schools in student growth trajectories are captured directly by the student characteristics that are included in the model. In this model, systematic differences in student growth trajectories that are not captured by student characteristics included in the model are absorbed by the estimated value-added effects. As discussed in Appendix B, one of the key advantages of including three or more achievement outcomes for each student is that it is possible (in some circumstances) to better control for differences across schools in the student-level determinants of achievement growth than in a model based on two achievement outcomes.
[10] See also the following for discussion of alternative value-added models: Ballou et al (2004); Boardman and Murnane (1979); Hanushek (2005); McCaffrey et al (2004); Meyer (1996, 2006, 2007); Sanders and Horn (1994); and Willms and Raudenbush (1989).
[11] In the value-added models we have developed for many districts and states, our district and state partners have preferred models that relax these assumptions and appropriately allow for student mobility, mid-year testing, and grade retention. Later in this section we discuss how the core value-added model can be expanded to accommodate specific local conditions.

The model is defined by three equations, a "best linear predictor" (Goldberger, 1991) value-added model defined in terms of true student post and prior achievement – y_{1i} and y_{0i}, respectively – and measurement error models for observed post and prior achievement – Y_{1i} and Y_{0i}, respectively:

$$y_{1i} = \zeta + \lambda y_{0i} + \beta' X_i + \alpha' S_i + e_i \tag{1}$$

$$\begin{aligned} &\text{Pretest measurement error model:} \quad Y_{0i} = y_{0i} + v_{0i} \\ &\text{Posttest measurement error model:} \quad Y_{1i} = y_{1i} + v_{1i} \end{aligned} \tag{2}$$

where v_{1i} and v_{0i} represent the measurement errors in post and prior achievement, respectively. The model includes the following components: true prior achievement with slope parameter λ, X_i = a vector of student characteristics with slope parameter vector β, S_i = a vector of student enrollment indicators, α = a vector of value-added school effects (where α_k = the value-added effect for school *k*), ζ = an intercept, and e_i = the error in predicting post achievement, given the explanatory variables included in the model.[12,13,14]

A model defined in terms of measured achievement Y_{0i} and Y_{1i} can be obtained using equations (1) and (2) as the building blocks. In particular equation (2) implies that y_{0i} and y_{1i} are given by:

---

[12] The model could also include school-level variables such as school average values of student characteristics. We discuss this model option in Appendix D. See Meyer (1996) and Willms and Raudenbush (1989) for further discussion of this issue.

[13] See Appendix C for a discussion of value-added models for multiple growth years.

[14] Since the value-added model presented in the text is a best linear predictor, the error term *e* is uncorrelated with the explanatory variables by definition. See Appendix B for discussion of statistical strategies that allow for possibly systematic differences across schools in the unobserved determinants of student achievement growth.

$$\begin{aligned} y_{0i} &= Y_{0i} - v_{0i} \\ y_{1i} &= Y_{1i} - v_{1i} \end{aligned}. \tag{3}$$

Substituting these equations into (1) yields an equation defined in terms of measured achievement:

$$\mathrm{Y}_{1i} = \zeta + \lambda Y_{0i} + \beta' X_i + \alpha' S_i + \varepsilon_i \tag{4}$$

where the error term includes three parts, the original error component plus both measurement error components:

$$\varepsilon_i = e_i + v_{1i} - \lambda v_{0i} \tag{5}$$

As discussed in Meyer (1992, 1999) and Fuller (1987), estimating equation (4) without controlling for pretest measurement error yields biased estimates of all parameters, including the value-added school effects. This bias stems from the fact that measurement error in prior achievement causes the error term (which includes the measurement component v_{0i}) to be correlated with measured prior achievement. The desired parameters, as defined in equation (1) can be estimated consistently if external information is available on the variance of measurement error for prior achievement. This information is typically reported in the technical manuals for published assessments.

The value-added school effect α_k can be interpreted as the mean effect of school *k* on growth in student achievement after controlling for student characteristics X_i and true prior achievement. The school effect parameter α_k in the value-added model is typically centered around zero (in the benchmark year at each grade level). Value-added indicators reported in this metric are often referred to as the "beat the average" (BTA) indicators, since the average indicator value (equal to zero) corresponds to district (or state) average productivity. Alternatively, value-added can be centered around the average district gain (in the benchmark year at each grade level). In this case, the indicator can be interpreted as the "predicted gain" in achievement for a given school, given the counterfactual assumption that the school serves students who are identical, on average, to the students served by the district as a whole.

The Coefficient on Prior Achievement: λ

One of the important features of the value-added model that we have considered thus far is that it allows for the possibility that the coefficient on prior achievement (λ) is not equal to one. The model would be simpler to estimate if it was appropriate to impose the parameter restriction $\lambda = 1$,[15] but there are at least four factors that could make this restriction invalid. First, λ could be less than one if the "stock" of knowledge, skill, and achievement captured by student assessments is not totally "durable," but rather is subject to decay. Let δ = annual durability rate of student achievement so that the annual decay rate equals $(1-\delta)$ (Meyer, 2006).

Second, λ could differ from one if school resources are allocated differentially to students as a function of true prior achievement, as captured by a resource allocation parameter ρ. If resources are tilted relatively toward low-achieving students – a remediation strategy – then $\rho < 1$. The opposite would be true if resources were tilted toward high-achieving students (Meyer, 2006). Combining these two factors yields a coefficient on prior achievement equal to:

$$\lambda^* = \delta + \rho \tag{6}$$

Third, λ could differ from one (or from λ^*, as defined above) if posttest and pretest scores are measured on different scales, perhaps because the assessments administered in different grades are from different vendors and scored on different test scales. In this case, the

[15] Imposing the parameter restriction $\lambda = 1$ yields the following model: $\mathrm{Y}_{1i} - Y_{0i} = \zeta + \beta' X_i + \alpha' S_i + \varepsilon_i$. In this restricted model, the error term is not correlated with the right hand side variables and thus it is not necessary to control for measurement error when estimating the model. The presence of test measurement error in the error term does, of course, have the undesired effect of reducing the precision of parameter estimates.

coefficient on prior achievement partially reflects the difference in scale units between the pretest and posttest. Fourth, the different methods used to score assessments could, in effect, transform posttest and pretest scores so that the relationship between post and prior achievement is nonlinear. In this case, a linear value-added model might still provide a reasonably accurate approximation of the achievement growth process, but the coefficient on prior achievement (as in the case of the third point) is affected by the test scaling.

To see this, consider a value-added model – with the same structure as equation (1) – defined in terms of latent unobserved test scores z_1 (true latent post achievement) and z_0 (true latent prior achievement):

$$z_{1i} = \zeta^* + \lambda^* z_{0i} + \beta^{*\prime} X_i + \alpha^{*\prime} S_i + e_i^* \quad (7)$$

The parameters in this model are distinguished from the parameters of equation (1) by the superscript "*." True scores (measured without error), corresponding to measured pretest and posttest scores, are given by (possibly nonlinear) transformations of the latent achievement scores:

$$\begin{aligned} y_{0i} &= f_0(z_{0i}) \\ y_{1i} &= f_1(z_{1i}) \end{aligned}. \quad (8)$$

Note that the transformation functions could be the same, if the properties of the scoring/scaling algorithm are similar at different grades. Latent prior achievement is correspondingly given by:

$$z_{0i} = f_0^{-1}(y_{0i}) \quad (9)$$

where f_0^{-1} is the inverse function.

