D. Measurement

What does the research suggest about ways to structure rewards so that teachers are not penalized for working in low-performing schools?

When making high-stakes decisions like determining educator salaries, all teachers must be evaluated fairly. Though many models link student achievement to teacher effectiveness, it is essential to use models that do not penalize teachers who work in low-performing schools.

Attainment models capture student achievement at one point in time. This indicator of achievement is often compared to an established target, such as percent proficient on a state assessment. Comparing teachers, grades, or schools based on one test score from one year is an example of attainment modeling. Many states and districts require a certain percentage of their students to test into the proficient category or higher on their state assessment. Student achievement is then measured as the "percent proficient" for a given school, state, or district. For example, District X may be 62 percent proficient, while District Y is 84 percent proficient. This is an example of an attainment model of evaluation: it measures student achievement against a target goal—the proficient standard on the state assessment. This model, however, fails to control for prior student achievement or student characteristics.

Simple growth models—sometimes referred to as gain models—measure student achievement by taking into account prior student performance. In determining teacher effectiveness, a school using a growth model would take a student's test score from the current year and compare it to last year's score to see how much progress the student made. For example, a fifth-grade teacher's students begin the year reading at a second-grade level. At the end of the year, they are all reading at a fourth-grade level. Though her students technically still test below a fifth-grade reading level, this teacher was actually able to advance her students two years. Thus, the teacher would be considered quite effective as determined by a growth model. Growth models give teachers credit if their students are improving, even if they are still not high achievers. However, growth models fail to account for student characteristics that affect student achievement.

Value-added models take into account prior student performance while controlling for outside factors, such as poverty, family background, and previous teachers' effectiveness. Controlling for these factors allows schools and districts to isolate the effect that a particular teacher—or in some cases, a group of teachers—had on a student and give a clearer representation of that teacher's true effectiveness (Braun, 2010; Braun, Chudowsky, & Koenig, 2010; Harris, 2007; Kane & Staiger, 2008; Meyer, 1996; Rivkin, 2007; Sass, 2008).

Examples of characteristics that value-added models can control for include family income, parental involvement, English language proficiency, peer characteristics, and school quality.

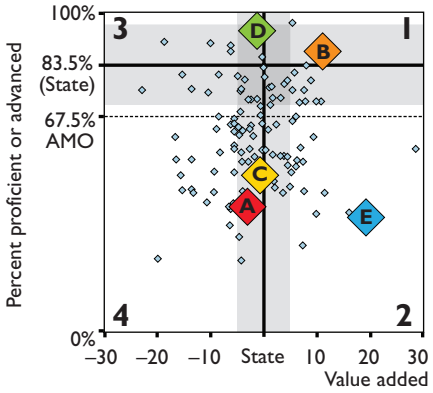
Students in low-performing schools often enter the classroom far behind where a state or district expects them to be, and, consequently, their test scores reflect that. Since value-added models look at prior student achievement, teachers who are assigned low-performing students are not penalized if their students show appropriate growth (Goldhaber & Hansen, 2008; Meyer, 1996; Sass, 2008; Winters, Greene, Ritter, & Marsh, 2008). Much like growth models, value-added models evaluate students based on how far they progressed during the year, rather than on their scores at the end of the year. What really distinguishes value-added from other models, though, is its ability to account for student characteristics.

Value-added modeling is optimal for evaluating teacher performance in any type of school, including low-performing schools, because it measures individual student growth over time and can be adjusted to control for outside factors such as student-family characteristics and former teacher's effectiveness. In a pay-for-performance system that uses value-added modeling, teachers are rewarded for increasing student achievement, regardless of their student's backgrounds and prior achievement. Pay systems that reward teachers when students reach a certain level of proficiency will disproportionately favor teachers who work in affluent schools, which does not provide an incentive for teachers to work in high-poverty, low-performing schools.

