

Making Sense of the Metrics: Student Growth, Value-added Models, and Teacher Effectiveness

Kimberly O'Malley, Ph.D., Katie McClarty, Ph.D., Tracey Magda, Ph.D., and Kelly Burling, Ph.D.

The nation is showing an unprecedented focus on increasing the rigor in education and preparing students for college. The college readiness trend is driving changes in the ways in which the nation uses student test data. Educational data are no longer limited to static data snapshots showing the status of a student performance at one point in time. Instead data are linked grade-to-grade and course-to-course, to create a longitudinal measure of student performance. Inferences about student progress are now made using status as well as growth models. As an illustration of the national focus on longitudinal data, the United States Department of Education (2010) publication, *A Blueprint for Reform: The Reauthorization of the Elementary and Secondary Education Act* noted, "Instead of a single snapshot, we will recognize progress and growth" (p. 2).

States have previously focused on snapshots of student performance and have drawn inferences about progress from those snapshots, assuming that passing in one grade/course meant that students were on track to passing in the next grade/course. The problem is that data supporting those assumptions were not typically provided. In some instances when states did analyze longitudinal data from a system built for static interpretations, the results proved surprising. For example, states transitioning from a horizontal to vertical scale have found that when passing standards are put on a vertical scale and comparisons are made across grades, passing standards for grade level can be lower than the passing standards for the prior grade level. The new national trend is to enhance our ability to draw inferences about student growth by collecting more direct evidence from longitudinal student data.

The use of longitudinal data expands beyond informing about student progress to evaluating teachers and educational leaders. President Obama has repeatedly highlighted the need for teacher effectiveness measures and offered incentives for

those who are willing to implement them. The Department of Education awarded billions of dollars from the Race to the Top fund to 11 states and the District of Columbia in 2010. In granting the awards, the Department evaluated state applications for which 28% of the points were dedicated to a section entitled "Great Teachers and Leaders." As part of the application requirements, states had to develop and describe a system for assessing teacher effectiveness that included student achievement data and provided annual effectiveness ratings for all teachers. States awarded the Race to the Top funds are currently working to implement their plans for teacher effectiveness systems, with most relying on student growth measures as essential measures in their systems.

The college readiness trend is driving changes in the ways in which the nation uses student test data.

The use of student score changes in different applications has led to confusion in the use of terms and concepts. Terms such as student growth, value-added models, and teacher effectiveness are often used interchangeably. The differences in these three measures are significant. Using one when another is intended has impeded the nation's ability to develop these measures well and to use the information in optimal ways.

The goal of this paper is to define student growth, value-added models, and teacher effectiveness, the three terms that are often confused. Furthermore, the paper will compare and contrast features of these three measures and identify next steps needed for advancing the use of these measures for educational reform.

Student growth measures focus on performance of individual students, addressing questions about how much a student progressed and if the student is on track, where on track

typically means on track for passing the next year or two. Student growth measures are developed using student score variables from two or more years. Results from most growth measures produce a label and a score, where the label informs about whether students are on track and the score informs about the gain or expected gain. For example, some student growth models report an on-track indicator such as whether the student made sufficient performance-level transitions (e.g., from high below basic to middle basic) such that the student is on track to reach the proficiency performance level within three years. Regarding the score that is reported from student growth measures, some states report an actual scale-score gain, whereas others report a growth percentile indicating how a student's growth compares to the growth of students with similar score history.

When used in accountability systems, student growth measures can credit students for being on track to reach some future goal or for having achieved a certain level of growth. The ways in which student growth measures are used in federal and state accountability systems vary from state to state, where the variations have led to differential impacts on accountability ratings. Most student growth measures do not include student characteristics such as gender and ethnicity. The reason is that the growth measures calculate the growth a student is expected to make. By including gender or ethnicity in these models, the expectation for a student's growth would be based in part on the student's gender or ethnicity. If a student's gender or ethnicity is one that has been associated with fast growth, that student will have a higher growth

expectation than if that student was of a different gender or ethnicity. Many, including researchers and the United States Department of Education, have questioned the fairness of defining the growth we expect of a student based on student characteristics for which the student has no control. Instead, student growth models define expected growth for all students the same way, requiring the same high expectations of students regardless of student characteristics like gender and ethnicity.

The goal of this paper is to define student growth, value-added models, and teacher effectiveness, three terms often confused.

Value-added models are focused on the effects of teachers and leaders within schools on student score gains, addressing the question of whether students in given classes or schools grow more or less than expected. Some states (e.g., Delaware) and districts, are currently using averages of student growth measures in a value-added manner. For these states and districts, average class or school score gains are used as measures to evaluate educators. More often though, states (e.g., Tennessee, Ohio) and districts (e.g., Dallas Independent School Districts, Tulsa Public Schools) are implementing more formal value-added models, or sophisticated statistical models designed to calculate an empirical estimate of the effect a teacher or leader has on student score gains beyond what is expected.