Given equations (7) - (9), the value-added model, written in terms of true scores y_{1i} and y_{0i} is given by:

$$y_{1i} = f_1\left[\zeta^* + \lambda^* f_0^{-1}(y_{0i}) + \beta^{*\prime} X_i + \alpha^{*\prime} S_i + e_i^*\right] \tag{10}$$

This model is not actionable since the transformation functions are unknown, but it can be approximated by a linear (Taylor series) approximation around the district (or state) means of the regressors:

$$y_{1i} = \overline{f}_1 + \left(\frac{m_1}{m_0}\right)\lambda^*(y_{0i} - \overline{\overline{y}}_0) + m_1\left[\beta^{*\prime}(X_i - \overline{\overline{X}}) + \alpha^{*\prime} S_i + e_i^*\right] \tag{11}$$

where $\overline{\overline{y}}_0$ = the district mean of prior achievement, $\overline{\overline{X}}$ = the district mean of student characteristics, $\overline{f}_1$ = the function f_1 evaluated at district means, and $m_1 \equiv \overline{f}_1'$ and $m_0 \equiv \overline{f}_0'$ (scale multipliers) are the first derivatives of the functions f_1 and f_0 evaluated at district means, respectively.[16] The key result is that equation (11) is equivalent to the linear value-added model defined by (1); the parameters of this model adjust in response to the particular scaling algorithms used to score/scale assessments. In particular, the coefficient on prior achievement is the product of the ratio of two scale multipliers (m_1 / m_0) and the latent pretest coefficient, and the latent school performance effect and the other components of the model are multiplied by the posttest multiplier m_1:

$$\lambda = \left(\frac{m_1}{m_0}\right)\lambda^* \equiv \left(\frac{m_1}{m_0}\right)(\delta + \rho) \tag{12}$$

$$\begin{aligned} \alpha &= m_1\alpha^* \\ \beta &= m_1\beta^* \\ e_i &= m_1 e_i^* \end{aligned} \tag{13}$$

The degree to which the latent parameters are affected by implicit scale transformation depends on the shape of the scaling functions *f*. As an example, consider the scaling functions portrayed

[16] The district means of the other variables are zero, given the normalizations used in the model.

in Figure 1 (panels a and b).[17] The scaling functions in panel (a) exhibit negative curvature and thus the ratios $\left(\frac{m_1}{m_0}\right)$ and posttest multipliers m_1 are less than one. The opposite is true in panel (b) where the scaling functions exhibit positive curvature. The bottom line is that the parameters of a value-added model are not invariant to the scaling algorithms used to score the pretest and posttest assessments.

In summary, we have considered four factors that could make it problematic to impose the parameter restriction that the coefficient in prior achievement equals one: durability/decay in achievement, differential resource allocation, differences in the pretest and posttest test scales, and nonlinearity in the test scaling algorithm. Later in the paper we consider the consequences of imposing the restriction that $\lambda = 1$ and other parameter restrictions.

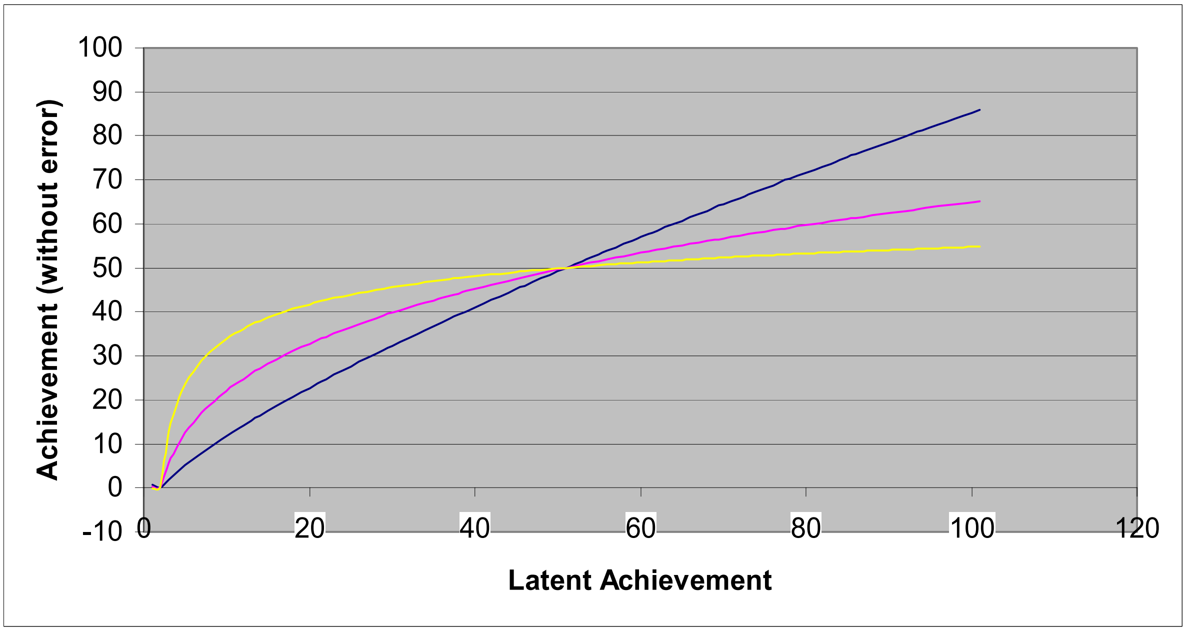
[17] The scaling functions are all examples of Box-Cox transformations (with different transformation parameters), modified to fit on the same graph. The Box-Cox transformation is given by:

$$y = f(z) = \frac{z^{\eta} - 1}{\eta}$$

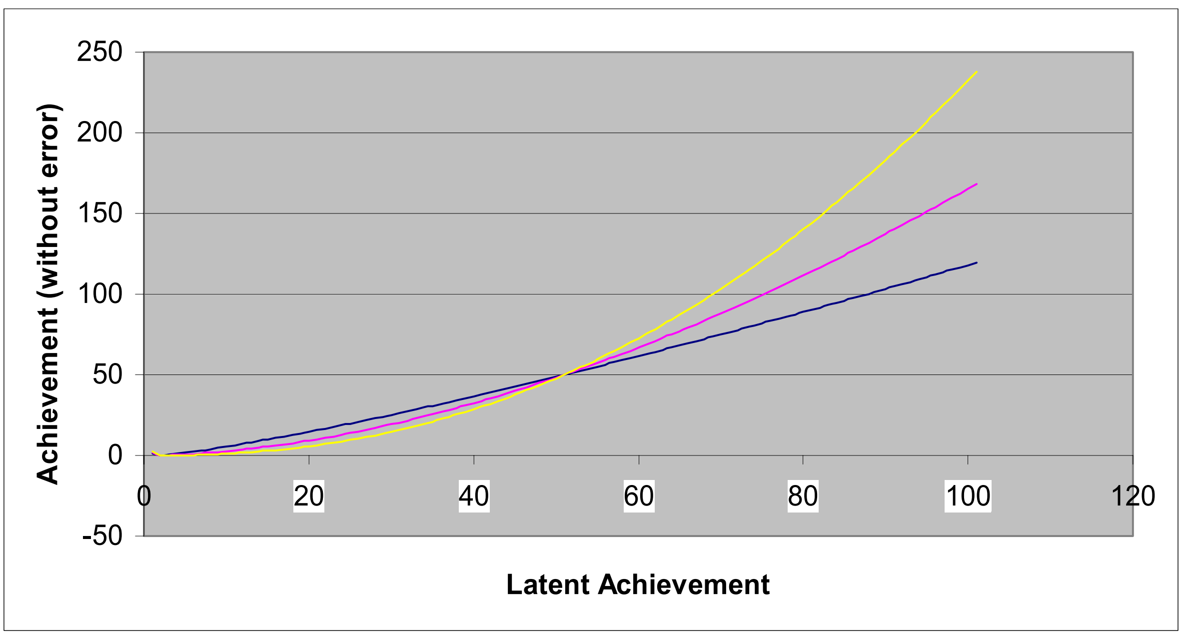
with scaling parameter η

Figure 1. Hypothetical Transformations of Latent Test Scales

(a) Negative Curvature



(b) Positive Curvature



Does Achievement Growth Differ for Students with Different Student Characteristics?

