



POSITION STATEMENT: VALUE-ADDED MEASURES IN TEACHER EVALUATION



PURPOSE

To determine the efficacy of the use of data from student test scores, particularly in the form of Value-Added Measures (VAMs), to evaluate and make key personnel decisions about classroom teachers.

ISSUE

Currently, a number of states are either adopting or have adopted new or revamped teacher evaluation systems, which are based in part on data from student test scores in the form of value-added measures (VAMs). Some states mandate that up to 50 percent of the teacher evaluation must be based on data from student test scores. States and school districts are using the evaluation systems to make key personnel decisions about retention, dismissal, and compensation of teachers and principals.

At the same time, the Every Student Succeeds Act requires that states adopt and implement challenging academic content and achievement standards. These new standards are intended to raise the bar from having every student earn a high school diploma to the much more ambitious goal of having every student be on target for success in college, career, and life.



POSITION STATEMENT: VALUE-ADDED MEASURES IN TEACHER EVALUATION

The assessments accompanying these new standards depart from the old, much less expensive, multiple-choice style tests—shifting to assessments with constructed responses. These new assessments demand higher-order thinking and up to a two-year increase in expected reading and writing skills. Not surprisingly, the newness of the assessments combined with increased rigor has resulted in significant drops in the number of students reaching proficient levels on assessments aligned to the new standards.

New teacher evaluation systems demand the inclusion of student data at a time when scores on new assessments are dropping.

Herein lies the challenge for principals and school leaders. New teacher evaluation systems demand the inclusion of student data at a time when scores on new assessments are dropping. The fears accompanying any new evaluation system have been magnified by the inclusion of data that will get worse before it gets better. Principals are concerned that the new evaluation systems are eroding trust and are detrimental to building a culture of collaboration and continuous improvement necessary to successfully raise student performance to college- and career-ready levels.

Specific questions have arisen about using VAMs to retain, dismiss, and compensate teachers. VAMs are statistical measures of student growth. They employ complex algorithms to figure out how much teachers contribute to their students' learning, holding constant factors such as demographics. At first glance, it would appear reasonable to use VAMs to gauge teacher effectiveness. Unfortunately, policymakers have acted on that impression over the consistent objections of researchers who have cautioned against this inappropriate use of VAMs.

In a 2014 report, the American Statistical Association urged states and school districts against using VAM systems to make personnel decisions. A statement accompanying the report pointed out the following:

- VAMs are generally based on standardized test scores and do not directly measure potential teacher contributions toward other student outcomes.
- VAMs typically measure correlation, not causation: Effects—positive or negative—attributed to a teacher may actually be caused by other factors that are not captured in the model.
- Under some conditions, VAM scores and rankings can change substantially when a different model or test is used, and a thorough analysis should be undertaken to evaluate the sensitivity of estimates to different models.
- VAMs should be viewed within the context of quality improvement, which distinguishes aspects of quality that can be attributed to the system from those that can be attributed to individual teachers, teacher preparation programs, or schools.
- Most VAM studies find that teachers account for about 1 to 14 percent of the variability in test scores, and that the majority of opportunities for quality improvement are found in the system-level conditions. Ranking teachers by their VAM scores can have unintended consequences that reduce quality.



POSITION STATEMENT: VALUE-ADDED MEASURES IN TEACHER EVALUATION

Another peer-reviewed study funded by the Gates Foundation and published by the American Educational Research Association (AERA) stated emphatically, “Value-added performance measures do not reflect the content or quality of teachers’ instruction.” The study found that “state tests and these measures of evaluating teachers don’t really seem to be associated with the things we think of as defining good teaching.” It further found that some teachers who were highly rated on student surveys, in classroom observations by principals, and through other indicators of quality had students who scored poorly on tests. The opposite also was true. “We need to slow down or ease off completely for the stakes for teachers, at least in the first few years, so we can get a sense of what do these things measure, what does it mean,” the researchers admonished. “We’re moving these systems forward way ahead of the science in terms of the quality of the measures.”

