

Running head: VAMs and REACH

Value-Added Models and Adequate Yearly Progress: Combining Growth and Adequacy in a
Standards-based Environment

Harold C. Doran

Council of Chief State School Officers

Paper Presented at the Annual CCSSO Large-Scale Assessment Conference

Boston, Massachusetts

Please do not cite or distribute without permission from author

June 8, 2004

Harold C. Doran

Council of Chief State School Officers

One Massachusetts Ave., NW, Suite 700

Washington, DC 20001-1431

202.336.7075

haroldd@ccsso.org

Value-Added Models and Adequate Yearly Progress: Combining Growth and Adequacy in a
Standards-based Environment

Abstract

Value-added models have received increased attention in light of current test-based accountability legislation. However, prior attempts to use growth models have allowed for students to grow without ever reaching an acceptable standard of performance. This paper presents a value-added model that can be fit given many of the data elements maintained by a school district or state education agency and a new AYP model, referred to as the Rate of Expected Academic Change, or REACH Score. Rather than allowing for students to grow without reaching an acceptable standard of performance, this method requires that all students, not successive student groups, reach the proficient standard within a defined period of time. Therefore, REACH is consistent with the design principles of the No Child Left Behind but suggests that individual students serve as the unit of analysis with their growth trajectories used to measure progress towards a standard.

Keywords: accountability; adequate yearly progress; value-added model

Value-Added Models and Adequate Yearly Progress: Combining Growth and Adequacy in a Standards-based Environment

Introduction

On March 24, 2004, 16 Chief State School Officers collectively sent the U.S. Secretary of Education a letter presenting a solution to the methodological problems of the No Child Left Behind Act (NCLB) of 2001 (O'Connell et al, 2004). In their letter, the chiefs suggest that the status-based method for measuring student and school progress required by NCLB is a central problem to the Act and that more appropriate methods, such as a growth model, should be considered for its improvement.

Status-based accountability models, the most common methods used in accountability systems, compare successive cohorts of students using aggregate level scores. The problems and limitations of status-based accountability designs are well known (Doran & Izumi, 2004; Meyer, 1997). Because students are not randomly assigned to schools or classrooms, preexisting differences exist across school populations to begin with. As such, comparing performance differences across schools or classrooms using status based models is likely to illustrate differences due to initial status or other irrelevant student characteristics, not instructional quality. Clearly, knowing where a student *is* tells you nothing unless you know where that student *was*. Therefore, it should not be controversial that the most logical method for measuring the extent to which students are learning is through longitudinal design.

Consequently, value-added models (VAM) have received increased attention as researchers, policymakers, and practitioners explore new methods for comparing schools or teachers (Lockwood, Doran, & McCaffrey, 2003; Sanders, Saxton, & Horn, 1997; Thum, 2003). VAMs, unlike cross-sectional methods of analysis, consider individual students as the unit of analysis and estimate gains in student learning. As such, they more reasonably align

with the notion of learning and are more likely to reflect the effectiveness of schools or teachers than cross-sectional methods of analysis.

However, knowing whether VAMs can provide educational benefit requires an understanding of the models, what they are estimating, and how they can be used to judge the adequacy of individual student growth, or effectiveness of a school. Therefore, this paper presents a general value-added model that can be fit given many of the common data elements maintained by a local education agency or state department of education (Doran & Lockwood, 2004). In this VAM, the primary object of inference is the individual student and his rate of change towards an outcome of *value*. This rate of change serves as the basis for a new model for defining Adequate Yearly Progress (AYP) referred to as the Rate of Expected Academic Change, or REACH. The REACH model is consistent with the principles of NCLB—that all students reach proficiency within a specific timeline—but considers individual students as the unit of analysis and the timeline is structured to allow for true longitudinal design.

In the discussion that follows, I briefly review the problems associated with methods of measuring growth based on the equipercentile assumption that have been used in Title I research (Davis, 1991). Subsequently, I suggest that all inferences regarding student achievement should be made with respect to standards, not other students. Therefore, the modeling approach presented places less emphasis on causal inferences and greater emphasis on the descriptive information regarding student growth rates extracted from the statistical analysis.