A second important feature of the value-added model presented above is that it includes explicit measures of student characteristics (as represented by X_i in equation (1)). Since most district or state value-added systems are based on administrative data bases (as opposed to special purpose data collections), value-added models generally include a limited number of student measures, for example, poverty status (participation in free or reduced-price lunch), participation in special education, participation in an English Language Learner (ELL) program, gender, or race/ethnicity.

Including measures of student characteristics in a value-added model serves two purposes. First, including these measures in a value-added model makes it possible to decompose district-wide differences in achievement gain by student group (e.g., low and high poverty) into distinct "growth gap factors":

- Direct (within school) effects.
- Indirect (within school) effects via correlation with other student characteristics (for example, participation in special education).
- Indirect (between school) effects via differences in the productivity of schools that serve different groups of students.

These statistics are important for policy purposes because they capture the channels via which differences in student attainment by poverty status and other student characteristics emerge over time. Over time a district can monitor changes in these factors to evaluate the success of policies and programs designed to reduce attainment and growth gaps. [18]

[18] We are currently conducting research and working with school districts on the concept and application of "value-added growth gaps."

The second purpose for including characteristics in a value-added model is to "control" for differences across schools in the student composition of schools so that estimates of school performance α_k reflect differences in school productivity, rather than differences in school composition.[19] In other words, control variables (including prior achievement) are included in the model to achieve, to the extent possible, "apples and apples" school comparisons rather than "apples and oranges comparisons."[20]

Value-Added Model Features to Accommodate Local Conditions

The value-added model presented above can be viewed as a "core" model (in the class of "T2" value-added models) – core in the sense that: (1) it includes most, perhaps all, of the statistical features required for a model (of its type)[21], and (2) the model is based on maintained assumptions with respect to the timing of testing, student mobility, missing data, etc., that allow the model to be as simple as possible. Given our experience working with many districts and states, we have found that it is typically valuable to augment the core value-added model presented in this paper by incorporating model features designed to address local violations of these assumptions. In this section, we briefly discuss two model features that have often been

[19] In T3+ models, it is possible (under some conditions) to indirectly control for differences across schools in the student-level determinants of achievement growth. Given that, as we have suggested, one of the purposes of value added is to measure "growth gaps", we would argue that, even in cases when it is possible to indirectly control for differences in student characteristics, value-added models can be used to measure differences in achievement growth correlated with student characteristics.

[20] In the model considered in this paper we have included student-level measures of student characteristics (X_i), but have not included school-level (or classroom-level) measures of these variables or other school-level variables, for example, the proportion of students in poverty (by school). We discuss the option of including school-variables in a value-added model in Appendix D. See Meyer (1996) and Willms and Raudenbush (1989) for further discussion of this issue.

[21] Core features, as presented in the text, include: control for student-level differences in student characteristics, allowance for the possibility that the coefficient on prior achievement is not one, interpretability of the model given implicit transformation/rescaling of student test scores (relative to a hypothesized latent test score), and control for measurement error in prior achievement.

added to the core model: (1) expansion of the model to allow for student mobility during the school year and (2) generalization of the model to allow for mid-year testing (as opposed to testing at the beginning or end of the school year).

With respect to the first issue, student mobility can be introduced into the core value-added model with only a slight tweak in the definition of the school variables $S_i = [S_{ik}]$ included in the core model. In the core model the school variable S_{ik} is set to one if student *i* attended school *k* during the school year, zero otherwise. In order to accommodate students who changed schools during the school year, S_{ik} is redefined so that it measures the fraction of time that student *i* attended school *k* during the school year. We refer to this variant of the value-added model as the "dose model." Appendix E shows how the dose model can be derived from the core value-added model expressed in the form of a continuous time growth curve model.

There are two alternatives to explicitly measuring "enrollment dose" and neither alternative is very appealing. First, mobile students could be excluded from the value-added measurement system. In general, we strongly believe that it is problematic from a policy perspective to systematically exclude students from a measurement system that serves an evaluation and accountability function. Systematic exclusion of mobile students (or any other student group) from an accountability system creates an incentive for "agents" to allocate fewer resources to this group. Creating an incentive of this type is a bad idea even if we believe (as we do) that few educators would respond to this incentive. The second alternative is to arbitrarily "assign" all mobile students to a single school in a given school year (for example, the school that a student attended on the day they were assessed). The problem with this option is that it could result in substantially biased school effect estimates for schools with modest to high mobility. In short, if there is substantial student mobility in a district (a likely circumstance), then it is best to use a value-added model that measures school enrollment as a dose variable. If the data needed to estimate a value-added dose model is not available in a district, we recommend that the district make it a high priority to begin collecting and warehousing this data.

Second, many states and districts currently administer their assessments during the middle of the school year, rather than at the beginning or end of the school year, as assumed in the core model.[22] In this case, it is necessary to expand the core model to incorporate two growth periods:

- a "spring" growth period, defined by the test date to the end of the school year and
- a "fall" growth period, defined by the beginning of the school year to the test date.

The school enrollment (or dose) variables corresponding to these periods are given by the vectors S_{0i} and S_{1i}, respectively. As discussed in Appendix E, a value-added model that encompasses growth over two part-year periods (defined in terms of true test scores, measured without error) is given by:

$$y_{1i} = \zeta + \lambda y_{0i} + \beta' X_i + \alpha_{0s}{}' S_{0i} + \alpha_{1f}{}' S_{1i} + e_i \,. \tag{14}$$

where the spring and fall school effects are given approximately by:

$$\begin{aligned} \alpha_{0s} &= p\alpha_0 \\ \alpha_{1f} &= (1-p)\alpha_1 \end{aligned} \tag{15}$$

and α_0 and α_1 represent the full-year productivity of schools in grades *0* and *1*, respectively, p = the fraction of the school year covered by the spring period, and $(1-p)$ = the fraction of the school year covered by the fall period. We refer to this model as the "mid-year assessment value-added model."

[22] For example, the Illinois state assessment (ISAT) is administered in March and the Wisconsin state assessment (WKCE) is administered in November.

As above, this model could be simplified by excluding all students who change schools from one school year to the next or from one testing date to the next.[23] In this case, the school enrollment variables would be identical (that is, $S_{0i} = S_{1i} \equiv S_i$) and the model would include only a single (vector) enrollment variable S_i. The coefficient on that variable would absorb both the spring and fall school effects; that is:

$$\alpha_{0s,1f} = \alpha_{0s} + \alpha_{1f} \tag{16}$$

As discussed above, it is generally not a good idea to systematically exclude students from value-added systems.

