In his textbook on statistics, George Box wrote, “essentially, all models are wrong, but some are useful.” This is true of the growth models used to represent student academic progress in schools and districts. These models purport to describe how much growth is occurring among and within groups of students. They represent a major shift by the U.S. Department of Education (USED) from its original No Child Left Behind accountability approach for calculating adequate yearly progress (AYP). The AYP approach didn’t recognize and credit schools with growth of students who were classified below proficient, regardless of the amount of progress they made toward proficiency.
USED uses five growth models to calculate AYP of schools and districts: value added, value table, trajectory, projection, and student growth percentile. Most of the models involve complex statistical calculations to describe growth of the student or student group in comparison with expected growth or with growth of other groups.

### Five Models

1. **Value added model (VAM).** Created by statistician William Sanders in 1992 and first used for accountability in Tennessee, VAM isolates teacher effect from other variables that have an impact on student learning. VAM has several variations, each with its own strengths and weaknesses. One approach is calculating, for each student, the difference between scale scores or normal curve equivalents (NCE) from one year to the next, which generates a gain score. The average of these gain scores, ideally using scores from at least three prior years, is added to the most recent scale score or NCE, with the result representing the predicted score for the student in the current year.

   If the student’s actual score or NCE is greater than predicted, the model assumes positive teacher effect. If it is less than predicted, negative effect is indicated. The difference between the actual and predicted scores is calculated for each student, and the average of those differences represents teacher effect. These are further aggregated to calculate school effect.

   If the effect is statistically significant (either positive or negative), it is attributed to the teacher or school. In this model, each student becomes his or her own control, with the predicted score being based on that student’s past performance with other teachers. An important concept in assessment is the standard error of measurement (SEM). This statistical quantity represents the precision of a score. That is, there is a 95 percent probability that the true scores for the students (actual achievement level) will fall within ± 2 SEM of the reported score from the test. SEM can have an impact on VAM in that smaller SEM contributes to a smaller standard error of the value added estimate for the classroom.

2. **Value table model.** This model assigns points to the school for each student who moves from a lower proficiency category (e.g., below basic or basic) to a category closer to or at proficient. Points are also assigned for proficient and advanced students who maintain proficiency. No points are awarded for students who remain in or regress to a category below proficient. The school’s value table score is the average of all student value table scores. Scale scores on the test are not used to calculate growth and have no impact on AYP, except as cut points related to proficiency categories. Each state using this model determines the categories, the number of points in each category, and the value score necessary for the school to achieve AYP.

3. **Trajectory model.** Starting with the student’s most recently measured performance level, a linear path of scale scores is created that, if achieved, will culminate in classification of the student as proficient, typically after three or four years. Student progress can be tracked in relationship to the trajectory line, with each student’s actual score being compared with the score required by the trajectory. AYP is met if the school has a sufficiently large proportion of students who are either proficient or are showing progress on or above the trajectory line.

4. **Projection model.** This model uses statistical calculations to predict the likelihood that the student’s current growth trajectory will result in proficiency by a target date. The calculations used to plot the trend line predicting future performance include all of the student’s past scores; these calculations may also include scores of academic peers. If a student’s past performance suggests he will not reach proficient by the target year, this student would count against AYP for the school in the current year. If the slope of the student’s trend line shifted to intercept proficiency by the target grade, his or her score would support AYP for the school even though the student would not yet have achieved proficiency.

5. **Student growth percentile model (SGP).** Also known as the Colorado Growth Model, this approach starts by grouping students with historically similar scale scores, designated as academic peers. The growth of those students is compared and percentiles assigned to each student. Using quantile regression analysis of past scores for the academic peers, a prediction of future performance is made for each student. Growth is expressed as a percentile of the scores of the student’s academic peers.

   Application of this analysis to AYP is conducted differently by each state,
Growth Measurement in Action

Bryan Garcia, principal of Puesta del Sol Elementary in New Mexico’s Rio Rancho Public Schools, practices data-based decisionmaking. Decisions made by the school’s teachers and leaders are informed by measurement data on student academic growth. Garcia believes that a comprehensive assessment system—drawing on day-to-day formative assessments, interim assessments, and summative assessments—has been one of the most important keys to supporting student growth and improving professional practice.