Adequate Yearly Progress and The Equipercentile Assumption

Judging the adequacy of growth rates has been the source of great discussion including models based on aggregate scores (Schwarz, Yen, & Schafer, 2001), coefficients from linear models (Raudenbush & Willms, 1995; Meyer, 1997), and recommendations for including a more stringent alpha level for judging the rate of improvement (Hill & DePascale, 2003).

A method commonly used in Title I research relies on the equipercentile assumption (Davis, 1991), otherwise referred to as Normal Educational Growth. In this model a student must maintain or exceed the same position in the distribution over time. This gain is justified on the basis that a student is taking a more difficult test with more difficult content.

It was common to compute this gain using adjacent Normal Curve Equivalent (NCE) Scores (Davis, 1991; Thum, 2003). For example, the NCE pretest score for student i at time t in subject p was subtracted from this student's posttest score at time t to compute a raw gain as:

$$G_{ip} = NCE_{tip} - NCE_{t-1,ip} \quad (1)$$

Because the NCE score scale is not continuous, a gain of "0" was considered the expected gain, or "one year's growth", not "no growth." The problems with this approach are numerous. Among the many statistical inefficiencies, G_{ip} can only be obtained when students have adjacent scores, two waves of data is statistically inefficient (cite Rogosa here), and the simple method of subtracting scores completely confounds measurement error and any other systematic or random error with the gain attributed to the school.

Last, and in my view the substantive problem, is the fact that a student is not expected to reach an acceptable standard of performance. For example, a student need only to perform as badly as he did previously to be considered adequate. Consider, for example, a student who had an initial score of 10 and maintained this same position the subsequent year. This gain would be considered adequate even though he is still performing at the low end of the score distribution. It is difficult to conceive this growth as progress toward any valuable standard of academic performance.

Consider the empirical cumulative distribution functions (CDF) displayed in the left panel of Figure (1). They represent the Grade 4 and 5 scales on the Stanford 9 Math test (Harcourt Brace, 1997). The right panel of the figure displays the horizontal gap measured between the

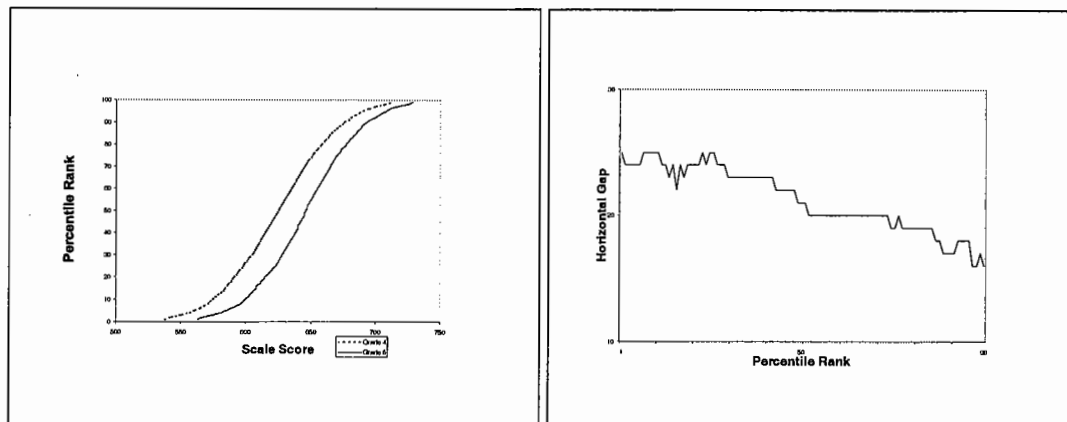


Figure 1: Stanford 9 Grades 4 and 5 Math Empirical Cumulative Distribution Functions and Horizontal Gap

two CDFs at each fixed percentile rank (Holland, 2002). This gap is an indicator of the amount of growth, in scale score units, a student must make in order to maintain the same position in the distribution. It is evident from the right panel that growth varies as a function of initial status. That is, the amount of growth required to maintain the same position in the distribution at percentile rank 25 is not the same as that at 75. If we assume an interval scale (an assumption that is likely to be untenable), then it appears that making normal growth from Grade 4 to 5 is easier for students at the higher end of the distribution than it would be for others.

This method has both statistical inefficiencies as well as a problematic definition of adequacy. Primarily, students may grow, but never reach an acceptable standard of academic performance.