The preferred strategy is to include all students in the sample and estimate all of the parameters of the mid-year assessment model. It is possible to estimate both the spring and fall school effect estimates if at least some students change schools from one school year to the next. These estimates can be used separately (if they are sufficiently precise) or combined (as in equation (16)) to produce very precise estimates of the "spring plus fall" school effect.

Summary

In this section we have defined a core value-added model within the class of T2 value-added models and illustrated how it can be customized to accommodate local conditions such as mid-year testing. In the next section we use this model as a standard to evaluate other methods for measuring school performance.

[23] This strategy, if applied to transition grades such as 8th to 9th grade, would typically eliminate most, if not all, students from the sample – not a good idea.

III. An Analysis of Alternative Approaches to Measuring Grade-Level School Performance

In this section we discuss and evaluate the four most common alternatives to measuring the performance of schools, teachers, and administrators, as originally proposed by Teacher Incentive Fund grantees: (1) gain, (2) average achievement (an attainment indicator), (3) proficiency rate (an attainment indicator) and (4) gain in proficiency status (movement across proficiency categories) (a gain-type indicator). We evaluate these alternative approaches in terms of the core value-added model presented in the previous section. Although we generally prefer: (1) value-added models based on at least three achievement outcomes for each student ("T3+" models) and (2) models customized to accommodate local conditions (as discussed in the previous section), we selected this model as a point of comparison for two reasons. First, it is simpler to explain much of the logic of the value-added method using this model. Second, models based on two achievement outcomes are somewhat better suited to exploring the strengths and limitations of alternative methods of measuring school performance since these methods also are based on two (or fewer) student achievement outcomes.

Gain and Average Achievement Indicators

The task of evaluating average gain and average achievement as indicators of school performance is much facilitated by the fact that these indicators are "statistically nested" within the value-added model, in the sense that the indicators are obtained by imposing restrictions on the parameters of the value-added model. A shared feature of the three indicators is that they are all based on means of school-level variables. In order to compare the three indicators it is helpful to treat them comparably with respect to centering. Our approach is to center average school-level achievement and gain around the district means of these variables so that the district-level means of these variables equal to zero. By design, the district mean of the value-added effect α_k equals zero. The district and school-level variables used in our analysis are defined as follows:

Variable	**District Average**	**School Average**	**District Centered School Average**
Post Achievement	$\overline{\overline{Y}}_1$	$\overline{Y}_{1k}$	$\overline{Y}_{1k} - \overline{\overline{Y}}_1$
Prior Achievement	$\overline{\overline{Y}}_0$	$\overline{Y}_{0k}$	$\overline{Y}_{0k} - \overline{\overline{Y}}_0$
Gain	$\overline{\overline{Y}}_1 - \overline{\overline{Y}}_0$	$\overline{Y}_{1k} - \overline{Y}_{0k}$	$(\overline{Y}_{1k} - \overline{Y}_{0k}) - (\overline{\overline{Y}}_1 - \overline{\overline{Y}}_0)$
Value-Added Performance	0	α_k	α_k
Student Characteristics	$\overline{\overline{X}}$	$\overline{X}_k$	$\overline{X}_k - \overline{\overline{X}}$

Equation (4) – the core value-added model defined in terms of measured post and prior achievement – provides all of the information needed to link value-added, gain, and average achievement. This equation implies that estimated value-added performance is given by:

$$\begin{aligned}\hat{\alpha}_k &= \overline{Y}_{1k} - (\hat{\zeta} + \hat{\lambda}\overline{Y}_{0k} + \hat{\beta}'\overline{X}_k) \\ &= (\overline{Y}_{1k} - \overline{\overline{Y}}_1) - \hat{\lambda}(\overline{Y}_{0k} - \overline{\overline{Y}}_0) + \hat{\beta}'(\overline{X}_k - \overline{\overline{X}})\end{aligned} \tag{17}$$

where all of the parameters are "hatted" to indicate that they are parameter estimates. This, in turn, yields two key results (written in equation form and in words):

Average Gain	=	$(\lambda - 1)\begin{pmatrix}\text{Average Prior} \\ \text{Achievement}\end{pmatrix}$	+	Average Growth Effect of Student Characteristics	+	Value-Added Productivity

$$\left[(\overline{Y}_{1k} - \overline{Y}_{0k}) - (\overline{\overline{Y}}_1 - \overline{\overline{Y}}_0)\right] = (\hat{\lambda} - 1)(\overline{Y}_{0k} - \overline{\overline{Y}}_0) + \hat{\beta}'(\overline{X}_k - \overline{\overline{X}}) + \hat{\alpha}_k \tag{18}$$

First of all, the gain indicator differs from value-added productivity in two ways. One, it absorbs growth differences across schools, if any, due to differences in student characteristics. Two, differences in average prior achievement across schools "leak" into average gain if the coefficient on prior achievement (λ) does not equal one. Fortunately, these are conditions that can empirically be checked. If the conditions are violated and the magnitude of the violation is substantively large, then the gain indicator will fail to accurately measure school performance.

Average Post Achievement	$=$	$\lambda\begin{pmatrix}\text{Average Prior}\\ \text{Achievement}\end{pmatrix}$	$+$	Average Growth Effect of Student Characteristics	$+$	Value-Added Productivity

$$(\overline{Y}_{1k}-\overline{\overline{Y}}_{1})=\hat{\lambda}(\overline{Y}_{0k}-\overline{\overline{Y}}_{0})+\hat{\beta}'(\overline{X}_{k}-\overline{\overline{X}})+\hat{\alpha}_{k} \tag{19}$$

Similarly, average post achievement differs from value-added productivity in two ways. As in the case of the gain indicator, average post achievement absorbs growth differences across schools, if any, due to differences in student characteristics. Secondly, average post achievement, as expected, absorbs differences across schools in average prior achievement as long as student achievement is a cumulative growth process; that is, as long as $\lambda > 0$.[24] Unless average prior achievement and the average student characteristics factor happen to be the same across schools or perfectly correlated with value-added productivity (all unlikely circumstances), average post achievement and other attainment indicators are likely to be a highly inaccurate measure of school performance. Meyer (1996) presents additional evidence on why attainment indicators generally fail as measures of school performance.

Proficiency Rate

The proficiency rate, the primary indicator used in the NCLB accountability system, is an attainment indicator and thus subject to the same criticisms as in the case of the average achievement indicator. An additional weakness of the proficiency rate, as an indicator of school performance, is that it discards information on student achievement by collapsing achievement into two parts: (1) achievement (Y_{1i}) greater than or equal to the threshold (cut point) between basic and proficient (say, D) and (2) achievement less than this threshold. States and districts

[24] In our experience, estimates of λ typically are much closer to one then they are to zero. They may differ from one for the reasons discussed in the previous section.

typically measure proficiency levels in terms of a very limited number of discrete categories, for example: (1) minimum, (2) basic, (3) proficient, and (4) advanced. As discussed in the next section, replacing "fine-grained" measurement of student achievement with "coarse" discrete levels makes it difficult to accurately measure achievement growth over time.

