D. Measurement

What does research suggest about ways to measure teacher effectiveness so that determination of performance-based rewards is accurate, reliable, and defensible?

Research suggests that there is wide variation in the effectiveness of teachers, yet traditional measures of teacher quality (i.e., education, degrees, and certification status) are not strongly associated with student achievement. This would suggest that the measures of teacher quality most commonly used by school districts are missing the mark in identifying highly effective teachers. Hanushek and colleagues (2005) and Ballou and Podgursky (2000), for example, found that uncertified teachers perform, on average, at roughly the same level as certified teachers. Hanushek and Rivkin (2004), Murnane (1975), and Ehrenberg and Brewer (1994) found that earning a master's degree has negligible impact on teacher effectiveness, with some exceptions for subject-specific degrees.

Related research suggests that the way that traditional pay systems for teachers (according to years of experience and highest degree earned) have little to do with how effective they are at raising student achievement. A growing number of researchers argue that linking teacher pay to student achievement and other measures of teacher productivity could improve the quality of the teaching force and lead to higher levels of student performance (Mohrman, Mohrman, & Odden, 1996; Odden, 2000; Odden & Kelley, 2002).

One common way to measure teachers' effectiveness is by measuring the academic achievement of the students whom they teach. However, if the compensation system bases rewards on the percentage of students who reach a certain level of proficiency or attainment, teachers in the lowest performing schools will be less likely to earn performance awards, even if their students make significant gains during the school year. Performance pay structured in this way creates a powerful disincentive for teachers to work in schools that serve students who are most in need of highly effective teachers.

An alternative way to measure teacher effectiveness is through value-added modeling. A value-added approach that measures growth is preferable to more traditional measures of proficiency or attainment. Value-added modeling rewards teachers according to the amount of academic growth that students make over the course of a school year. This methodology can distinguish teachers who are highly effective at increasing student learning, regardless of students' beginning levels of academic achievement. Sanders' research on value-added assessment in Tennessee was the first to measure teacher effectiveness in this way (Rivers, 1999; Sanders & Horn, 1994, 1998; Sanders & Rivers, 1996).

Further research suggests that teacher impact has a cumulative effect on student learning and that assigning low-performing students to a series of highly effective teachers can raise student test scores dramatically (Mendro, 1998; Rivers, 1999; Sanders & Rivers, 1996).

Research shows that value-added measures are a more accurate gauge of teacher performance than attainment indicators such as average test scores or the percentage of students who reach a predetermined test score or proficiency rate. This is because attainment indicators cannot control for student characteristics or other external influences on growth that a value-added model can include (Braun, 2010; Braun, Chudowsky, & Koenig, 2010; Meyer, 1996; Rivkin, 2007; Sass, 2008).

However, it should be noted that even value-added measures have a certain degree of "noise"—measurement error that reduces the validity of performance distinctions between teachers and schools—that needs to be considered. For example, some researchers have found that teacher-level value added varies considerably from year to year (Goldhaber & Hansen, 2008; Koedel & Betts, 2009; McCaffrey, Sass, & Lockwood, 2008; Sass, 2008). There is a growing consensus that one year of value-added data may not be sufficiently reliable for making high-stakes decisions about individual teacher effectiveness—particularly when determining teacher tenure or hiring/firing decisions (Schochet & Chiang, 2010). When multiple years of value-added data are not available, a viable alternative would be to reward groups of teachers when student performance improves (e.g., all teachers in a school, all teachers in a grade, or all teachers by subject, department, or team).

Designing an accurate, reliable, and defensible teacher compensation system requires that school systems "control for baseline test scores so that teachers are held accountable for their ability to raise achievement, not for students' preexisting knowledge and skills" (Gordon, Kane, & Staiger, 2006). Important factors to control include student and family characteristics, student mobility, and test imperfections (Ballou, Sanders, & Wright, 2004).

Factors such as selection bias and assignment bias can also decrease the accuracy of teacher productivity measures. Selection bias means that students self-select to certain schools based on housing patterns and where families live. Assignment bias means that principals may affect measures of teacher productivity by the way that they assign students to teachers. For example, a principal may assign more difficult students to his or her strongest teachers. Other principals may reward their strongest teachers with better performing, better-behaved students (Rivkin, 2007). However, this would not preclude school systems from using grade-level value-added measures to reward teacher productivity (e.g., rewarding all fourth-grade teachers when fourth-grade reading scores improve).