A General Value-Added Model

A mixed-effects linear model forms the basis for the statistical model presented in the remainder of this paper. The model takes the following general form $Y_i = X_i\beta + Z_i\Theta_i + \epsilon_i$ (Pinheiro & Bates, 2000), where Y_i is an n_i -dimensional response vector of test score data, X_i

is an $n_i \times p$ design matrix, β is the p -dimensional vector of fixed effects, Z_i is the $n_i \times q$ design matrix for the q -dimensional vector of Θ_i random effects, and ϵ_i is the n_i -dimensional within-group error term. The fixed effects, $X_i\beta$, are referred to as the structural portion of the model and the random effects, $Z_i\Theta_i + \epsilon_i$, as the stochastic portion of the model.

These models are collectively known as hierarchical linear models, mixed models, or multilevel models and extend Ordinary Least Squares (OLS) techniques by accounting for the non-zero intraclass correlation likely to exist when students share similar instructional settings (e.g., classrooms, schools). In addition, the method captures the entire vector of student observations regardless of missing observations and “shrinks” parameter estimates to adjust for unreliability. Given the advantages afforded, mixed linear models provide an adequate platform for estimating the growth rates of students nested (or crossed) within teachers and schools.

The general structure of the data considered are repeated observations nested within students nested within schools. Models considering teacher effects are currently ignored as most districts or state departments of education do not maintain the unique teacher identification numbers that would be required to link student scores to teachers over time.

Fitting the model proposed requires at least two test scores provided for individual students, unique student and school identification numbers that link scores to individual students over time, and a continuous, developmental scale such that $(Y_{tij} - Y_{t-1,ij})$ represents increased academic achievement. Although not a mathematical assumption, the model also requires that the test score data are well aligned to curricular goals and objectives such that the output produces information that is actionable by practitioners within the schools.

Model Specification

The first objective is to formulate and fit a statistical growth model that accurately reflects the observed data and supports the types of inferences desired. Because it is reasonable to

assume that students and schools start in different locations on the ability scale and grow at unique rates, the model includes random effects for the intercept and slope at each level of the hierarchy. Therefore, the following general specification is formulated to estimate the school and student effects for purely nested designs:

$$Y_{ti} = [\beta_0 + \beta_1(\text{time})] + [\theta_{0j} + \theta_{1j}(\text{time}) + \delta_{0j(i)} + \delta_{1j(i)}(\text{time}) + \epsilon_{ti}] \quad (2)$$

where i indexes student, j indexes school, t indexes time, and the notation $j(i)$ is used to indicate that student i is nested within school j . The error terms are assumed $\epsilon \sim N(0, \sigma^2)$ with a general positive-definite covariance matrix for the school and student random effects, respectively:

$$(\Psi) = \text{Var} \begin{bmatrix} \theta_{0j} \\ \theta_{1j} \end{bmatrix} = \begin{bmatrix} \psi_{00} & \psi_{01} \\ \psi_{10} & \psi_{11} \end{bmatrix} \quad (3)$$

$$(\Omega) = \text{Var} \begin{bmatrix} \delta_{0j(i)} \\ \delta_{1j(i)} \end{bmatrix} = \begin{bmatrix} \omega_{00} & \omega_{01} \\ \omega_{10} & \omega_{11} \end{bmatrix} \quad (4)$$

The fixed portion of the model includes the grand mean (β_0) and the main effect for *time* (β_1) separated from the entire stochastic portion using brackets.

Equation (2) is the unconditional multivariate latent growth model that can be extended in numerous ways contingent upon the desired inferences and the available data. For example, Y_{ti} can be reduced to $D_i = Y_2 - Y_1$ when two data points are provided¹, can be reformulated as doubly-multivariate when more than one response variable is provided, include conditional standard errors of measurement, include additional covariates, such as gender and ethnicity, and even be extended to account for crossed random effects at the school or teacher levels. The interested reader may consult Doran and Lockwood (2004) and Lockwood, Doran, and

¹In this univariate model, the researcher would also remove the slope parameter.

McCaffrey (2003) for a full description of this model, its extensions, and the code necessary for fitting it in the **nlme** library, a free statistical software package.