Gain in Proficiency Status (Movement across Proficiency Categories)

Several TIF grantees measure the productivity of schools, teachers, and administrators with an indicator that measures gain in proficiency status. Although the specific details of this indicator differ across grantees, a simple proficiency gain indicator might be defined as follows: (1) award a school (or other unit) a full point if a student's proficiency status increases from one year to the next and (2) award a half point if a student's proficiency status does not change from one year to the next. The primary advantages of this indicator are that it takes prior achievement into account, it is directly linked to the proficiency outcomes used in NCLB accountability indicators, and it is relatively easy to compute. The problem with this indicator is similar to the problem with proficiency ratings: increases in student achievement only "count" if they are large enough to enable a student to cross a proficiency threshold.

Figures 2 and 3 provide a hypothetical example of how measuring gain in proficiency status could provide misleading information on school productivity. Figure 2 graphs hypothetical data on student post and prior achievement for all schools in a small district at a given grade in a given school year. Vertical and horizontal lines are superimposed on the graph to identify students who scored at one of four proficiency levels (L1, L2, L3, or L4) on each of the two assessments. Students with a gain in proficiency, no change in proficiency, or a decline in proficiency are noted on the graph. As is evident in Figure 2, the proficiency gain indicator (as defined above) has some major weaknesses. First, all students initially at the lowest proficiency level maintain or *increase* their proficiency status. Conversely, all students initially at the highest proficiency level maintain or *decrease* their proficiency status. This is an automatic consequence of the fact that proficiency status is bounded at the top and bottom. Most student assessments are explicitly designed to avoid this problem; few students typically score near the ceiling (maximum) or floor (minimum) of a well-designed assessment. The second major weakness with the proficiency gain indicator is that it is not neutral with respect to a students' initial achievement; students with minimal differences in student achievement may end up in very different proficiency gain categories.

Figure 3 provides more direct evidence on this latter point. The figure reports the proficiency gain rating defined above, but disaggregated by prior achievement level, for a school (or set of schools) with an average value-added rating and with students ranging from the lowest to the highest level of prior achievement. The figure also reports a proficiency gain rating where points are only awarded for increases in proficiency status.[25] As indicated in Figure 3, there is a pronounced saw-tooth pattern to the data, driven by the fact that some students are just short of crossing a proficiency threshold. The key point is that the proficiency gain rating for a given school depends on the exact distribution of students, by prior student achievement, in that school. In addition, both of the hypothetical rating systems depicted in Figure 3 are strongly biased against schools that serve relatively higher achieving students (as evidenced by the sharp downward slope in the gain ratings). This result, as discussed above, is due to the fact that proficiency levels are capped at the floor and ceiling.[26]

---

[25] Note that the aggregate proficiency gain rating (not depicted on the graph) is given simply by the average of the disaggregate proficiency gain ratings (weighted by the number of students at each level of prior achievement).

[26] The specific problem of bias against schools that serve high-achieving students can be at least partly alleviated by awarding more points for students who repeatedly score at the highest level.

Figure 2. Distribution of Prior and Post Achievement and Gain in Proficiency Status

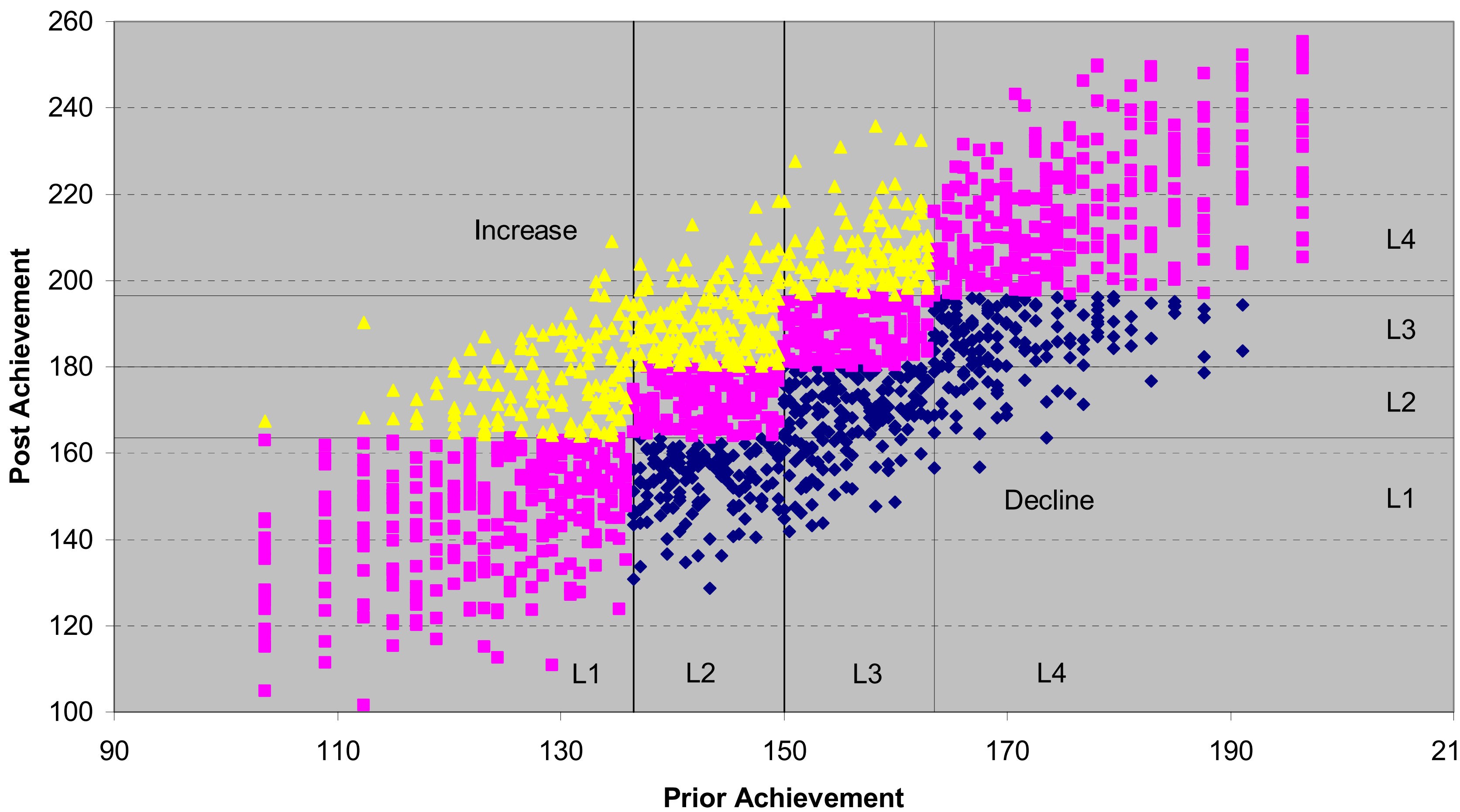
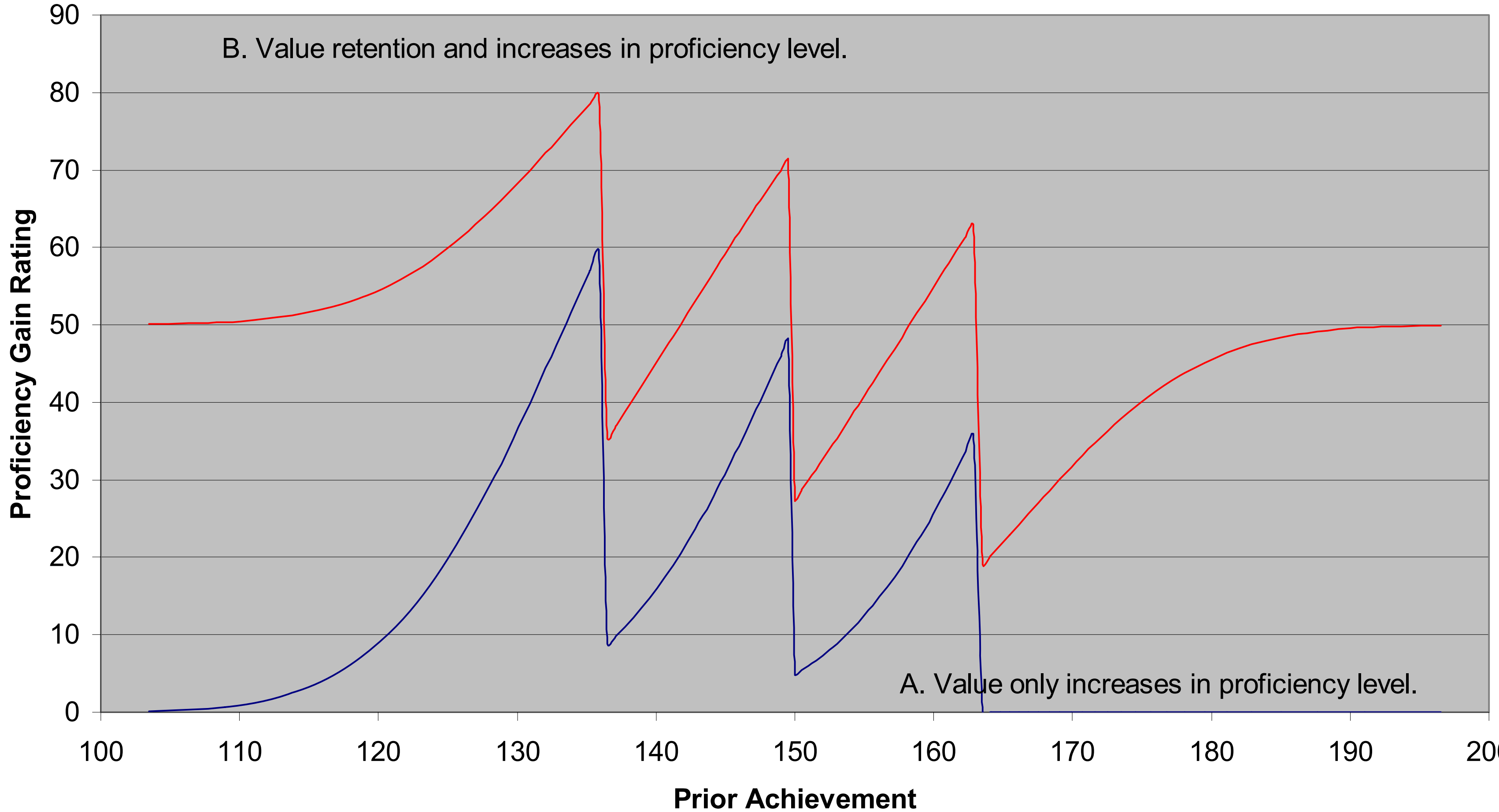