Accurate, reliable, and defensible teacher compensation systems require states and school districts to collect and track student test data over time. Data systems must be robust and must be capable of tracking student test data for multiple years and matching individual teachers to the students, grades, and subjects that they teach (Thorn, 2002; Thorn, Glover, & Watson, 2007). Some of the unresolved challenges of using value-added models include how to measure the effectiveness of teachers who do not teach tested subjects and grades and how to

assess the effectiveness of teachers in the lower elementary grades because students do not usually take standardized achievement tests before grade 3.

Research suggests that evaluations of classroom performance based on observations may be sufficiently valid measures of teacher effectiveness when value-added assessment cannot be used. Milanowski (2003), for example, found that Cincinnati's teacher evaluation system was fairly accurate at identifying teachers whose students had higher than expected levels of performance, a finding that underscores the value of properly constructed and implemented performance evaluations. (See also Heneman, Milanowski, Kimball, & Odden, 2006.) A more recent study on the Cincinnati teacher evaluation system (TES) by Kane, Taylor, Tyler, and Wooten (2010) produced further evidence that classroom observations can correctly capture teacher performance. They found that teachers' TES scores were positively and non-trivially linked to student achievement.[1]

Research by Jacob and Lefgren (2005) also supports the assertion that evaluators' judgments of teacher performance can be predictive of student achievement. However, observation systems need to be designed and implemented carefully to combat tendencies to rate nearly all teachers at the same level.

In any case, value-added measures of student achievement should not be the sole measures that states and districts use to evaluate teacher productivity. Multiple measures of student achievement and multiple assessments of teacher performance are recommended. A new two-year research study funded by the Gates Foundation will soon shed more light on the qualities that make teachers effective and how to assess them. The Measures of Effective Teaching (MET) project, launched in fall 2009, will analyze videotaped observations of teachers in their classrooms, student and teacher surveys, pedagogical content knowledge tests, and other sources of data to identify indicators of effective teaching. Nearly 4,000 teachers from five or six school districts, including Charlotte-Mecklenburg, NC, and New York City, will participate in the research study. The Gates Foundation expects to release its findings—along with suggested measurement instruments—in the summer of 2011 (Gates Foundation, 2010).

[1] The TES rubric has four ratings: unsatisfactory, basic, proficient, and distinguished. Kane et al. (2010) found that moving up one ranking (e.g., from basic to proficient) resulted in a student achievement gain of about one-sixth to one-fifth of a standard deviation.

References

Ballou, D., & Podgursky, M. (2000). Reforming teacher preparation and licensing: What is the evidence? *Teachers College Record, 102*(1), 5–27.

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics, 29*(1), 37–65.

Braun, H. (2010). *Issues in measuring student growth and conducting productivity analyses.* Washington, DC: Educational Testing Service. Retrieved Aug. 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.

Ehrenberg, R., & Brewer, D. (1994). Do school and teacher characteristics matter? Evidence from high school and beyond. *Economics of Education Review, 13*(1), 1–17.

Gates Foundation, The. (2010, June). *Working with teachers to develop fair and reliable measures of effective teaching.* Seattle, WA: The Gates Foundation. Retrieved Dec. 11, 2010, from http://www.gatesfoundation.org/highschools/Documents/met-framing-paper.pdf

Goldhaber, D., & Hansen, M. (2008). *Is it just a bad class? Assessing the stability of measured teacher performance.* Center on Reinventing Public Education Working Paper 2008-5. Seattle, WA: University of Washington, Center on Reinventing Public Education.

Gordon, R., Kane, T., & Staiger, D. (2006). *Identifying effective teachers using performance on the job.* Washington, DC: The Brookings Institution.

Hanushek, E., & Rivkin, S. (2004). How to improve the supply of high quality teachers (pp. 7–25). *In* D. Ravitch (Ed.), *Brookings papers on education policy*. Washington, DC: The Brookings Institution.