Researcher Bruce Baker cautions against using VAMs even when test scores count less than 50 percent of a teacher’s final evaluation. Using VAM estimates in a parallel weighting system with other measures like student surveys and principal observations “requires that VAMs be considered even in the presence of a likely false positive. New York legislation prohibits a teacher from being rated highly if their test-based effectiveness estimate is low. Further, where VAM estimates vary more than other components, they will quite often be the tipping point—nearly 100 percent of the decision even if only 20 percent of the weight.”

Still, other researchers believe VAMs are flawed at their very foundation. The use of test scores via VAMs assumes “that student learning is measured by a given test, is influenced by the teacher alone, and is independent from the growth of classmates and other aspects of the classroom context. None of these assumptions is well supported by current evidence” (Darling-Hammond et al., 2012). Other factors including class size, instructional time, home support, peer culture, and summer learning loss impact student achievement. Darling-Hammond points out that VAMs are inconsistent from class to class and year to year. VAMs are based on the false assumption that students are randomly assigned to teachers. VAMs cannot account for the fact that “some teachers may be more effective at some forms of instruction ... and less effective in others.” Recent analysis by Jesse Rothstein has further enforced this conclusion, demonstrating that even the newer, more sophisticated VAMs remain importantly biased and cannot be relied upon to effectively evaluate teachers.



NASSP GUIDING PRINCIPLES

- As an instructional leader, “the principal’s role is to lead the school’s teachers in a process of learning to improve teaching, while learning alongside them about what works and what doesn’t” for those students in that school.
- NASSP believes that teaching is a complex craft and evaluation of effective teaching should be based on close examination of a variety of variables through quantitative and qualitative data, take into account the context in which a teacher works, and not be limited to students’ performance on standardized tests.
- *Building Ranks™: A Comprehensive Framework for Effective School Leaders* highlights the principal’s role in human capital management and encourages principals to adopt strategies to recruit, retain, and continually develop teacher talent.



POSITION STATEMENT: VALUE-ADDED MEASURES IN TEACHER EVALUATION

- The teacher evaluation system should aid the principal in creating a collaborative culture of continuous feedback and ongoing critical conversations that will lead to incremental improvements in teaching and learning.
- Assessment for teacher and student learning is critical to continuous improvement of teachers.
- Data from student test scores should be used by schools to move students to mastery and a deep conceptual understanding of key concepts as well as to inform instruction, target remediation, and focus review efforts.
- NASSP supports recommendations for the use of “multiple measures” to evaluate teachers as indicated in the 2014 “Standards for Educational and Psychological Testing” measurement standards released by leading professional organizations in the area of educational measurement, including the AERA, American Psychological Association, and the National Council on Measurement in Education.



RECOMMENDATIONS FOR FEDERAL, STATE, AND DISTRICT POLICYMAKERS

- Successful teacher evaluation systems should employ “multiple classroom observations across the year by expert evaluators looking to multiple sources of data and provide meaningful feedback to teachers.”
- Districts and states should encourage the use of Peer Assistance and Review programs, which use expert mentor teachers in supporting novice teachers and struggling veteran teachers, and which have been proven to be an effective system for improving instruction.
- States and districts should allow the use of teacher-constructed portfolios of student learning, which are being successfully used as a part of teacher evaluation systems in a number of jurisdictions.
- VAMs should be used by principals to measure school improvement and determine the effectiveness of programs and instructional methods.
- VAMs should be used by principals to target professional development initiatives.
- VAMs should not be used as a primary source to make key personnel decisions about individual teachers.
- States and districts should provide ongoing training for principals in the appropriate use of student data and VAMs.
- States and districts should make student data and VAMs available to principals at a time when decisions about school programs are being made.
- States and districts should provide the resources and time principals need in order to determine the best use of data and help teachers better understand how to use VAM data to measure their own effectiveness or that of their programs.
- The U.S. Department of Education should support ongoing research to establish the validity and reliability of comprehensive teacher evaluation programs, further examine the efficacy of value-added measures for teacher evaluation, and support adequate training and professional development of evaluators to ensure fidelity of implementation of evaluation models found to be effective in improving teaching and learning.