

Methods of Evaluating Teacher Effectiveness

Now that nearly all teachers are meeting the criteria to be considered “highly qualified,” policy conversations are turning to issues of teacher effectiveness. Ensuring that teachers meet the federal requirements to be considered highly qualified is the foundation upon which teaching and learning is built. The next step is determining whether teachers are providing instruction in ways that will lead to high levels of student achievement (i.e., teacher effectiveness).

LEARNING FROM RESEARCH

In *Approaches to Evaluating Teacher Effectiveness: A Research Synthesis* published by the National Comprehensive Center for Teacher Quality (TQ Center), ways of evaluating teachers were compared (Goe, Bell, & Little, 2008). This brief compares two of the methods discussed in that synthesis—value-added measures and classroom observations—and discusses the advantages and drawbacks associated with these methods. Although classroom observations have been commonly used for evaluating teachers for many decades, value-added models are becoming an increasingly popular method of determining teacher effectiveness.

In This Brief

This brief is intended to help regional centers and state policymakers as they consider evaluation methods to clarify policy, develop new strategies, identify effective teachers, or guide and support districts in selecting and using appropriate evaluation methods for various purposes.

Authors

This brief was written by Laura Goe, Ph.D., and Andrew Croft of ETS.

Promoting Students' Academic Achievement

Ensuring that students are achieving in tested subjects in tested grades is only part of what effective teachers do on the job. Promoting students' academic achievement is arguably the most important component of their jobs, but teachers contribute to their students' development in myriad ways. For example, teachers help students learn to work cooperatively with peers; conduct themselves appropriately in classrooms and schools; resolve differences peacefully; and understand their roles as citizens in classrooms, schools, communities, and society as a whole. Teachers also have responsibilities beyond direct instruction, such as working with colleagues to identify at-risk students and develop plans to support them.

Teachers contribute significantly to the establishment and maintenance of supportive, learning-centered environments in their classrooms and schools and work with parents and the community to support educational opportunity and success. Moreover, their relationships within schools (e.g., mentoring new teachers, serving on curriculum committees, providing leadership for extracurricular activities) may not directly impact student learning, but they create an environment conducive to successful teaching and learning. See Goe, Bell, and Little's (2008) five-point definition of effective teachers to learn more about the key responsibilities of effective teachers.

Goe et al. (2008) developed a five-point definition of teacher effectiveness by analyzing research, policy, and standards that addressed teacher effectiveness. After the definition had been developed, Goe et al. (2008) consulted a number of experts and strengthened the definition based on their feedback.

"The five-point definition of teacher effectiveness consists of the following:

- Effective teachers have high expectations for all students and help students learn, as measured by value-added or other test-based growth measures, or by alternative measures.
- Effective teachers contribute to positive academic, attitudinal, and social outcomes for students such as regular attendance, on-time promotion to the next grade, on-time graduation, self-efficacy, and cooperative behavior.
- Effective teachers use diverse resources to plan and structure engaging learning opportunities; monitor student progress formatively, adapting instruction as needed; and evaluate learning using multiple sources of evidence.
- Effective teachers contribute to the development of classrooms and schools that value diversity and civic-mindedness.
- Effective teachers collaborate with other teachers, administrators, parents, and education professionals to ensure student success, particularly the success of students with special needs and those at high risk for failure" (Goe et al., 2008, p. 8).

Measuring Teacher Effectiveness

Given the broad manner in which teacher effectiveness can be defined, it is not surprising that multiple methods for evaluating teachers exist. These include principal evaluations; analyses of classroom artifacts (i.e., ratings of teacher assignments and student work); teaching portfolios; teacher self-reports of practice, including surveys, teaching logs, and interviews; and student ratings of teacher performance. Although these various methods have their functions and appropriate uses, this brief will focus on two of the most widely used measures of teacher effectiveness: value-added models and classroom observation.

Both types of measures focus primarily on teachers' contributions to student learning but with very different lenses. Value-added measures can be defined as "a collection of complex statistical techniques that use multiple years of students' test score data to estimate the effects of individual schools or teachers" (McCaffrey, Lockwood, Koretz, & Hamilton, 2003, p. xi).