For example, extending this model to account for student and school characteristics and students migrating across k schools would take the following general form:

$$Y_{ti} = [\beta_0 + \beta_1(\text{time}) + \lambda'_j + \phi'_{ij}] + \left[\sum_{t=0}^T \sum_{j=1}^K \eta_{tij} \theta_{0j} + \theta_{1j}(\text{time}) + \delta_{0j(i)} + \delta_{1j(i)}(\text{time}) + \epsilon_{ti} \right] \quad (5)$$

where λ'_j is a vector of school characteristics and ϕ'_{ij} is a vector of student characteristics, which, for illustration are treated as fixed effects. η is a measure of the proportion of the school year student i attended school j during year t . If student i was in school j for the entire school year, then $\eta=1$. If student i was not in school j for any part of the year, then $\eta=0$. Otherwise, η falls between 0 and 1. The double summation permits for the effect of school j to persist undiminished into the future consistent with the model proposed by Raudenbush and Bryk (2002). However, the actual persistence of school (or teacher) effects may not behave quite so. Consequently, a “decay” parameter as described by McCaffrey et al (2004) may be appropriate.

Equation (5) presents an interesting modeling strategy, even for students properly nested within a single school. In this formulation, θ_{0jt} is the effect of school j at time t and $\theta_{0j,t+1}$ is the effect of school j one year later. If this model is used for students that remain in the same school, then θ_{0jt} describes how the school has a differential impact on student achievement as a student progresses in age. For example, certain grade levels may have a larger impact on student learning gains than others. However, if all students are properly nested within one higher level unit and the school effects are constant across time, then Equation (5) would reduce to Equation (2).

The Traditional VAM and Causal Inference

Traditionally, the parameter of interest in a value-added study is θ_{1j} , the “effect” of school j . That is, the effect of attending school j is to adjust the growth rate from β_1 to $(\beta_1 + \theta_{1j})$. As

such, θ_{1j} is a measure of relative instructional quality. In other words, judgment regarding the performance of school j is made with respect to other schools included in the analysis.

This definition seems problematic for at least two reasons. First, a researcher draws causal inferences regarding the quality of a school by comparing school j to a “typical” school (i.e., $\theta_{1j}=0$). A student attending a school with a deviation less than β_1 (i.e., $\theta_{1j} < 0$) is perceived to be less fortunate than a student attending a school with a positive random effect (i.e., $\theta_{1j} > 0$). In other words, one is inextricably lead to believe that if student i had attended school j rather than a typical school (or the school he actually attended), then his growth rate would have been equivalent to that observed for school j .

Under randomized conditions, with student ability equally distributed across all schools and classrooms, it may be a reasonable assessment to judge the gain as the causal effect of attending School A as opposed to School B. However, the impact of selection bias precludes a researcher from making this causal association. In the absence of knowing whether good teachers make good students or good students make good teachers, these causal claims are unwarranted.

However, comparing the varying growth rates of students or schools and making normative comparison does not provide the type of diagnostic information that is useful in a standards-based environment. More importantly, β_1 is sensitive to which schools are included in the analysis. For example, if a second school district were added to a large-scale analysis β_1 will shift. If it were to shift downwards, a school perceived to be adding less value than expected could end up appearing to be a high performing school.

The REACH Score: A Student Level Contextual Model

Rather than seeking causal inferences based on relative instructional quality, it seems more relevant and meaningful to extract the descriptive information regarding individual student growth rates and consider their progress towards a defined standard of academic

performance. Rubin, Stuart, and Zanutto (2004) note :

We advocate a position of taking the current value-added models at face-value and considering their parameter estimates as descriptive measures that we hope are of some relevance to the question of educational assessment. The real question then is, do these descriptive measures, or proposed reward systems based upon them, improve education? (p. 111)

However, it is not the measures themselves that improve education, but the actions educators and/or the public take as a result of data. This suggests that practitioners should be provided with meaningful and relevant information that can be used to improve the quality of instructional delivery. Second, the accountability system must also provide accurate and reliable information reflecting the quality of the educational program. In other words, the data should serve an internal function to support appropriate *classroom action*, i.e., instructional consequences. At the same time the data should serve an external function to support appropriate *public action*, or social consequences that originate external from the school (Doran, 2003).

Consequently, the value-added results should be placed within a context that is educationally and socially meaningful. Additionally, results from the VAM should also provide the basis for instructional consequences such that educational practitioners can judge whether students are making adequate progress towards an outcome of *value*?