Figure 3. A Hypothetical Proficiency Gain Rating for a School with Average Productivity by Student Prior Achievement



IV. Summary and Conclusions

Teacher Incentive Fund grantees use a wide range of methods to evaluate the performance of schools, teachers, and administrators. A plurality use value-added models of one type or another, but others use simpler approaches, including gain, proficiency gain (movement across proficiency categories), average achievement, and proficiency rates.

In this paper we provided a detailed description of a value-added model in the class of models based on two longitudinal achievement outcomes for each student (posttest and pretest) ("T2" models). We also considered design features that could enhance the ability of the model to accurately measure school performance at a given grade level. We did so for two reasons: to stimulate districts and states to think critically about value-added models and to posit a standard that could be used to evaluate other methods for measuring school performance. Appendix B provides further discussion of more general value-added models based on more than two longitudinal achievement outcomes.

Table 4 summarizes the features of six alternative models/indicators of school performance. As discussed in the paper (and indicated in Table 4), attainment measures such as average achievement or proficiency rates are likely to deviate substantially from value added under realistic conditions. Virtually any measure that takes into account prior student achievement is likely to be strongly preferred to attainment.

The verdict on the simple average gain indicator is mixed. The good news is that this indicator is likely to be similar to a (T2 or T3+) value-added indicator under specific conditions, in particular:

- The assumptions of the core value-added model are reasonable. These assumptions include:

- Students are tested at the beginning or end of the school year so that growth in student achievement does not cut across two school years.
- Students do not change schools during the school year.
- No students are retained in grade or skip a grade from one school year to the next.

- The coefficient on prior achievement (λ) is close to one. We considered four factors that could make it problematic to impose this restriction: durability/decay in achievement, differential resource allocation, differences in the pretest and posttest test scales, and nonlinearity in the test scaling algorithm.
- Differences in student characteristics across schools (see equation (18)) account for little of the variation in average achievement growth.

It is evident that that these conditions could be violated in some realistic conditions. If the violations are substantively large, then the gain indicator may fail to accurately measure school performance.

The proficiency gain indicator is similar in most respects to the average gain indicator. The specific problem with this indicator is that increases in student achievement only "count" if they are large enough to enable a student to cross a proficiency threshold.

The value-added model is highly attractive from a technical perspective. As indicated in Table 4, a value-added model, particularly one that is enhanced and customized to local conditions, is capable of handling numerous real world problems that might otherwise threaten the validity of a simpler model or indicator. In our work developing value-added systems with district and state partners, we have generally followed the model design rule presented in Meyer (2007): "Simpler is better, unless it is wrong." This rule implies that it is desirable, when feasible, to test the validity of restrictive assumptions and relax them in favor of a more complex model if the restrictions are rejected. The value-added model presented in this paper and the related models considered in Appendix B provide an attractive framework for selecting a specific model that accommodates important local conditions and provides valid and reliable estimates of the performance of schools, teachers, and administrators.

In future work we plan to address the feasibility of using value-added, gain, and related indicators in districts that have not traditionally had the resources or technical capacity to build and administer models and indicators of this type. One possible solution to this challenge is for states or consortia of districts to build value-added systems that could serve multiple districts both small and large.

Table 4. Comparison of the Features of Alternative Models/Indicators of School Performance

	Value-Added		Gain		Attainment	
Model Feature	**T2 Model with Extended Features**	**Core T2 Model**	**Average Gain[27]**	**Proficiency Level Gain**	**Average Achievement**	**Proficiency Rate**
Measures average attainment gaps by student group	✓	✓	✓	✓	✓	✓
Measures growth gaps by student group	✓	✓	✓	✓		
Decompose district-wide differences in growth gaps by student group into distinct growth gap factors	✓	✓				
Measures performance in terms of achievement growth	✓	✓	✓	✓		
Responsive to changes in achievement at all points	✓	✓	✓			
Measures average attainment gaps by student group	✓	✓	✓	✓	✓	✓
Measures value-added growth gaps by student group	✓	✓				
Allows for decay in achievement over time or resource allocation correlated with prior achievement	✓	✓				
Robust to transformation of test score (linear approximation)	✓	✓				
Controls for test measurement error, if needed	✓	✓	NA		NA	
Controls for differences across schools in school composition (control at level of within school effect)	✓	✓				
Accommodates local conditions (examples) Dose Mid-year testing	✓					

NA = not applicable

[27] Average gain is defined (and can be computed) only if post and prior achievement are measured on the same test scale.