The graph at the right depicts value-added scores for a large, urban district. The X axis shows schools' average value-added scores, while the Y axis shows the schools' average attainment (defined as percent proficient or advanced on the state assessment). Each data point represents a different school in the district. Data points A, C, and D have very similar value-added scores, but quite different attainment scores. School D is above the state average, while Schools A and C fall far below it. This example demonstrates how misleading attainment measures can be; they often disproportionately favor high-performing schools. An even more interesting comparison, though, is School E and School D. While School E's students clearly fall well below the state average of percent proficient, its teachers have one of the highest value-added scores in the entire district. As this district's data shows, it is clear that teachers who work in low-performing schools (like School E) will not be penalized when evaluated using a value-added model.

Several studies have addressed the reliability of value-added models to accurately assess teacher effectiveness in both high- and low-performing schools. The Tennessee Value-Added Assessment

Value-added and percent proficient scores:
Fourth Grade Reading



Quadrant 1 – High value-added, high attainment
Quadrant 2 – High value-added, low attainment
Quadrant 3 – Low value-added, high attainment
Quadrant 4 – Low value-added, low attainment

Meyer et al. (2009)

System (TVAAS) system was one of the first systems in the country to use value-added measurement (the Sanders model) statewide for its schools. When calculating teacher effectiveness, the state found that highly effective teachers were effective across all student demographics—regardless of prior achievement, race/ethnicity, or economic status (Sanders & Horn, 1998; Sanders & Rivers, 1996).

Goldhaber and Hansen (2008) found that value-added estimates of teacher effectiveness remained the same when teachers moved from low-minority to high-minority districts. They write, "One might imagine teaching in high-minority schools may unduly penalize our estimates of teacher performance, or inappropriately reward teachers with low minority levels; however, there appears to be no detectable change in this analysis."

In a study of the Little Rock, Arkansas, pay-for-performance program, Winters, Greene, Ritter, and Marsh (2009) found that teachers in low-performing schools actually had higher value-added scores than those in high-performing schools. Five schools in the district implemented a pay-for-performance system that gave bonuses to teachers based on their student's improved achievement. In comparison to the control group—comprised of similar, untreated schools in the Little Rock School District—students in the pay-for-performance schools who had previously scored the lowest on the state assessments showed the largest learning gains. Additionally, the performance-pay model had the greatest positive impact on the lowest performing teachers.[1]

While many statisticians believe value-added modeling is the most objective, fair way to evaluate teachers, they also caution against using these results cavalierly (Braun, 2005; Goldhaber & Hansen, 2008; Harris, 2007; Rivkin, 2007). There will always be a certain amount of noise—measurement error that affects the validity of distinctions by effectiveness between teachers and schools—when calculating value-added scores. Additionally, the data systems required to perform value-added modeling are very robust. Many schools and districts do not currently have the information systems infrastructure to accommodate such data collection (Braun, 2010; Thorn, 2001). Districts with high rates of student mobility may find it difficult to produce value-added estimates because they lack data on incoming and transfer students, though researchers are now developing viable solutions to this problem (Meyer & Dokumaci, 2010). Even for districts that are able to calculate value-added scores for their teachers, a growing body of research now suggests that value-added scores vary considerably from year to year (Goldhaber & Hansen, 2008; Koedel & Betts, 2009; McCaffrey, Sass, Lockwood, & Mihaly, 2009; Sass, 2008).

Most agree, though, that value-added measurements, particularly when combined with multiple measures of evaluation—like principal evaluations, classroom observations, etc.—are the most accurate indicators of teacher effectiveness (Braun 2005; Goe, Bell, & Little, 2008; Goldhaber & Hansen, 2008; Kane & Staiger, 2008; Rivkin, 2007). Using value-added indicators levels the playing field for all teachers.

1 Performance was measured by the average test score gain of students in the teacher's classroom in the baseline year.

References

Braun, H. (2005). *Using Student Progress to Evaluate Teachers: A Primer on Value-Added Models.* Washington, DC: Educational Testing Service.

Braun, H. (2010). *Issues in Measuring Student Growth and Conducting Productivity Analyses.* Washington, DC: Educational Testing Service. Retrieved August 27, 2010, from http://www.k12center.org/rsc/pdf/BraunPresenterSession2.pdf

Braun, H., Chudowsky, N., & Koenig, J. (Eds.). (2010). *Getting Value Out of Value-added: Report of a Workshop.* Washington, DC: The National Academies Press.