Rather than seeking a causal inference from the random effects, or making a normative judgment on growth based on initial status, all inferences regarding student growth should be made within a standards-based context. However, determining whether growth is adequate is highly value-laden and must fit within an espoused system of educational and social values. The model presented here assumes that the statistical model provides an accurate

representation of student i 's growth rate that can be used to measure the progress made towards an acceptable standard of academic performance.

Given the parameterization of the growth model specified in Equation (2), the estimated true growth rate ($ETGR$) student i in school j is the sum of the main effect for $time$ and the school and student random effects:

$$ETGR_i = \beta_1 + \theta_{1j} + \delta_{1j(i)} \quad (6)$$

Equation (6) estimates a gain in the scale score metric, which is often an arbitrary scale designed to operationalize a latent trait. Consequently, scale score gains are difficult to interpret. Therefore, the second objective of the VAM is to place each student's observed growth rate in context, providing practitioners and policymakers with a tool to judge the "adequacy" of the gain estimated from the statistical growth model. One particular way of achieving this goal, with a slight modification to the way the NCLB timeline is structured, is by determining a Rate of Expected Academic Change, or REACH Score, for each student. In particular we ask, "given this student's current location on the ability scale, how much does she need to grow each year in order to be proficient by the time she leaves this school?" This question is operationalized in the following manner:

$$REACH_i = \frac{\lambda_k - Y_{ti}}{T - \alpha_i} \quad (7)$$

where λ represents the lower bound cutscore for proficiency on test k at the highest grade in a school, T is the highest grade in the school (or end of a specified timeline), and α_i represents the current grade level of student i .² The REACH model requires that performance categories such as "Proficient" or "Advanced" are defined as ranges along the scale score continuum. For

²The end of the timeline can be easily modified. For example, the State of Colorado has recently modified the REACH timeline to be Grade 10, rather than the highest grade in the school.

this model to be realistic, it is assumed that the proficient cutpoint is an ambitious, but attainable for all students.

The student's estimated true growth rate from Equation (6) is then compared to the value produced in Equation (7). Thus, the extent to which a student is making "adequate" progress is judged by the following ratio:

$$REACH\ Ratio = \frac{ETGR_i}{REACH_i} \quad (8)$$

From this perspective, a REACH Ratio of "1" or greater indicates that, given the students historical record of growth, this student is likely to be proficient by the time she leaves the highest grade in the school. A REACH Ratio less than "1" indicates that, unless instruction is modified for this student, she is unlikely to reach the Proficient cutpoint by the time she leaves the highest grade in the school.

The left panel of Figure (2) illustrates the idea of a REACH Ratio less than "1" and the right panel illustrates the idea of a REACH Ratio greater than "1". In the figure below, the dark line is the estimated true growth rate using the data points obtained from grades 1 to 3 and the dashed line is the extrapolated growth trajectory. The line extended from the current score to the lowerbound cutscore for proficiency is the REACH Score, or the rate at which this student must grow to reach the proficient cutpoint by Grade 5.

While the ultimate goal is for each student to reach the Proficient standard by a specified timeline (e.g., the highest grade in a school), the REACH Ratio can be used as an intermediate goal, judging the adequacy of each student's yearly progress.

An Example of the REACH Method

Table 1 illustrates four hypothetical students in an elementary school where Grade 5 is the highest grade in the school and 500 represents the lower bound cutscore for Proficiency.

The data suggest that both Students A and B are likely to reach the proficient cutpoint

