CECR

Center for
Educator Compensation
Reform

Understanding the Basics of Measuring Student Achievement

March 2012

Jenna Scott
Jackson Miller
Westat

The work described in this paper was supported by the U.S. Department of Education through the Center for Educator Compensation Reform. The opinions expressed are those of the authors and do not necessarily reflect the view of the U.S. Department of Education, the Center for Educator Compensation Reform, or the institutional partners of the Center. Comments and suggestions are welcome.

The Center for Educator Compensation and Reform (CECR) was awarded to Westat — in partnership with Learning Point Associates, an affiliate of American Institutes for Research; Synergy Enterprises, Inc.; J. Koppich and Associates; and the University of Wisconsin — by the U.S. Department of Education in October 2006.

The primary purpose of CECR is to support Teacher Incentive Fund (TIF) grantees in their implementation efforts through provision of sustained technical assistance and development and dissemination of timely resources. CECR also is charged with raising national awareness of alternative and effective strategies for educator compensation through a newsletter, a Web-based clearinghouse, and other outreach activities.

This work was originally produced in whole or in part by the CECR with funds from the U.S. Department of Education under contract number ED-06-CO-0110. The content does not necessarily reflect the position or policy of CECR or the Department of Education, nor does mention or visual representation of trade names, commercial products, or organizations imply endorsement by CECR or the federal government.

Center for
Educator Compensation
Reform

Allison Henderson, Director
Phone: 888-202-1513
E-mail: cecr@westat.com

34601.0312.83670507

Understanding the Basics of Measuring Student Achievement

State educational agencies (SEAs), local educational agencies (LEAs), and schools can use many approaches to measure student achievement and educator effectiveness. Seven of the primary approaches include attainment, gain, student percentile growth, standard value-added, customized value-added, student learning objectives (SLOs), and teacher portfolios. This paper (1) briefly describes each of the approaches, (2) lists a menu of features SEAs, LEAs, and schools should consider including in their system for measuring student growth, (3) provides a list of conditions SEAs, LEAs, and schools should take into account if they choose to use a value-added model (VAM) (see Attachment 1), and (4) defines the key terminology needed for understanding and discussing measures of student achievement (see Attachment 2).

The Primary Approaches for Measuring Student Achievement and Educator Effectiveness

Attainment and Gain

Attainment and gain are two basic methods of measuring student achievement. States have relied on attainment indicators to demonstrate progress under No Child Left Behind (NCLB). Some states have moved to gain scores to measure the annual change in student or educator effectiveness.

Attainment

Until recently, most Federal, state, and local educational agencies have focused on attainment as the measure for student achievement and educator effectiveness. Attainment scores reflect student performance on a particular assessment at a single point in time. These measures are easy to compute and widely used by school systems to determine performance related to benchmarks such as Adequate Yearly Progress (AYP).

Gain (Simple Growth)

Whereas attainment scores focus on student performance at one point in time, and are thus cross-sectional measures of performance, gain measures take a longitudinal approach. Gain in student test scores is the difference between student performance on a post-test and the same group of students' performance on the corresponding pre-test.

Growth Models

Gain scores are only effective at capturing changes in student or educator achievement when the assessments measure some of the same content across grades. If this is not the case, then growth models may be a more effective measure of performance. Growth scores, which account for the potential differences between tests across grades, use statistical models to compare the difference between actual and expected growth. The following three examples provide a summary of growth models SEAs, LEAs, and/or schools use in educator evaluation systems.

Student Percentile Growth

Student growth percentiles allow SEAs and/or LEAs to compare test score growth across groups of academic peers, which are students with similar test score histories in the same grade and subject. SEAs and LEAs calculate percentile growth scores by ranking each student's growth with all other students who have similar student achievement histories. For example, students who grow at the median rate would earn a rank of 50, indicating that they performed better than half of their academic peers.

Standard Value-Added

Standard value-added models attempt to capture the contribution of a school, classroom/teacher, or other education unit to growth in student achievement. By controlling for prior student test scores, standard models will take into account some of the nonschool factors that contribute to student achievement. However, standard models may exclude certain factors that affect student achievement (e.g., student demographics), which could threaten an SEA's, LEA's, or school's ability to compare student outcomes across schools, grades, and/or classrooms. Despite leaving out some factors that may contribute to student achievement, standard models still provide a more valid comparison of student outcomes than attainment or gain-score models.

Customized Value-Added

Similar to the standard value-added model, customized value-added models can capture the contribution of a school, classroom/teacher, or other educational unit to growth in student achievement. However, unlike standard value-added models, customized value-added models may consider the effects of nonschool factors that contribute to student achievement in specific states and/or districts, such as:

- differences in prior student achievement
- student, teacher, and principal mobility
- student demographics (e.g., race, ethnicity, sex, free or reduced-price lunch, English language learner, special education) and
- student attendance data

In addition to considering the nonschool factors, value-added may also take into account many LEA, school, and classroom-level factors, such as:

- district and/or school resource allocation
- principal characteristics (e.g., level of education, years of experience) and
- teacher characteristics (e.g., level of education, years of experience)

Alternative Teacher Evaluation Measures

In addition to attainment, gain, and value-added growth models, SEAs, LEAs, and schools can also evaluate teachers by using alternative measures of student growth. In some cases, it may be beneficial to use alternative measures of student achievement in conjunction with growth models. The following examples provide two potential alternative methods of evaluating educators based on student growth.