References

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics, 29*(1), 37-66.

Boardman, A. E., & Murnane, R. J. (1979). Using panel data to improve estimates of the determinants of educational achievement. *Sociology of Education, 52*, 113-121.

Fuller, W. A. (1987). *Measurement Error Models*, New York: John Wiley & Sons.

Goldberger, A.S. (1991). *A Course in Econometrics*, Boston: Harvard University Press.

Hanushek, E. A. (2005). *The market for teacher quality*. Cambridge, Mass: National Bureau of Economic Research.

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*(1), 67-101.

Meyer, R. H. (1992). *Applied versus traditional mathematics: New econometric models of the contribution of high school courses to mathematics proficiency.* (Discussion Paper No. 966-92). University of Wisconsin-Madison: Institute for Research on Poverty.

Meyer, R. H. (1996). Value-added indicators of school performance. In E. A. Hanushek & D. W. Jorgenson (Eds.), *Improving the performance of America's schools* (pp. 197-223). Washington, D.C.: National Academy Press.

Meyer, R. H. (1997). Value-added indicators of school performance: A primer. *Economics of Education Review*, *16*(3), 283-301.

Meyer, R. H. (1999). The production of mathematics skills in high school: What works? In S. Mayer & P. Peterson (Eds.), *Earning and learning: How schools matter* (pp. 169-204). Washington, D.C.: Brookings Institution.

Meyer, R. H. (2001). *An evaluation of the effectiveness of the urban systemic initiative and other academic reforms in Texas: Statistical models for analyzing large-scale datasets.* (Report prepared for the National Science Foundation). University of Wisconsin-Madison: Wisconsin Center for Education Research.

Meyer, R. H. (2006). *Student and school performance targets: a framework for unifying the objectives of NCLB and the logic of value-added.* Report prepared for conference with the Institute for Research on Poverty. University of Wisconsin-Madison: Wisconsin Center for Education Research

Meyer, R.H. (2007). *Using Value-Added Measures of Student Achievement Growth in Teacher Compensation: Pitfalls and Promises*. Report prepared for the First Annual Meeting of the Teacher Incentive Fund Project, Center for Education Compensation Reform.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: applications and data analysis methods (2nd ed.).* Thousand Oaks, CA: Sage.

Sanders, W. L., & Horn, S. P. (1994). The Tennessee value-added assessment system (TVAAS): Mixed model methodology in educational assessment. *Journal of Personnel Evaluation in Education*, *8*, 299-311.

Searle, S. R., Casella, G., & McCulloch C. E. (1992). Variance components. John Wiley & Sons, New York.

Todd, E. T., & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal,* 113, F3–F33

Willms, D. J., & Raudenbush, S. W. (1989). A longitudinal hierarchical linear model for estimating school effects and their stability. *Journal of Educational Measurement, 26*(3), 209-232.

Appendix A

Approaches to Measuring the Performance of Schools, Teachers, and Administrators in Teacher Incentive Fund Proposals

The table below lists the approach to measuring the performance of schools, teachers, and administrators specified in each Teacher Incentive Fund proposal. We classified the different approaches into six different categories (listed in italics for each project):

- Value-added
- Gain
- Movement across proficiency categories
- Proficiency rates or attainment
- Gain/movement/proficiency hybrid
- Individual learning plans

The page numbers in the table reference the appropriate pages in the grantees' proposals, which were the basis for this survey. When information outside the proposals was used, the source of the information is footnoted.

Round I Grantees	Performance Measurement Approach
Alaska	*Movement across proficiency categories*; meeting individual learning plans (p. 18)
Chicago	*Value added* (pp. 7-8)
Dallas	*Value added* (pp. 9-10)
Eagle County	*Value added* (pp. 6-8).
Guilford County	*Value added* (p. 14).
Houston	*Value added* and comparable improvement, a measure of adjusted gain.[28] Proposal did not include value added (pp. 15-20).
Mare Island	*Proficiency rates and attainment* (pp. 4-7).
NLNS Charter Schools	*Value added* (pp. 12-13).
NLNS D.C.	*Value added* (p. 12).
NLNS Memphis	*Value added* (pp. 12-13).
Northern New Mexico	*Gain* (p. 10).
Ohio	Proposal expands TAP, which uses *value added*, in Cincinnati and Columbus and TRACS in Toledo, and starts CTIS in Cleveland (pp. 10-15). Cincinnati uses value added.[29] Cleveland will (p. 15).
Philadelphia	*Value added* (pp. 10-12)
Denver	Proposal extends ProComp (p.1), which requires teachers to meet objectives based that may be based on *proficiency or gain*.[30]
South Carolina	*Value added* (pp. 9, 19)
Weld County	Payouts have been on accreditation indicators, which focus on *proficiency rates and attainment*.[31] Proposal references proficiency rates and gain (pp. 14-15).

[28] Houston Independent School District, *Houston Independent School District ASPIRE Awards Model: 2006-07 School Year*, at <http://www.houstonisd.org/ResearchAccountability/Home/Teacher%20Performance%20Pay/Teacher%20Performance%20Pay/Board%20Items/ASPIRE_AWARDS_OVERVIEW.pdf >; Houston Independent School District, *Methods for the ASPIRE Award Model for 2006-07*, at <http://www.houstonisd.org/ResearchAccountability/Home/Teacher%20Performance%20Pay/Teacher%20Performance%20Pay/Board%20Items/ASPIRE_Award_Model_methods.pdf>; Texas Education Agency, *2006 Accountability Manual*, at <http://www.tea.state.tx.us/perfreport/account/2006/manual/index.html>

[29] "Cincy and Toledo advance with five-year federal grant," *Catalyst Ohio* 9(1), Sept./Oct. 2007, at <http://www.catalyst-cleveland.org/assets/20070901/files/sepoct07.pdf>

[30] Denver Public Schools, *Checklist for Student Growth Objectives*, *Elementary Classroom Sample Objectives*, *Middle School Sample Objectives*, and *High School Sample Objectives*, at <http://my.dpsk12.org/objectives/>

[31] Weld County School District*, Teacher Incentive Fund Project: The Basics*, at <http://www.ftlupton.k12.co.us/information/default.asp?NavPageID=66078>; Colorado Department of Education, *School and District Accreditation*, at < http://www.cde.state.co.us/index_accredit.htm>.