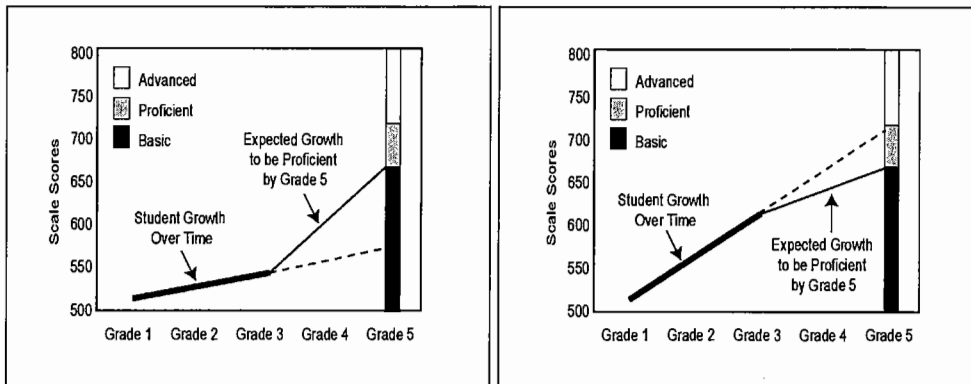


Figure 2: Sample REACH Ratios

Table 1: Sample REACH Ratio Computation

Student	y	ETGR	α	λ	$(\lambda - y)$	REACH	REACH Ratio
A	400	60	3	500	100	50	1.2
B	460	45	4	500	40	40	1.1
C	440	25	4	500	60	60	.42
D	510	40	5	500	Proficient	*	*

unless their growth trajectories decline. Student C, however, is unlikely to be Proficient given his rate of change unless instruction is modified for this individual. This illustrates how the REACH Ratio can serve a diagnostic purpose for instructional planning.

Note that a REACH Ratio is not produced for students already in the highest grade, Grade 5 (or for students in grades 1 to 4 that have reached the proficient cutpoint). Because this student has no more time left to grow, she either reached the standard or not.

The School Model

The REACH method is the contextual model used to judge the adequacy of a student's rate of change. At the school level, a slightly different method for evaluating performance is posed. At this higher level of aggregation the percentage of students at or above the Proficient cutpoint (PAC) is computed across all tested grade levels in the school and use the estimated growth rate of the school computed from Equation (9) below.

$$ETGR_j = \beta_1 + \theta_{1j} \quad (9)$$

Combining these data, we can organize all schools on a two-dimensional coordinate plane $(ETGR_j, PAC_j)$. As such, we can compare the performance of all schools in the following multidimensional manner:

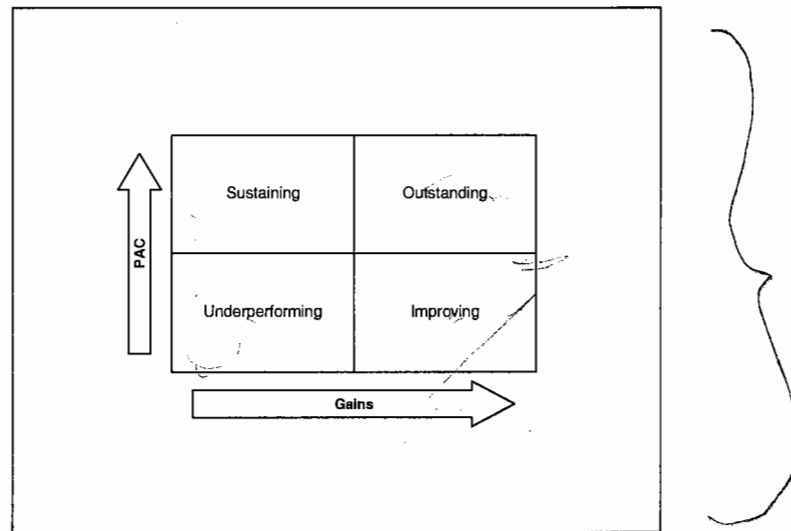


Figure 3: Multidimensional perspective of school accountability

where the y-axis represents the percentage of students within the school at or above proficient and the gains are presented along the x-axis as determined by Equation (9). The horizontal line within the matrix corresponds to the fixed effect for time (β_1) and the vertical line corresponds to the percentage of students at or above a cutpoint (PAC), such as the proficient cutpoint. In the matrix below, a school at the origin would have 50 percent of its students at or above PAC and a growth rate equal to β_1 .

When data are organized as such, schools in the *Underperforming* quadrant have few students at or above proficient and are making low year-to-year gains. Schools in the *Improving* quadrant also have a low percentage of students at or above proficient, but are making substantial yearly gains. Schools in the *Sustaining* quadrant have a high percentage of students

at or above proficient, but are making low gains. Last, schools in the *Outstanding* quadrant have a high percentage of students at or above proficient and are making strong yearly gains.

Organizing schools into the four quadrants allows for policymakers to differentiate between two schools that have the same percentage of students at or above proficient, but are making different gains. For example, Schools A and B may both have 45 percent of their students at Proficient, but School B may be making larger yearly gains with its students, indicating it is a more effective school than School A.