With the help of an instructional core team led by Diane Earnest, the school’s instructional coach, Garcia has used this approach to create and facilitate a data study process that moves beyond compliance and accountability to providing important information for teachers to assist with instructional planning. At multiple points in the year, grade-level based collaborative teaching teams (CTTs), with the support of the instructional core team, review a visual representation of where students are on two measures: predicted proficiency on the state exam and progress toward individual growth targets. The CTTs then use the data to discuss the learning needs of each student, including identification of which students need intervention and which are ready for more advanced material. Puesta del Sol educators also use the process to help students take ownership of their learning.

“For us, student engagement has been absolutely critical. When students set goals with teachers for their own learning, it’s very motivational,” Garcia says.

The strength of Puesta del Sol’s student growth scores has moved the school from a “D” to a “B” ranking in the New Mexico Public Education Department’s accountability system in just two years.
VAM, viewed as reducing bias and error, is increasingly used by states and districts as a major factor in teacher evaluation, retention, promotion, and compensation. Use of student assessment data for this purpose is controversial and has become a concern for principals and teachers nationwide.

In Problems With the Use of Student Test Scores to Evaluate Teachers, a 2010 Economic Policy Institute publication, a number of prominent scholars looked at the validity of VAM for teacher evaluation. Their findings described VAM as an imprecise and unstable measure of teacher effectiveness, particularly if the student assessment data are not of high quality or do not cover a sufficient number of years from which to predict student achievement with any degree of accuracy. The authors warned specifically of the likelihood that effective teachers will be classified by VAM as ineffective, and vice versa.

In the same year, a policy report from the Brookings Institute, Evaluating Teachers: The Important Role of Value-Added suggested that VAM should not be measured against an abstract ideal, but rather should be compared to other teacher evaluation methods to determine its potential usefulness. The authors suggest that, while far from perfect, VAM better predicts some teaching outcomes than other evaluation processes. When used in conjunction with principal observation and other measures of teacher performance, VAM increases the validity and reliability of the evaluation process and contributes to improvement of educator evaluation systems, according to the report.

Regardless of the research and debate on the use of VAM for accountability, policymakers have embraced the logical appeal of using student outcomes to measure the performance of the individuals with the greatest influence on those outcomes. Since it has become a reality for many principals and teachers, an analysis of VAM’s value beyond accountability is worthwhile.

VAM data can provide useful information for strengthening instruction, by providing a look at how students with specific characteristics respond to a teacher’s instruction. For example, William Sanders identified three common patterns related to low, average, and high-achieving students: the shed, reverse shed, and tent patterns.

The shed pattern represents higher growth for the most challenged students, meaning that the teacher may be contributing to closure of the achievement gap at the expense of the most advanced students. The reverse shed shows high teacher effect for the most advanced students, but little growth for the less able pupils. The tent pattern may indicate that the teacher is focusing on curriculum for the middle of the class. Each of these patterns indicates that two groups are not being appropriately challenged.

Grouping by achievement level is not the only way to look at teacher influence. Student results may be grouped by demographic characteristics, participation in specific programs, or a host of other classifications. VAM provides a way to look at these subgroups for instructional purposes to see what strategies produce the best results for targeted student populations. Knowing what patterns are present among your teachers is the first step in changing those patterns. Even the best teacher can and should improve.

No single data point constitutes a complete picture or establishes a trend. VAM data become more useful as multiple years are examined, revealing trends and changing patterns. Multiple years of VAM data, particularly when triangulated with principal observation, student and parent feedback, peer observation, self-reflection, and other information, can mitigate much of the instability associated with VAM. More importantly, it can help every teacher strengthen his or her performance.

One of the ongoing challenges with growth models and other data used for accountability is that teachers may be less inclined to embrace those data for instructional improvement if they also serve as a Sword of Damocles. Principals can help to mitigate this challenge and should understand that, even with their inherent limitations, these growth models can play a useful and important role in helping to increase student learning.

Raymond Yeagley, a former principal, is vice president and chief academic officer at Northwest Evaluation Association (NWEA).