NATIONAL EDUCATION ASSOCIATION
nea.org
Great Public Schools for Every Student

*Collective Bargaining & Member Advocacy*

1201 16th St., N.W. | Washington, DC 20036 | Phone: (202) 822-7080 | Fax: (202) 822-7833

# WHAT YOU SHOULD KNOW ABOUT VALUE-ADDED MODELS

The push to use student test scores as a means to evaluate and reward teachers has grown significantly over the past several years. Value-added models (VAMs) are statistical models that, in the education context, use test scores to project student growth over time. Some foundations, think tanks, and school district leaders are touting VAMs as the type of model best able to isolate and assess the effectiveness of a particular teacher.

While some regard value-added models to be accurate measures of teacher effectiveness, many statistical and academic experts have identified significant weaknesses in the precision of these models. Although some VAMs have the potential to provide insight into student achievement or instructional practice, there are many outstanding questions about the reliability and validity of these models and the estimates they produce. Accordingly, utilizing VAM results as a driver for high stakes decisions related to teacher evaluation or pay is extremely problematic. This factsheet summarizes the limitations associated with VAMs and includes references to complete reports and other resources at its conclusion.

**VAMs are not always designed to, or capable of, accurately isolating the effects of a single teacher on a student. One reason is that VAMs often do not reflect the realities of the classroom or students' learning environment.**

- In most school districts, the majority of teachers do not teach students who take standardized tests in their subject area. As a result, most teachers cannot be "valued" under a VAM based only on math and/or reading. Teachers who cannot be included in such a measure include most kindergarten through grade 2 teachers, some grade 3-8 teachers, most secondary school teachers as well as teachers of art, music, technology, physical education, and foreign languages. VAMs cannot measure these teachers' effectiveness at all.

- Students are often not assigned to classrooms and teachers on a random basis. VAM scores are influenced by non-random student assignments that result in certain teachers having students with better attendance, better standardized test performance, better behavior, or better academic records and performance. In addition, it is impossible to fully separate out the influences of students' other teachers, as well as school conditions, on students' learning.

- VAMs often do not consider student out-of-school learning such as parental input, tutoring, and summer and on-line programs. Even those VAMs that do factor in some student information factors do an incomplete job and cannot capture the wide variation in out-of-classroom student learning and experiences. If these factors are not adequately measured, teachers' scores will be inaccurate. Summer learning loss is also a significant problem and tends to affect high poverty students more significantly.

- VAMs are especially problematic for teachers who teach special education students, English language learners, and lower income students who may have more learning issues and limited access to the services outlined above. Test scores of these special populations tend to be more unreliable.

- Many district-wide policies beyond the teacher's control (curriculum, school physical conditions, school resources, class size) can impact student test scores. When these types of conditions vary across schools within a district, a teacher's VAM score can be adversely impacted.

- VAMs do not accurately reflect aspects of team teaching or the prevalence of pull outs where some students receive additional support.

- Many experts acknowledge that VAMs are better suited for elementary school rather than middle or secondary school teachers, where students are taught by several teachers at any given time. However, team teaching is common in many elementary schools and could muddy results.

- It is difficult to track students who have high mobility rates and assign a VAM score to a particular teacher.

**VAM results can be volatile from one year to the next, making it an unreliable measure for high stakes decisions.**

- Volatility is due to a number of complex statistical reasons, including the fact that VAMs based on changes from one year to the next are not usually based on tests that are "vertically scaled." This means that there is no real linkage in curriculum from one year to the next, which makes it difficult to adequately measure real student growth.

- Erratic results, where a teacher may be rated at the top of the rankings one year and then drop to the lower end the next, tend to occur when aggregate student test scores are based on smaller, rather than larger samples. Student samples will decrease over time due to student mobility, missing data, and other factors. Sample size may be small to start with in smaller, rural schools. As a result, teacher rankings could change dramatically over the course of multiple years.

- For teachers with a smaller number of students, a few students with poor performance can disproportionately decrease their ranking.

**VAMs may work against collaborative teaching and could create a host of unintended consequences.**

- If VAMs are going to be used to rank teachers against their peers based on test scores, then teachers would essentially be encouraged to collaborate less with colleagues. This would clearly undermine the efforts to build strong schools where administrators and teachers work together cohesively.

**Providing pay based on student test scores may encourage "Teaching to the Test" – and only in a couple of subjects.**

- Because most VAMs rely on just reading and math test scores, schools are focusing more on these subjects at the expense of other subjects such as science, social studies, foreign languages, history, and the arts. In addition, schools have less time to hone students' critical thinking, advanced writing, and problem-solving skills. Even within math and reading, state tests depend on students answering multiple choice questions, rather than comprehensive writing samples or complex problem solving. This leads to repetitive drilling to enhance students' test taking abilities but does not improve more advanced reasoning skills.