Recommendations for Policymakers and Practitioners

- If test-based accountability systems are to be part of legislation, then estimate the growth of individual students.
- Combine status and growth. Do not throw the baby out with the bath water. Gains provide a logical measure of progress, but students should be making progress towards an outcome of value. Accountability systems should examine school performance from multiple perspectives.
- Do not specify a particular value-added model or a software program within the legislative language. Instead, specify what inferences are desired and the questions that should be answered. For example, “a growth model that evaluates individual student progress towards an acceptable standard of performance by a certain point in time” would be appropriate legislative language. This permits for experts to consider the data and available models, consider what outcome is really of value, and more appropriately structure a timeline to permit for true longitudinal design.
- Insist on transparency. The value-added model should be subject to peer review by a well qualified team of technical experts. While there is no “one correct” method, the

decisions and rules used should be defensible from an educational, statistical, and policy perspective.

- Do not become too quantitative focused. This article rests on the assumption that test score data provide information that is sensitive to instruction. However, there is no assessment that can adequately describe instructional quality on its own.

Recommendations for the Measurement and Statistical Community

- Build the internal capacity of states and districts to fit and interpret value-added models. New software program, such as the **nlme** library in the R software package, HLM (Bryk, Raudenbush, & Congdon, 1996), and SAS Proc Mixed are suitable programs for building VAMs. The **nlme** library has recently been modified to accommodate arbitrarily large Z matrices, permitting for models with crossed random effects within a reasonable amount of computational time (Personal Communication, Douglas Bates). I am currently aware of four articles that provide the transparency required to fit a host of VAMs in software programs (Doran & Lockwood, 2004; Lockwood et al., 2003; Singer, 1998; Tekwe et al., 2004). These articles may form the basis for a series of short courses.
- Do not expect educators to become more statistically inclined. Instead, construct reporting methods that are easily understood and actionable. Communicating variance components, regression coefficients, or levels of statistical significance are likely not to result in classroom or public action. It is important that we construct methods from the complex models that can easily communicate to classroom teachers and school officials with no statistical training.
- Provide empirical research on two “little things”, the choice of covariance structure and the impact of the empirical Bayes (EB) estimates. In the popular TVAAS system,

multiple response variables are estimated simultaneously. In self-contained classroom setting (e.g., an elementary school classroom) all subjects will likely be taught by the same teacher. However, TVAAS fits a diagonal matrix for the teacher effects. Logic, as well as educational theory (e.g., thematic instruction), implies that the effects would not be independent. Second, EB estimates are often defended as providing superior estimates. However, when covariates are included as additional fixed effects the EB estimates “pull” a student’s score towards a conditional expectation. Consider a very small group of high performing minority students in a high performing school. Further assume that, on average across the data, minority students are high performing. The scores of these students would be adjusted towards the conditional mean of the minority students. It seems like something of a paradox to introduce covariates in an effort to adjust for systematic bias in the linear model, which actually may have the effect of introducing bias in the EB estimates.

Conclusion

The value-added model proposed above is primarily focused on the individual student and his rate of change as it relates to an outcome of value. Consequently, this paper argues that value-added models can provide descriptive measures of student growth. The REACH model then uses these parameters couching them within a contextual model. The REACH model proposed is consistent with the design principles of NCLB, and capitalizes on the growth rates obtained from a value-added model. A school level contextual model is also described that would combine multiple sources of data. This four-quadrant model allows for all schools to be organized onto a cartesian grid such that the achievement level of schools may be differentiated.

Technical Appendix

Based on the model for school effects, Equation (5), the predicted values for student i are obtained as follows:

$$\begin{aligned}\hat{Y}_0 &= \beta_0 + \theta_{0j1} + \delta_{0j(i)} \\ \hat{Y}_1 &= \beta_0 + \beta_1 + \theta_{0j1} + \theta_{0j2} + \theta_{1j} + \delta_{0j(i)} + \delta_{1j(i)} \\ \hat{Y}_2 &= \beta_0 + 2\beta_1 + \theta_{0j1} + \theta_{0j2} + \theta_{0j3} + 2\theta_{1j} + \delta_{0j(i)} + 2\delta_{1j(i)}\end{aligned}\tag{10}$$

Therefore, the gain for student i from year 1 to year 2 is represented in the first equation and the gain from year 2 to 3 is represented in the second.

$$\begin{aligned}\text{Gain}_i &= \beta_1 + \theta_{1j} + \delta_{1j(i)} + \theta_{0j2} \\ \text{Gain}_i &= \beta_1 + \theta_{1j} + \delta_{1j(i)} + \theta_{0j3}\end{aligned}\tag{11}$$

The yearly growth rate for student i is the sum of his mean growth trajectory plus a school effect that varies each year.