William Sanders is credited with developing value-added modeling for evaluating teachers, using it in Tennessee to determine that students in some teachers' classrooms were scoring higher than their previous test scores would have predicted (Sanders & Rivers, 1996).

Observation measures capture additional information about the specific strategies teachers use in their classroom, and they can be used for formative purposes, providing direction for teachers to strengthen their practice in specific areas. For example, results from a standards-based observation can help build teachers' awareness of their most successful teaching approaches and areas in which there is room for growth.

Requirements for Using Value-Added Measures to Evaluate Individual Teachers

- Student achievement test scores must be linked to individual teachers.
- All value-added models require students' achievement scores prior to the year for which the teachers' scores are being calculated, though models vary on how many years worth of scores are needed for an accurate prediction.
- Some models include students' gender, race, and socioeconomic background.
- Some models include information about teachers' experience.

What the Research Says

- The scores cannot be solely attributed to teachers' influence. Value-added measures are believed to provide a summary score of the "contribution of various factors toward growth in student achievement" (Goldhaber & Anthony, 2003, p. 38).
- Strong, consistent correlations between what teachers do in their classrooms (measured by observations) and value-added scores are not apparent (Kimball, White, Milanowski, & Borman, 2004).
- Researchers found that the majority of teacher effectiveness could not be explained by observable teacher characteristics. Teachers vary in their contribution to students' achievement score gains, but researchers have not been able to identify the cause of this variation (Rivkin, Hanushek, & Kain, 2005).
- Valued-added scores cannot be calculated for most teachers in a district or state because they teach subjects that are not tested or teach in lower elementary grades for which prior test scores are not available.

Value-Added Measures

Supporters of using value-added measures of teacher effectiveness contend that such models can accurately rank teachers within a district by their contributions to student learning. Value-added measures can indicate the following:

- That students of a particular teacher performed better than their previous achievement would have predicted
- Whether certain teachers' students consistently perform above or below predicted levels on standardized achievement tests

Teacher effectiveness rankings are calculated based on whether students meet, exceed, or fail to reach their predicted scores on the test. Teachers are compared with other teachers within their district. Rankings can be calculated only for teachers who have students with standardized test scores (usually mathematics and reading/language arts teachers). If a teacher's students perform better than predicted on standardized achievement tests, the teacher is credited with being effective, but if most of his or her students fail to make predicted gains, the teacher may be deemed less effective.

Value-added modeling is complex, and many experts urge caution in using the results for evaluating teacher effectiveness (e.g., Bracey, 2004; Braun, 2005; Kupermintz, 2003; McCaffrey, Koretz, Lockwood, & Hamilton, 2004; Thum, 2003). Because teachers are not randomly assigned to schools, and students are not randomly assigned to teachers, it is difficult to sort out how much student achievement growth is attributable solely to teachers' efforts and how much is attributable to other factors not included in the statistical model.

Classroom Observations

Classroom observations are the most common form of teacher evaluation and vary widely in how they are conducted and what they evaluate. Observations can be created by the district or purchased as a product. They can be conducted by a school administrator or an outside evaluator. They can measure general teaching practices or subject-specific techniques. They can be formally scheduled or unannounced and can occur once or several times per year. The type of observation method adopted, its focus, and its frequency should depend on what the administration would like to learn from the process.

Classroom observations provide a useful measure of teachers' practice but little evidence about whether students are actually learning. However, if the observation instruments are based on valid standards of effective teaching practice, they can be used as a source of evidence about individual teachers' effectiveness. The degree to which observations can or should be used for specific purposes depends on the instrument, how that instrument was developed, the level of training and monitoring raters receive, and the psychometric properties of the instrument.

The following conditions should be in place prior to using a classroom observation measure for evaluation:

- Use a high-quality observation instrument based on standards of effective teaching practice that include levels of performance.
- Allow teachers time and opportunity to familiarize themselves with the observation instrument so that they will understand what is expected.
- Train observers to use the instrument so that all observers are using it in the same way. The goal is to ensure that a teacher gets the same score no matter which rater conducts the observation. Furthermore, avoid potential rater bias (or the appearance of bias) by using trained raters.
- Calibrate observers. Calibration involves checking the scores of observers to ensure that they are not getting more stringent or lax in scoring over time, a condition called "rater drift."
- When the stakes are high, conduct multiple observations, preferably with different observers.
- For elementary teachers and other teachers of more than one subject, observing when they are teaching different subjects will help identify subject-specific strengths and weaknesses.
- Share ratings with the teachers, preferably as part of an individual development plan.

