Title: Options for measuring teacher effectiveness

Date: June 2010

Question: Could you provide a summary of the methods that are available for measuring teacher effectiveness and the associated tradeoffs?

Response:

This response addresses the following questions:

- What are some known methods for measuring teacher effectiveness that could be used in teacher evaluation systems?
- What do these methods afford in regards to reliability and validity; acceptability to stakeholders; cost; and usefulness to teachers and administrators?
- How could these methods be used as part of a teacher evaluation system?
- What are some examples of comprehensive teacher evaluation systems around the nation?
- What innovative teacher evaluation practices exist within California school districts?

In addressing the first three of these questions, this response draws significantly, though not exclusively, on two key reports from the National Comprehensive Center for Teacher Quality: *Approaches to Evaluating Teacher Effectiveness: A Research Synthesis* (2008), by L. Goe, C. Bell, and O. Little, and *A Practical Guide to Evaluating Teacher Effectiveness* (2009), by O. Little, L. Goe, and C. Bell.

1. **What are some known methods for measuring teacher effectiveness that could be used in teacher evaluation systems?**

Depending on its purpose or purposes, as well as other factors, a teacher evaluation system may combine multiple methods for measuring teacher effectiveness. Researchers Little et al. (2009) state that "a well-conceived system should combine approaches to gain the most complete understanding of teaching and that administrators and teachers should work together to create a system that supports teachers as well as evaluates them." Below are descriptions of six methods reviewed in Little et al., along with links to more information about each method on the National Comprehensive Center for Teacher Quality (NCCTQ) website.

Value-added models are statistical models used to gauge a teacher's contributions to students' test score gains. Students' prior achievement on standardized tests may be used to predict their achievement in a specific subject the next year. Thus, when the student

scores of a particular class improve on standardized achievement tests beyond what would be expected based on their past performance, the comparison of actual growth to expected growth is considered to be the *value added*. For more information, visit http://www3.learningpt.org/tqsource/GEP/GEPEvalType.aspx?tid=6.

Classroom observation enables an evaluator to measure discernable classroom processes, including general teaching practices, subject-specific techniques, or both. They can be conducted by a school administrator, peer evaluator, or outside evaluator.[1] Generally, classroom observations conducted by an administrator are used for summative purposes, most commonly for tenure or dismissal decisions for beginning teachers. For more information, visit http://www3.learningpt.org/tqsource/gep/GEPEvalType.aspx?tid=1.

Analysis of classroom artifacts can help an evaluator determine the quality of classroom instruction through consideration of a set of artifacts that may include lesson plans, assignments, assessments, scoring rubrics, and student work. Although the examination of teacher lesson plans or student work is often part of teacher evaluation procedures, few systems use a validated protocol to measure the quality of instruction and the process would benefit from one. Examples of a protocol for analyzing classroom artifacts are the Intellectual Demand Assignment Protocol (IDAP)[2] or the protocol used in the Instructional Quality Assessment (IQA) toolkit.[3] For more information, visit http://www3.learningpt.org/tqsource/gep/GEPEvalType.aspx?tid=2.

Teacher portfolio is similar to an analysis of artifacts but differs in two ways: 1) it is the teacher, rather than the evaluator, who assembles the collection of materials for a portfolio, and 2) the materials in the portfolio are intended to capture a broader understanding of teacher effectiveness, serving as evidence of teaching practices, contributions to school community, and student progress. A portfolio may contain, for example, teacher lesson plans, schedules, assignments, assessments, student work, teacher reflection, videotape of classroom instruction, family contact logs, notes from parents, documentation of professional development, and special awards or recognition. In addition to being used for evaluation, portfolios often are designed to promote teacher reflection and improvement, requiring teachers to justify their selections and how their artifacts relate to the standards. Two programs that use portfolios for evaluation are the National Board for Professional Teaching Standards (NBPTS) certification program and Connecticut's Beginning Educator Support and Training (BEST). For more information, visit NCCTQ at http://www3.learningpt.org/tqsource/gep/GEPEvalType.aspx?tid=3.

Self-report of practice is an evaluation method that asks teachers to report on what they are doing in the classroom. The reporting may take the form of surveys, instructional logs,

---

[1] *Enhancing Teacher Evaluation: Effective Practices for Evaluating Teachers of All Students.* National Comprehensive Center for Teacher Quality Webcast on April 14, 2010. Experts Lynn Holdheide of Vanderbilt University and Andrew Croft of ETS suggest that, when evaluating teachers of students with disabilities and English language learners, evaluators should be knowledgeable about the pedagogy and practices of teaching these at-risk populations.
[2] Developed by the Consortium on Chicago School Research.
[3] Developed by UCLA's National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

interviews, or some combination of the above. Depending on its purpose, self-reporting may be targeted to gather information on a specific pedagogy or content area or may be focused on broad, overarching aspects of teaching. In reviewing a teacher's self-report, an evaluator may use a checklist of observable behaviors and practices, possibly including rating scales to assess the extent to which certain teaching practices are used, or may require the teacher to record an exact count of how frequently one uses a particular practice or standard. For more information, visit http://www3.learningpt.org/tqsource/gep/GEPEvalType.aspx?tid=4.