At the school level, the predicted values are:

$$\begin{aligned}\hat{Y}_0 &= \beta_0 + \theta_{0j1} \\ \hat{Y}_1 &= \beta_0 + \beta_1 + \theta_{0j1} + \theta_{0j2} + \theta_{1j} \\ \hat{Y}_2 &= \beta_0 + 2\beta_1 + \theta_{0j1} + \theta_{0j2} + \theta_{0j3} + 2\theta_{1j}\end{aligned}\tag{12}$$

This makes the school level gain:

$$Gain_j = \beta_1 + \theta_{1j} + \theta_{02j}$$

$$Gain_j = \beta_1 + \theta_{1j} + \theta_{0j3}$$

(13)

*

References

- Bryk, A. S., Raudenbush, S., & Congdon, R. (1996). *HLM: Hierarchical linear and nonlinear modeling with the HLM/2L and HLM/3L programs*. Chicago, IL: Scientific Software International, Inc.
- Davis, A. (1991). Upping the stakes: Using gain scores to judge local program effectiveness in Chapter 1. *Educational Evaluation and Policy Analysis*, 13(4), 380-388.
- Doran, H. C. (2003). Adding value to accountability. *Educational Leadership*, 61(3), 55-59.
- Doran, H. C., & Izumi, L. (2004). *Putting education to the test: A value added model for the state of California* (Tech. Rep.). Sacramento, California: Pacific Research Institute.
- Doran, H. C., & Lockwood, J. (2004). Fitting value-added models in R. *Submitted for publication*.
- Harcourt Brace. (1997). *Stanford 9 achievement test series ninth edition spring norms book*. San Antonio: Harcourt Brace Educational Measurement.
- Hill, R. K., & DePascale, C. A. (2003). Reliability of no child left behind accountability designs. *Educational Measurement: Issues and Practice*, 22(3), 12-20.
- Holland, P. W. (2002). Two measures of change in gaps between the CDFs of test-score distributions. *Journal of Educational and Behavioral Statistics*, 27(1), 3-17.
- Lockwood, J., Doran, H. C., & McCaffrey, D. F. (2003). Using R for estimating longitudinal student achievement models. *The Newsletter of the R Project*, 3(3), 17-23.
- Meyer, R. H. (1997). Value-added indicators of school performance: A primer. *Economics of education review*, 16(3), 283-301.
- Pinheiro, J., & Bates, D. (2000). *Mixed-effects models in S and S-Plus*. New York, NY: Springer.
- Raudenbush, S. S., & Willms, J. (1995). The estimation of school effects. *Journal of educational and behavioral statistics*, 20(4), 307-335.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Second ed.). Newbury Park, CA: Sage.
- Rubin, D., Stuart, E. A., & Zanutto, E. (2004). A potential outcomes view of value-added assessment in education. *Journal of Education and Behavioral Statistics*, 29(1), 103-116.
- Sanders, W. L., Saxton, A., & Horn, S. (1997). The Tennessee value-added assessment system: A quantitative, outcomes-based approach to educational assessment. In J. Millman (Ed.), *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* Thousand Oaks, CA: Corwin Press.
- Schwarz, R. D., Yen, W., & Schafer, W. D. (2001). The challenge and attainability of adequate yearly progress. *Educational Measurement: Issues and Practice*, 20(3), 26-32.
- Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of educational and behavioral statistics*, 24(4), 323-355.

- Tekwe, C. D., Carter, R. L., Ma, C.-X., Algina, J., Lucas, M., Roth, J., Ariet, M., Fisher, T., & Resnick, M. B. (2004). An empirical comparison of statistical models for value-added assessment of school performance. *Journal of educational and behavioral statistics*, 29(1), 11-36.
- Thum, Y. M. (2003). Measuring progress towards a goal: Estimating teacher productivity using a multivariate multilevel model for value-added analysis. *Sociological Methods and Research*, 32(2), 153-207.