Though many variations of these more formal value-added models exist, most of the models estimate the value-added measure for a teacher or leader using three steps:

1. Determine the amount of growth expected for the teacher's class (or the leader's school)
2. Calculate the amount of growth the teacher's class actually made
3. Define the difference as the "value" that the teacher added

Unlike student growth models, value-added models are typically created by including student test scores, student demographics, and possibly teacher or school-level variables. In these measures, student characteristics are included because they help make it possible to compare teachers when those teachers have classes of students



with different characteristics. The outcome of a value-added measure is typically a label indicating whether the teacher added measurable value. The score from most formal value-added models indicates how the score gains of the students of a particular teacher compare with score gains of other teachers after accounting for student and teacher characteristics.

Whereas student scores are the only variables used in student growth models, student scores are included along with student and teacher variables in most formal value-added models. Teachers often have no control over the demographic features and characteristics of their students, so including these variables in the models allows the value-added estimates to be made accounting for class composition.

Teacher effectiveness measures are focused on teachers, addressing the question about whether a teacher is effective or ineffective. Using multiple measures (e.g., content knowledge, communication skills, value-added scores, etc.)

obtained through multiple methods (e.g., observations, surveys, portfolios, assessments, etc.), a teacher receives an overall effectiveness rating. For example, teachers could be rated as satisfactory, needs improvement, or unsatisfactory. Teachers could receive a single score on a composite measure or a separate score on each of the measures used. Teacher effectiveness ratings and scores under this approach can be used to identify which teachers and in which areas teachers might need professional development or for teacher accountability.

Whereas some districts like Washington, DC have already established measures of teacher effectiveness, most states and districts are currently creating these metrics. Because relatively few teacher effectiveness measures exist compared to the prevalence of both growth and value-added models, there is limited empirical evidence to guide their development. Table I provides a summary of the distinguishing features of growth measures, value-added models, and teacher effectiveness.

Table I: Distinguishing Features of Student Growth, Value-Added, and Teacher Effectiveness Measures.

FEATURES	GROWTH	VALUE-ADDED	TEACHER EFFECTIVENESS
FOCUS	Student	Teacher/leader	Teacher/leader
QUESTION ANSWERED	How much did the student progress? Is the student on track?	Did the students grow more, or less, than expected?	Is the teacher effective or ineffective in improving student outcomes?
OUTPUT	1. Score/label showing student score gain or predicted gain 2. Yes/No for on track	Value showing how score gains of the students of that teacher or leader compare with average score gains	Overall effectiveness rating (e.g., satisfactory, needs improvement, unsatisfactory) and/or score
VARIABLES	Student scores only	Student scores and demographics Teacher demographics	Multiple measures (e.g., content knowledge, communication, value-added score) Multiple methods (e.g., observation, survey, portfolio, assessments)

Moving from the use of static snapshots of student performance to a combination of static and longitudinal measures offers the opportunity for our nation to focus on where each student is, the progress students made to get there, and student growth needed for college and career readiness. Furthermore, evaluations of teachers and leaders can now be made considering student growth and progress to educational goals. These are good goals. However, reaching them will require that educators, administrators, and policy-makers understand what these new measures are and how they are created.

Once the basics are understood, educators, administrators, and policy-makers can identify which variations within each of these measures (e.g., value-table approach to growth versus a projection measure) best match the data conditions and the political context within each state or district. They can apply and interpret these measures in appropriate ways for improving student learning, providing professional development for teachers, and holding teachers and leaders accountable for preparing our nation's students for college and careers.

References

Duncan, A. & Martin, C. (2010). *ESEA blueprint for reform*. Washington, DC: U.S. Department of Education, Office of Planning, Evaluation and Policy Development.

About Pearson

Pearson, the global leader in education and education technology, provides innovative print and digital education materials for preK through college, student information systems and learning management systems, teacher licensure testing, teacher professional development, career certification programs, and testing and assessment products that set the standard for the industry. Pearson's other primary businesses include the Financial Times Group and the Penguin Group. For more information about the Assessment & Information group of Pearson, visit <http://www.pearsonassessments.com/>.

About Pearson's Research Bulletins

Pearson's research bulletins help clarify specific assessment-related topics for educators, parents, students, researchers and policy makers through brief and non-technical discussions. Pearson's publications in .pdf format may be obtained at: <http://www.pearsonassessments.com/research>.