Hanushek, E., Kain, J., O'Brien, D., & Rivkin, S. (2005). *The market for teacher quality*. NBER Working Paper 11154. Cambridge, MA: National Bureau of Economic Research.

Heneman, H.G. III, Milanowski, A., Kimball, S. M., & Odden, A. (2006). *Standards-based teacher evaluation as a foundation for knowledge- and skill-based pay.* CPRE Policy Brief RB-45. Philadelphia: Consortium for Policy Research in Education.

Jacob, B. A., & Lefgren, L. (2005). *Principals as agents: Subjective performance measurement in education. NBER* Working Paper No. 11463. Cambridge, MA: National Bureau of Economic Research.

Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2010). *Identifying effective classroom practices using student achievement data* (NBER Working Paper 15803). Cambridge, MA: National Bureau of Economic Research. Retrieved January 11, 2011, from http://www.nber.org/papers/ w15803.pdf

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. http://economics.missouri.edu/working-papers/2009/WP0902_koedel.pdf

McCaffrey, D. F., Sass, T. R., & Lockwood, J. R. (2008). *The intertemporal stability of teacher effect estimates.* Nashville, TN: National Center on Performance Incentives. http://www.performanceincentives.org/data/files/news/PapersNews/McCaffrey_et_al_2008.pdf

Mendro, R. (1998). Student achievement and school and teacher accountability. *Journal of Personnel Evaluation in Education, 12*(3), 257–267.

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.

Milanowski, A. (2003). *The criterion-related validity of the performance assessment system in Cincinnati.* Madison, WI: University of Wisconsin, Consortium for Policy Research in Education.

Mohrman, A., Mohrman, S., & Odden, A. (1996). Aligning teacher compensation with systemic school reform: Skill based pay and group-based performance awards. *Educational Evaluation and Policy Analysis, 18 (1),* 51–71.

Murnane, R. (1975). *Impact of school resources on the learning of inner city children*. Cambridge, MA: Ballinger.

Odden, A. (2000). New and better forms of teacher compensation are possible. *Phi Delta Kappan, 81*(5), 361–366.

Odden, A., & Kelley, C. (2002). *Paying teachers for what they know and do: New and smarter compensation strategies to improve schools.* Thousand Oaks, CA: Corwin Press.

Rivers, J. (1999). *The impact of teacher effect on student math competency achievement*. Unpublished doctoral dissertation, University of Tennessee, Knoxville, TN.

Rivkin, S. (2007). *Value-added analysis and education policy*. Brief 1. Washington, DC: National Center for Analysis of Longitudinal Data in Education Research, The Urban Institute.

Sanders, W., & Horn, S. (1994). The Tennessee value-added assessment system: Mixed methodology in educational assessment. *Journal of Personnel Evaluation in Education, 8*(3), 299–311.

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.

Sanders, W., & Rivers, J. (1996). *Cumulative and residual effects of teachers on future students' academic achievement*. Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.

Sass, T.R. (2008, November). *The stability of value-added measures of teacher quality and implications for teacher compensation policy.* Research Brief. Washington, DC: National Center for Analysis of Longitudinal Data in Education Research, The Urban Institute. Retrieved from http://www.urban.org/UploadedPDF/1001266_stabilityofvalue.pdf

Schochet, P. Z., & Chiang, H. S. (2010). Error rates in measuring teacher and school performance based on student test score gains (NCEE 2010-4004). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

References continued

Thorn, C. (2002). *Data use in the school and classroom: The challenges of implementing data-based decision making inside schools*. Madison, WI: Wisconsin Center for Education Research, University of Wisconsin-Madison.

Thorn, C., Glover, R., & Watson, J. (2007). *Information technology considerations*. Washington, DC: Center for Educator Compensation Reform, Office of Elementary and Secondary Education, U.S. Department of Education.

This synthesis of key research studies was written by:

Cynthia D. Prince, Vanderbilt University; Julia Koppich, Ph.D., J. Koppich and Associates; Tamara Morse Azar, Westat; Monica Bhatt, Learning Point Associates; and Peter J. Witham, Vanderbilt University.

We are grateful to Michael Podgursky, University of Missouri, and Anthony Milanowski, University of Wisconsin-Madison, for their helpful comments and suggestions.

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

32074.0210.83670507