Different Evaluation Methods for Different Purposes

How should teacher effectiveness be evaluated? Table 1 provides a brief comparison of the advantages and disadvantages of value-added and classroom observation measures.

There are many different reasons for evaluating teacher effectiveness, and many different consequences are attached to those evaluations. The reasons and consequences should be clearly established before deciding upon appropriate

methods and instruments. Table 2 presents some of the purposes of evaluating teachers, along with methods that would be useful for collecting appropriate evidence. For further information on all types of evaluation methods mentioned in Table 2, see *Approaches to Evaluating Teacher Effectiveness: A Research Synthesis* (Goe et al., 2008).

Table 1. A Comparison of Value-Added Measures and Classroom Observation for Teacher Evaluation

	Advantages	Disadvantages
Value-Added Measures	<ul style="list-style-type: none"> • Relatively inexpensive (after initial infrastructure costs) • Focuses solely and directly on student learning • Relatively objective • Comparable across schools, districts, and even states (if they are using the same statistical methods and achievement tests) 	<ul style="list-style-type: none"> • Costly to build necessary data system; generally requires hiring experts to set it up and conduct the analyses • No information about what effective teachers do in the classroom • No information to help “bad” teachers improve • No information for some teachers (e.g., special education, art, music, early elementary)
Classroom Observation	<ul style="list-style-type: none"> • High face-validity and teacher buy-in • Allows teachers to understand and participate in the evaluation process • Useful for formative evaluation, particularly for novice teachers • Based on “best practices” 	<ul style="list-style-type: none"> • Costly due to personnel costs • May not take student achievement into account • Scores determined by evaluators with different levels of training • May be affected by whether measures are used for high-stakes or low-stakes evaluation

Table 2. Evaluation Purposes and Methods

Purpose	Value-Added	Classroom Observation	Interviews, Surveys	Administrative Judgment
Find out whether grade-level or instructional teams are meeting specific achievement goals.	X			
Determine whether a teacher's students are meeting achievement growth expectations.	X			
Establish whether a new teacher is meeting performance expectations in the classroom.		X		
Gather information in order to provide new teachers with guidance related to identified strengths and shortcomings.		X		
Examine the effectiveness of teachers in nonacademic subjects (e.g., art, music, and physical education).		X		
Examine the effectiveness of teachers in lower elementary grades for which no test scores from previous years are available to predict student achievement (required for value-added models).		X		
Determine the types of assistance and support a struggling teacher may need.		X		
Gather information to determine what professional development opportunities are needed for individual teachers, instructional teams, grade-level teams, etc.	X	X	X	X
Gather evidence for making contract renewal and tenure decisions.	X	X	X	X
Determine whether a teacher's performance qualifies him or her for additional compensation or incentive pay (rewards).	X	X	X	X
Gather information on a teacher's ability to work collaboratively with colleagues to evaluate the needs of and determine appropriate instruction for at-risk or struggling students.				X
Establish whether a teacher is effectively communicating with parents/guardians.				X
Determine how students and parents perceive a teacher's instructional efforts.				X
Determine who would qualify to become a mentor, coach, or teacher leader.	X	X	X	X

Note. "X" indicates appropriate measures for the specified purpose.