Student Learning Objectives

Student learning objectives (SLOs) are goals set by teachers that specify what students will know, or be able to perform, after completing a quarter, semester, or school year. Teachers set SLOs at the beginning of the quarter, semester, or school year in order to assess whether students achieved the set goals (SLOs). Teachers set their targets based on a thorough review of available data reflecting their students' baseline skills. Most often teachers' targets are set and approved after collaboration and consultation with colleagues and administrators.

Teacher Portfolios

Teacher portfolios are a set of materials that represent teachers' practice as it relates to student learning. Teacher practice extends beyond simply teaching a class and includes all activities that affect student learning. For example, teacher portfolios may include, but are not limited to, a summary of a teacher's experiences and responsibilities, examples of the teacher's students' work, statements about the teacher's goals and objectives for the course, a discussion of the teacher's instructional methods and strategies, statements about the teacher's future goals, and a summary of the teacher's professional development activities. Administrators base their evaluation of educator effectiveness on the commitments educators make in their portfolio.

Menu for Student Growth Considerations

Table 1 provides a list of features SEAs, LEAs, or schools may want to consider including in their student growth model. Completing the table may help guide the development and implementation of a model. In addition, if the SEA, LEA, or school chooses to partner with a vendor to develop its model, then Table 1 may help the vendor identify the needs and capacity of the SEA, LEA, or school. In order to ensure that Table 1 is used effectively, SEAs, LEAs, or schools should provide a detailed justification for each of their answers.

Table 1.

Model Feature	Yes	No	Uncertain	Justification
Should the model be responsive to changes in achievement for all potential scores along the test scale? For example, should the model be responsive to students who score low and high on the assessment?				
Should the model be sensitive to difference in assessment scales across grades and subjects?				
Should the model measure growth by student group (i.e., historic NCLB sub-groups)?				
Should the model consider the relationship between student achievement and school distribution of resources and/or the effect of long breaks from classroom learning?				
Should the model consider the differences across schools and/or classrooms for student demographics (i.e., historic NCLB sub-groups)?				
Should the model accommodate for school year mobility, attendance, and student testing dates (e.g., mid-year testing)?				
Should the model report the uncertainty in the measure (i.e., the confidence interval around the student achievement growth estimate)?				
Should the model account for uncertainty in test measurement error, meaning the difference between what students actually know and what the test indicates that they know?				
Should the model be easy to compute and explain?				

Attachment 1

Value-Added Checklist

The table below will help an SEA, LEA, and/or school determine which type of value-added model is best suited for its state, district, or school. The table lists some of the factors and conditions that an SEA, LEA, and/or school may want to consider when deciding on a value-added model. Since the list is not exhaustive, the SEA, LEA, and/or school may want to consider other issues in conjunction with the questions below. In addition, due to the unique nature of each SEA, LEA, and/or school, not every question will apply in all cases. However, answering all of the questions that apply to the SEA, LEA, and/or school may prove useful for informing decisions regarding the selection of a value-added model and/or vendor.

Value-Added Checklist				
#	Question	Yes	No	Comments
1	Have we selected the specific student achievement assessments and rubrics (such as state accountability tests, locally administered assessments, evaluation rubrics, portfolios, and student learning objectives) upon which we want to base our VAM?	☐	☐	
2	Have we ensured that the assessments we have selected are constructed in a manner that validly measures content knowledge and skill and permits vertical (over time) scores to be compared from one year or grade level to the next year or grade level?	☐	☐	
3	Mid-year testing substantially increases the difficulty of correctly assigning measures of classroom productivity. Are our high-stakes tests, if used as part of the compensation system, close to the beginning or end of the school year?	☐	☐	
3a	If we do not test near the end of the school year, do we want to take into account the fact that the annual growth period (say, March to March) cuts across two school years and, typically, two different teachers or sets of teachers?	☐	☐	
3b	Do we need to include students who changed schools over the summer (that is, within the annual testing interval if tests are not administered near the end or beginning of the school year)?	☐	☐	
4	Do we want the model to support a new compensation system that will use school, grade, or classroom value-added information?	☐	☐	
4a	Will our model support the level we have selected?	☐	☐	