Goe, L., Bell, C., & Little, O. (2008). *Approaches to Evaluating Teacher Effectiveness: A Research Synthesis.* Washington, DC: National Comprehensive Center for Teacher Quality.

Goldhaber, D., & Hansen, M. (2008). *Is It Just a Bad Class? Assessing the Stability of Measured Teacher Performance.* CPRE Working Paper 2008-5. Seattle, WA: Center on Reinventing Public Education.

Harris, D. N. (2007). *The Policy Uses and "Policy Validity" of Value-Added and Other Teacher Quality Measures.* San Francisco, CA: Educational Testing Service.

Kane, T. J., & Staiger, D. O. (2008). *Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation.* NBER Working Paper 14607. Cambridge, MA: National Bureau of Economic Research.

Koedel, C., & Betts, J. R. (2009, July). *Does Student Sorting Invalidate Value-Added Models of Teacher Effectiveness? An Extended Analysis of the Rothstein Critique.* Working Paper WP-09-02. Columbia, MO: University of Missouri-Columbia, Department of Economics. Retrieved August 27, 2010, from http://economics.missouri.edu/working-papers/2009/WP0902_koedel.pdf

McCaffrey, D., Sass, T., and Lockwood, J., & Mihaly, K. (2009). *The Intertemporal Stability of Teacher Effect Estimates.* NCPI Working Paper Series No. 2009–03. Nashville, TN: National Center on Performance Incentives.

Meyer, R. (1996). Value Added Indicators of School Performance. *In* E. Hanushek & D. Jorgenson (Eds.), *Improving America's Schools: The Role of Incentives.* Washington, DC: National Academy Press.

Meyer, R. H., & Dokumaci, E. (2010). *Value-added Models and the Next Generation of Assessments*. Washington, DC: Educational Testing Service. Retrieved August 27, 2010, from http://www.k12center.org/rsc/pdf/MeyerDokumaciPresenterSession4.pdf

Meyer, R. H., Dokumaci, E., Geraghty, E. L., McLaughlin, S., & Morgan, E. A. (2009). *Wisconsin Value-Added Model: A Demonstration Project Final Report*. Madison, WI: Wisconsin Center for Education Research, University of Wisconsin–Madison. Retrieved August 27, 2010, from http://dpi.wi.gov/oea/pdf/vademons.pdf

Rivkin, S. G. (2007). *Value-Added Analysis and Education Policy.* National Center for Analysis of Longitudinal Data in Education Research. Retrieved August 27, 2010, from http://urbaninstitute.org/UploadedPDF/411577_value-added_analysis.pdf

Sanders, W., & Horn, S. (1998). Research Findings from the Tennessee Value-Added Assessment System (TVAAS) Database: Implications for Educational Evaluation and Research. *Journal of Personnel Evaluation in Education, 12*(3), 247-256.

References continued

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and Residual Effects of Teachers on Future Student Academic Achievement.* Research Progress Report. Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.

Sass, T. R. (2008). *The Stability of Value-Added Measures of Teacher Quality and Implications for Teacher Compensation Policy.* National Center for Analysis of Longitudinal Data in Education Research. Retrieved August 27, 2010, from *http://www.urban.org/uploadedpdf/1001266_stabilityofvalue.pdf*

Thorn, C. A. (2001). Knowledge Management for Educational Information Systems: What Is the State of the Field? *Education Policy Analysis Archives, 9*(47). *http://epaa.asu.edu/epaa/v9n47/*

Winters, M., Greene, J. P., Ritter, G., & Marsh, R. (2008). *The Effect of Performance-Pay in Little Rock, Arkansas on Student Achievement.* NCPI Working Paper 2008-02. Nashville, TN: National Center on Performance Incentives.

This synthesis of key research studies was written by:

Erin Quinn, Westat.

The Center for Educator Compensation Reform (CECR) was awarded to Westat — in partnership with Learning Point Associates, Synergy Enterprises Inc., Vanderbilt University, 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.

CECR

Center for
Educator Compensation
Reform

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

31047.0910.83670407