Creating a Strong Evaluation System

Some states have statewide policies for teacher evaluation, whereas others allow districts to establish their own policies. Even when districts establish their own policies, state policymakers are often called upon to make recommendations. State policymakers should consider the following steps when creating or advising districts in creating an evaluation system or revamping an existing system:

- Involve teachers and stakeholders in developing the evaluation system.
 - Involvement increases teacher/stakeholder buy-in and validity of the system.
- Consider different teaching contexts and how the evaluation system will accommodate them.
 - Early elementary teachers cannot be evaluated with value-added models.
 - Nontested subjects cannot be evaluated with value-added models.
- Start with an instrument that is already valid and reliable, and adapt it if necessary.
 - Keep adaptations to a minimum because the instrument was validated as a whole—not in pieces.
- Use multiple indicators, not just an observation score.
 - There are many other important things you can measure economically (see the five-point definition of effective teachers on p. 2).
 - Use appropriate weights to give more importance to the most significant components of the system (e.g., on-time graduation may be weighted for secondary teachers and not weighted for elementary teachers).
- Set aside funds to support training and calibrating of observers.
- Measure what is most important to you, your administrators, your teachers, and other education stakeholders.
 - The system will drive improvement as teachers strive to improve in areas they know will be measured as part of the evaluation.
 - Ensure that what teachers are striving for is truly important in your definition of successful teaching.
- Give teachers opportunities to improve in the areas in which they score poorly.
 - Provide assistance in determining problem areas and planning strategies to address them.
- Differentiate among teachers.
 - Standards may be the same, but progress toward those standards should be compared with other similar teachers (e.g., lower elementary teachers may be evaluated with different rubrics than those used for evaluating secondary teachers, and acceptable performance for novice teachers may be at a lower level on the rubric when compared with experienced teachers).
- For high-stakes decision making, devise a system that involves multiple observations and multiple raters during the course of the year.
- For systems including measures of student achievement (value-added measures), establish whether the state's current testing system is valid for the purpose of conducting value-added analysis, and ensure that longitudinal linked student-teacher data is sufficient for conducting value-added analysis (i.e., data are accurate, and there are little "missing" data).

A SAMPLE OF EXISTING EVALUATION SYSTEMS

- The Beginning Educator Support and Training Program (BEST) (<http://www.ctbest.org>) [This Connecticut program is currently being revamped due to new legislation (see http://24.248.88.133/Resources/2008_BEST_C1.htm).]
- Delaware Performance Appraisal System (<http://www.doe.k12.de.us/performance/dpasii/default.shtml>)
- Florida District Performance Appraisal System Checklist (<http://www.fldoe.org/profdev/pa.asp>)
- Minnesota Q-Comp—Quality Teacher Compensation, Part of the National Institute for Excellence in Teaching (http://cfl.state.mn.us/MDE/Teacher_Support/QComp/index.html)
- New Mexico Evaluation Guidelines (http://www.teachnm.org/annual_assessment.html)
- North Carolina Public School Employee Evaluation Standards and Instruments (<http://www.ncpublicschools.org/fbs/personnel/evaluation/>)
- Ohio Value-Added Support (<http://portal.battelleforkids.org/Ohio/home.html?sflang=en>)
- South Carolina Performance Appraisal System (ADEPT) (<http://www.scteachers.org/ADEPT/index.cfm>)
- Ten Indicators of a Quality Teacher Evaluation Plan (<http://www.sde.ct.gov/sde/cwp/view.asp?a=2641&q=320432>)
- Tennessee Framework for Evaluation and Professional Growth Guidelines and Manuals (<http://www.state.tn.us/education/frameval/>)
- Wisconsin Master Educator Assessment Process and the Master Educator License (<http://dpi.wi.gov/tepd/wmeapsumm.html>)

Summary

Given that classroom observations and value-added measures have different strengths and weaknesses, the reason for the evaluation should be carefully considered before selecting the method. In addition, what to do with the results of the evaluation should

be determined in advance. Value-added measures can provide useful information; however, they provide little guidance for teachers who want to improve their practice. If the goal is to improve teacher practice, classroom observations may be more useful.