Round II Grantees	Performance Measurement Approach
Amphitheater	*Value added* or gain.[32] Policies still development as of the proposal (p. 32).
Beggs	*Proficiency rates and attainment* (pp. 21-27).
Prince George's County	*Proficiency rates* (pp. 10-12).
CEI - New York	Meeting *individual achievement plans* for students set by predictive models (pp. 8-12, 15-18).
Charlotte-Mecklenburg	*Proficiency rates* and High Growth (p. 10), an adjusted *gain* measure.[33]
Cumberland County	*Proficiency rates* and High Growth (pp. 8-13, 23-25), an adjusted *gain* measure.[34]
Brooke Charter School	*Proficiency rates and attainment, gains* (pp. 7-9)
Florence County	*Value added* (pp. 11-13).
Harrison	*Movement across proficiency categories* (pp. 6-9).
Hillsborough	*Movement across proficiency categories* (pp. 6-7). Incentives are given to teachers deemed effective using the STAR project's value tables.[35]
Lynwood	*Proficiency rates and attainment*; cross-sectional improvement in consecutive grades (pp. 9-12).
Lake County	*Value added* (pp. 7-9, 28).
Miami	Proposal awards principals using scorecards (p. 11). Scorecards for existing administrator incentive program focus on state accountability measures based on *proficiency, movement across proficiency categories*, and percent making *gain*.[36]
Orange County	*Gain* (pp. 9-10).
Pittsburgh	*Value added* (pp. 19-20 & 26-27).
School of Excellence in Education	*Value added* (pp. 23-25, 28), gain (pp. 17-18).
South Dakota	*Gain, attainment* (pp. 10-11).[37]
University of Texas	*Value added* (pp. 15-18).

[32] Amphitheater Unified School District, *Project Excell! Implementation Manual Phase 1 Design*, at <http://www.amphi.com/departments/teachlearn/projexcell/files/77EA6895EDD94187B0F2609608752002.pdf>
[33] Public Schools of North Carolina, *School-Based Management and Accountability Procedures Manual*, at <http://www.ncpublicschools.org/docs/schoolimprovement/2004abcmanual.pdf>
[34] Ibid.
[35] Florida Department of Education, *Special Teachers are Rewarded*, at <http://www.fldoe.org/PerformancePay/>
[36] Miami-Dade County Public Schools, *MEP Prototypes*, at <http://asp.dadeschools.net/Products/MEP_Prototypes/>
[37] South Dakota Department of Education, *South Dakota Incentives Plus*, at <http://www.doe.sd.gov/oatq/incentives_plus/index.asp>

Appendix C

Value-Added Models with Multiple Growth Years

In the context where there are multiple years of data, a value-added model could be structured in at least two different ways: (1) the coefficients on prior achievement and student characteristics could be restricted to be the same from year to year or (2) these coefficients could be allowed to vary across years. In the text we discuss how changes in the coefficients associated with student characteristics could be used to evaluate the success of policies and programs designed to reduce attainment and growth gaps. Below we discuss some aspects of the first model.

A Value-Added Model with Constant Coefficients on Prior Achievement and Student Characteristics over Time

A value-added model with multiple year data could include separate year effects ξ_{gt} for each year *t* (for a given grade level *g*) and the year effect for the benchmark year (typically the first growth year in the data series) would be normalized at zero. In this model specification the year effects ξ_{gt} would capture changes in the overall value-added productivity of the set of schools included in the analysis (typically a district, state, or consortium of schools). The school effects α_{gt} would be centered around zero for each grade and year and would capture the *relative* productivity of each school relative to the average school for that year (in a given grade). The two productivity indicators can be added together to produce an *absolute* productivity indicator $\alpha_{gt}^{ABSOLUTE} = \xi_{gt} + \alpha_{gt}$ that measures the productivity of each school relative to the average school in the benchmark year. The absolute value-added indicator is typically preferred because it allows schools to monitor whether they are improving over time (not relative to changes in the productivity of other schools over time). Of course, this indicator is valid only if test scores from different years are scored on the same scale; that is, there are no form effects.

Appendix D

Student and School-Level Variables in Value-Added Models

Meyer (1996) and Willms and Raudenbush (1989) discuss some of the conceptual and empirical issues involved in including student and school-level control variables in value-added models, for example, average poverty status. The primary concern with including school-level control variables is that the estimated coefficients on these variables could be substantially biased if school resources and "intrinsic" school productivity are not "assigned" to schools such that school-level control variables are uncorrelated with unobserved school productivity. This condition would be violated, for example, if high-performance teachers and administrators prefer to work in schools with low poverty. In this case, the coefficient on average poverty status would absorb this negative correlation, yielding an estimated coefficient biased in the negative direction.

Estimated value-added indicators would, of course, be biased if the coefficients on school-level control variables are biased. For example, if the coefficient on average poverty status is biased in the negative direction, the estimated value-added performance of schools that disproportionately serve high poverty students would be biased upward. This is problematic if one of the important purposes of a value-added system is to identify schools that are in need of improvement. Due to this concern, model designers need to be cautious about including school-level control variables in value-added models. This is an important area for further research. Note that the statistical concerns discussed above do not apply to student-level control variables, as represented by X_i in the text of the paper. These coefficients are estimated off of variation within schools or classrooms, for example, the contrast in achievement growth between low and high-poverty students within schools or classrooms. If resources are allocated within schools or classrooms in a way that is systematically related to student characteristics, then the coefficients on student-level control variables will capture these systematic patterns. The other role for student-level control variables is to proxy for the differences in resources provided to students by their families.[38]

Note: Appendices B and E are not included in this draft of the paper.

[38] We are currently working with a district to develop a differential effects value-added model in order to detect whether resources are allocated within schools or classrooms in a way that is systematically related to student characteristics.

Faculty and Research Affiliates

Matthew G. Springer
Director
National Center on Performance Incentives

Assistant Professor of Public Policy and Education
Vanderbilt University's Peabody College

Dale Ballou
Associate Professor of Public Policy and Education
Vanderbilt University's Peabody College

Leonard Bradley
Lecturer in Education
Vanderbilt University's Peabody College

Timothy C. Caboni
Associate Dean for Professional Education and External Relations
Associate Professor of the Practice in Public Policy and Higher Education
Vanderbilt University's Peabody College

Mark Ehlert
Research Assistant Professor
University of Missouri – Columbia

Bonnie Ghosh-Dastidar
Statistician
The RAND Corporation

Timothy J. Gronberg
Professor of Economics
Texas A&M University

James W. Guthrie
Senior Fellow
George W. Bush Institute

Professor
Southern Methodist University

Laura Hamilton
Senior Behavioral Scientist
RAND Corporation

Janet S. Hansen
Vice President and Director of Education Studies
Committee for Economic Development

Chris Hulleman
Assistant Professor
James Madison University

Brian A. Jacob
Walter H. Annenberg Professor of Education Policy
Gerald R. Ford School of Public Policy University of Michigan

Dennis W. Jansen
Professor of Economics
Texas A&M University

Cory Koedel
Assistant Professor of Economics
University of Missouri-Columbia

Vi-Nhuan Le
Behavioral Scientist
RAND Corporation

Jessica L. Lewis
Research Associate
National Center on Performance Incentives

J.R. Lockwood
Senior Statistician
RAND Corporation

Daniel F. McCaffrey
Senior Statistician
PNC Chair in Policy Analysis
RAND Corporation

Patrick J. McEwan
Associate Professor of Economics
Whitehead Associate Professor of Critical Thought
Wellesley College

Shawn Ni
Professor of Economics and Adjunct Professor of Statistics
University of Missouri-Columbia

Michael J. Podgursky
Professor of Economics
University of Missouri-Columbia

Brian M. Stecher
Senior Social Scientist
RAND Corporation

Lori L. Taylor
Associate Professor
Texas A&M University

NATIONAL CENTER ON
Performance Incentives

EXAMINING PERFORMANCE INCENTIVES
IN EDUCATION

National Center on Performance Incentives
Vanderbilt University Peabody College

Peabody #43
230 Appleton Place
Nashville, TN 37203

(615) 322-5538
www.performanceincentives.org

VANDERBILT
PEABODY COLLEGE