Value-Added Checklist				
#	Question	Yes	No	Comments
5	Do we want to include all students in the model?	☐	☐	
6	Do we want the model to be able to control for demographic characteristics? If so, which ones?	☐	☐	
7	Will we use some control variables that will be subject to manipulation (i.e., special education status)? If so, which ones?	☐	☐	
8	Do we need a model that will account for retained-in-grade and promoted students in the estimation of school effects?	☐	☐	
9	Do we need a model that will consider whether a student has attended a sufficient proportion of instructional days for a teacher to have a legitimate instructional impact?	☐	☐	
10	Do we need the model to include students who changed schools during the school year **and** take account of within-school-year mobility by defining school enrollments in the model as the fraction of the school year enrolled in a given school (dose model)?	☐	☐	
11	Have we determined our desired level of aggregation across units (schools, schools by grade, teacher teams, classroom/ teacher/school) for purposes of comparing productivity? **Note: Statistical precision is highest at the highest level of aggregation since precision increases with the number of students.**	☐	☐	
12	Do we need to use more than one year of test data to improve precision of value-added estimates?	☐	☐	
13	Do we need to be able to separately estimate the productivity of regular school (and teachers), summer school, after school, and NCLB Supplemental Education Services (SES)?	☐	☐	
14	Do we need a model that controls for any/all different types of special education status (i.e., type and severity of handicap)?	☐	☐	
15	Do we need a model that uses multiple years of longitudinal student data that takes into consideration the differences in student achievement across groups of academic peers?	☐	☐	
16	If we want to use prior test scores in our model, will we need the model to control for test measurement error?	☐	☐	

Value-Added Checklist				
#	Question	Yes	No	Comments
17	Will we change tests each year, so that students and staff are not familiar with the test items?	☐	☐	
18	Have we decided whether it is more important that the model be easy to explain to stakeholders or sufficiently complicated to reflect our actual circumstances?	☐	☐	
18a	If we pick a more complicated model, have we considered how we will communicate it to our stakeholders?	☐	☐	
19	Have we determined how much the model will cost? Have we compared the costs of different types of models?	☐	☐	
19a	If we determine we need a more complicated model, are we willing to pay a higher price for a more customized model?	☐	☐	

Attachment 2:

Definition of Terms

The terms below are important to understand when designing and implementing a performance-based compensation system.

Growth Models—a model that compares students' test scores from the current year with those of the previous year to demonstrate each student's progress over the course of the current year. For example, if a fifth-grade teacher's students begin the year reading at a second-grade level and end the year reading at a fourth-grade level, then, although his/her students technically test below a fifth-grade reading level, this teacher was actually able to advance his/her students two years.

Value-Added Models (VAMs)—a quasi-experimental growth model that yields estimates of the contribution of schools, teacher teams, classrooms, or individual teachers to growth in student achievement (or other student outcomes). These models control for other sources of student achievement growth, including prior student achievement, and some models also include student and family characteristics. The models produce value-added indicators under the counterfactual assumption that all schools serve the same group of students. The objective is to facilitate valid and fair "apples to apples" comparisons of student outcomes across schools given that the schools may serve very different student populations.

Expected Growth—a student's predicted growth based on his or her prior attainment. For example, a VAM will predict a student's test score based on his or her prior test scores. A VAM's estimation of expected growth helps to monitor the progress of students. A VAM would expect that students who scored the same on identical standardized tests in previous years (holding other factors constant, if applicable) would have the same test scores for the current academic year. When students in a classroom exceed their expected growth, the classroom teacher is said to have "high value-added."

Residual Growth—the difference in a student's expected growth score and actual (or observed) growth score. For example, although a VAM predicts what a student should score on a specific standardized test, a student may not receive the expected score and instead may obtain an actual test score that is lower or higher than the expected score. Statisticians refer to the difference between the expected score and actual score as residual growth.

Value-Added Estimate—the difference between a teacher's or school's actual effect on an outcome (e.g., student test scores) and the average effect of teachers or schools who work under similar conditions. For example, a value-added estimate may take into account differences in prior student achievement; student, teacher, and principal mobility; student demographics; and student attendance. By constructing a fair comparison group, a value-added estimate provides a sound measure of teacher effectiveness.

Teacher Effect—a VAM estimate that shows a teacher's contribution to students' growth. A VAM teacher effect estimate illustrates a teacher's contribution to students' growth by comparing his or her effect to the average/mean (or median, depending on the model construction) teacher effect at the school, district, and/or state level (holding other factors constant). Often, statisticians refer to residual growth as the "teacher effect."

Measurement Error—an inaccuracy in a test score. Standardized tests are not constructed foolproof and therefore have a level of inaccuracy (known as measurement error) that affects a student's test score. Certain VAM models are able to adjust or correct for measurement error via measurement error estimates.

Validity—a valid measure of teacher effectiveness that allows SEAs, LEAs, or schools to make reliable conclusions about a teacher's effect on a defined outcome (e.g., student performance on end-of-grade/course exams). Thus, a valid measure allows SEAs, LEAs, or schools to determine that teacher A is more/less effective than teacher B. If an SEA, LEA, or school uses a measure to make compensation, retention, and/or hiring decisions, then its model must provide an accurate, causal estimate of a teacher's effect on the defined outcome. In addition, the model should be accurate for all LEAs and schools using the measure to make high-stakes decisions.

Precision—a VAM's ability to produce accurate estimates.