- Research also shows that school districts may focus on those children "on the bubble" – that is, those who can at least easily reach average growth level with assistance rather than the neediest students who need the most help.

**VAMs have many unresolved technical and implementation issues that limit their utility as a primary source for high stakes decisions.** There are challenging technical issues to be resolved in creating a value-added model that meets a state's or school district's policy needs.

- The regression models that serve as the basis for VAMs are highly complex and vary among states and school districts, which must invest scarce resources into helping teachers, administrators, and the public understand them and determine how best to use the results. Training for administrators, teachers, human resources, IT, and other school employees is required to build understanding of a VAM within a school district.

- VAM models have rigorous data requirements. These models utilize multiple years of test data that must be comparable from year to year. In addition, there must be a capacity to track individual student scores from one year to the next (and sometimes from one district to another in the state). This capacity often requires a statewide student identification system. Also, school districts and states often need to upgrade their data systems to meet the needs of a VAM model.

- Implementing a VAM model requires a degree of human resources and psychometric expertise that most states and districts are not equipped with. Many states and school districts must contract out to acquire this expertise, which can be time consuming and expensive.

**It is extremely difficult for teachers or administrators to draw any useful guidance about improving teaching practice from VAMs and their results.**

- Teachers typically receive the results from the VAM model in the fall following the year in which their students' took the tests. They no longer teach these students and may not even teach the same grade. Because so much time has elapsed since the students took the tests, teachers struggle to determine how to apply the results to their current students.

**Additional Resources:**

Baker, Eva L, et al, *Problems with the Use of Student Test Scores to Evaluate Teachers,* EPI Briefing Paper, Briefing Paper 278, Washington, DC, August 2010.

Braun, H., *Using Student Progress to Evaluate Teachers: A Primer on Value Added Models,* Technical report, Princeton, NJ: Educational Testing Service, Policy Information Center, 2005. (Available at: www.ets.org/Media/Research/pdf/PICVAM.pdf)

Buddin, R., D. McCaffrey, S. Nataraj Kirby, and N. Xia. *Merit Pay for Florida Teachers Design and Implementation Issues,* Santa Monica, CA: RAND

Goldschmidt, P., et al, *Policymakers' Guide to Growth Models for School Accountability: How Do Accountability Models Differ*? The Council of Chief State School Officers, Washington, D.C., October 2005.

Hamilton, L., B. Stecher, J. Marsh, J. McCombs, A. Robyn, J. Russell, S. Naftel, and H. Barney, *Standards Based Accountability under No Child Left Behind: Experiences of Teachers and Administrators in Three States,* Santa Monica, CA: RAND Corporation, MG-589-NSF, 2007.

Harris, Douglas, *Would Accountability Based on Teacher Value Added Be Smart Policy? An Examination of the Statistical Properties and Policy Alternatives*, American Education Finance Association, 2009.

Kane, T., D. Staiger, *Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation,"* National Bureau of Economic Research, Working Paper 14607, December 2008.

Koedel, C., J. Betts. *"Does Student Sorting Invalidate Value-Added Models of Teacher Effectiveness? An Extended Analysis of the Rothstein Critique,"* National Bureau of Economic Research, April 2009.

Kupermintz, H. "Teacher Effects and Teacher Effectiveness: A Validity Investigation of the Tennessee Value Added Assessment System'," *Educational Evaluation and Policy Analysis,(25)*3, 287-298, 2003.

Martineau, J., "Distorting Value Added: The Use of Longitudinal, Vertically Scaled Student Achievement Data for Growth Based Value Added Accountability," *Journal of Educational and Behavioral Statistics,* Vol. 31, No. 1, 2006.

Miller, R. "Adding Value to Discussions About Value-Added: A New Framework for Talking About Teacher Effectiveness." *Center for American Progress*, Washington, D.C. December 2009.

National Research Council and National Academy of Education. (2010) *Getting Value Out of Value-Added: Report of a Workshop.* Committee on Value-Added Methodology for Instructional Improvement, Program Evaluation, and Educational Accountability, Henry Braun et al, Editors. Center for Education, Division of Behavioral and Social Sciences and Education, Washington, DC: The National Academies Press.

Rothstein, J. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement," Princeton University and National Bureau of Economic Research, May 2009.

Rowan, B., R. Correnti, and R. J. Miller, "*What Large-Scale Survey Research Tells Us About Teacher Effects on Student Achievement: Insights from the Prospects Study of Elementary Schools*," The Consortium for Policy Research in Education, November 2002.

See also a Daniel Willingham video: Merit Pay, Teacher Pay, and Value Added Models: You Tube http://www.youtube.com/watch?v=uONqxysWEk8

***To learn more, contact the NEA Collective Bargaining and Member Advocacy Department at 202-822-7080 or mailto:collectivebargaining@nea.org.***

***October 2010***