TQ CENTER RESOURCES

For more information on these and other measures of teacher quality and effectiveness, please see the following TQ Center reports, tools, and briefs:

Coggshall, J., Max, J., & Bassett, K. (2008). *Using performance-based assessment to identify and support high-quality teachers*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved March 3, 2009, from <http://www.tqsource.org/publications/keyIssue-June2008.pdf>

Goe, L. (2008). *Using value-added models to identify and support highly effective teachers*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved March 3, 2009, from <http://www2.tqsource.org/strategies/het/UsingValueAddedModels.pdf>

Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved March 3, 2009, from <http://www.tqsource.org/publications/EvaluatingTeachEffectiveness.pdf>

References

- Bracey, G. W. (2004). Value-added assessment findings: Poor kids get poor teachers. *Phi Delta Kappan*, 86(4), 331–333.
- Braun, H. I. (2005). *Using student progress to evaluate teachers: A primer on value-added models*. Princeton, NJ: ETS. Retrieved March 3, 2009, from <http://www.ets.org/Media/Research/pdf/PICVAM.pdf>
- Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved March 3, 2009, from <http://www2.tqsource.org/strategies/het/UsingValueAddedModels.pdf>
- Goldhaber, D., & Anthony, E. (2003). *Teacher quality and student achievement* (No. UDS-115). New York: ERIC Clearinghouse on Urban Education.
- Kimball, S. M., White, B., Milanowski, A. T., & Borman, G. (2004). Examining the relationship between teacher evaluation and student assessment results in Washoe County. *Peabody Journal of Education*, 79(4), 54–78.
- Kupermintz, H. (2003). Teacher effects and teacher effectiveness: A validity investigation of the Tennessee Value Added Assessment System. *Educational Evaluation and Policy Analysis*, 25(3), 287–298.
- McCaffrey, D. F., Koretz, D., Lockwood, J. R., & Hamilton, L. S. (2004). *The promise and peril of using value-added modeling to measure teacher effectiveness* (Research Brief No. RB-9050-EDU). Santa Monica, CA: RAND.
- McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability*. Santa Monica, CA: RAND. Retrieved March 3, 2009, from http://www.rand.org/pubs/monographs/2004/RAND_MG158.pdf
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417–458. Retrieved March 3, 2009, from <http://edpro.stanford.edu/Hanushek/admin/pages/files/uploads/teachers.econometrica.pdf>
- Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement* (No. R11-0435-02-001-97). Knoxville: University of Tennessee Value-Added Research and Assessment Center.
- Thum, Y. M. (2003). Measuring progress toward a goal: Estimating teacher productivity using a multivariate multilevel model for value-added analysis. *Sociological Methods & Research*, 32(2), 153–207. Retrieved March 3, 2009, from <https://www.msu.edu/~thum/Papers/SMR1103.pdf>

ABOUT THE NATIONAL COMPREHENSIVE CENTER FOR TEACHER QUALITY

The National Comprehensive Center for Teacher Quality (TQ Center) was created to serve as the national resource to which the regional comprehensive centers, states, and other education stakeholders turn for strengthening the quality of teaching—especially in high-poverty, low-performing, and hard-to-staff schools—and for finding guidance in addressing specific needs, thereby ensuring that highly qualified teachers are serving students with special needs.

The TQ Center is funded by the U.S. Department of Education and is a collaborative effort of ETS, Learning Point Associates, and Vanderbilt University. Integral to the TQ Center's charge is the provision of timely and relevant resources to build the capacity of regional comprehensive centers and states to effectively implement state policy and practice by ensuring that all teachers meet the federal teacher requirements of the No Child Left Behind (NCLB) Act.

The TQ Center is part of the U.S. Department of Education's Comprehensive Centers program, which includes 16 regional comprehensive centers that provide technical assistance to states within a specified boundary and five content centers that provide expert assistance to benefit states and districts nationwide on key issues related to the NCLB Act.



1100 17th Street NW, Suite 500
Washington, DC 20036-4632
877-322-8700 • 202-223-6690
www.tqsource.org

Copyright © 2009 National Comprehensive Center for Teacher Quality, sponsored under government cooperative agreement number S283B050051. All rights reserved.

This work was originally produced in whole or in part by the National Comprehensive Center for Teacher Quality with funds from the U.S. Department of Education under cooperative agreement number S283B050051. The content does not necessarily reflect the position or policy of the Department of Education, nor does mention or visual representation of trade names, commercial products, or organizations imply endorsement by the federal government.

The National Comprehensive Center for Teacher Quality is a collaborative effort of ETS, Learning Point Associates, and Vanderbilt University.