Student evaluation of teachers is most often conducted through a questionnaire that asks students to rate teachers on a Likert-type scale. Students may be asked to evaluate teachers on various aspects of teaching, from course content to teaching practices and behaviors. For more information, visit http://www3.learningpt.org/tqsource/gep/GEPEvalType.aspx?tid=5.

Collaboration/contribution to school community

In our review of the methods in preparation of this memo, we were able to identify three teacher evaluation rubrics used by either a state or district that include elements related to teachers' collaboration (i.e., impact on grade level, department, school team[4]) or their contribution to a school community. The rubrics are the Georgia CLASS Keys Standards,[5] Cincinnati Public Schools Standards & Rubrics,[6] and Chicago's Excellence for Teaching — Form 1 CPS Framework for Teaching.[7] All three trace their roots to Charlotte Danielson's *Framework for Teaching*,[8] in which Danielson presents four major domains of teacher responsibilities. One of them, *Teacher Professionalism,* addresses both teacher collaboration and professional contributions. Kane and colleagues (2010) note that the evidence for measuring practices in this fourth domain (e.g., teacher collaboration and contribution) is not observable during a classroom observation, which is when rubrics such as the three listed above are used to measure teacher effectiveness. Rather, to see what a teacher is doing in these two areas of professional responsibility, an evaluator must review documents, such as a professional learning log or family communication log, submitted by the teacher, as is done in Cincinnati's Teacher Evaluation System.

2. **What do these methods afford in regards to reliability and validity; acceptability to stakeholders; cost; and usefulness to teachers and administrators?**

For each evaluation method, issues of reliability and validity, acceptability to stakeholders, cost, and usefulness to teachers and administrators are addressed in the matrix below.

---

[4] Examples of teacher collaboration as described in the *LAUSD Teacher Effectiveness Task Force Final Report,* April 13, 2010.
[5] Georgia CLASS Keys Standards/http:www.doe.k12.ga.us/tss_teacher.aspx. Select *Class Keys Standards* to view and download the rubric.
[6] Cincinnati Public Schools Standards & Rubrics/http://www.cps-k12.org/employment/tchreval/stndsrubrics.pdf.
[7] Chicago's Excellence for Teaching - Form 1 CPS Framework for Teaching/
http://www.chicagoteacherexcellence.org/index.php?tray=topic&tid=gptop1&cid=1cps3.
[8] Charlotte Danielson's Framework for Teaching can be viewed at http://www.danielsongroup.org/theframeteach.htm.

Table 1. Summary of teacher evaluation methods

Method	Reliability and validity	Acceptability to stakeholders	Cost	Usefulness to teachers and administrators
Value-added models	Little is known about the validity of value-added scores for identifying effective teaching, though research using value-added models suggests that teachers differ markedly in their contributions to students' test score gains. However, correlating value-added scores with teacher qualifications, characteristics, or practices have yielded mixed results and few significant findings. Teachers vary in effectiveness, but research has not determined why. Vertical test alignment is assumed (i.e., tests are measuring essentially the same thing from grade to grade).	Value-added measures are controversial because they measure *only* teachers' contributions to student achievement gains on standardized tests. Entails little burden at the classroom or school level because most data are already collected for NCLB purposes.	Requires no classroom visits because linked student/teacher data can be analyzed at a distance.	Provides a way to evaluate teachers on their *contribution* to student learning, which most measures do not. Value-added scores are not useful for formative purposes because teachers learn nothing about *how* their practices contributed to (or impeded) student learning. Models are not able to sort out *teacher* effects from *classroom* effects. May be useful for identifying outstanding teachers whose classrooms can serve as "learning labs," as well as struggling teachers in need of support.
Classroom observation	Some highly researched protocols have been linked to student achievement, though associations are sometimes modest. Choosing or creating a valid and reliable protocol as well as training and calibrating raters are essential to obtaining valid results. Research and validity findings are highly dependent on the instrument used, sampling procedures, and the training of raters.	Is credible—generally considered a fair and direct measure by stakeholders.	Expensive due to cost of observers' time; intensive training and calibrating of observers adds to expense, but are necessary for validity.	Provides rich information about classroom behaviors and activities. Depending on the protocol, can be used in various subjects, grades, and contexts. Generally is feasible and can be one useful component in a system used to make summative judgments and provide formative feedback. Assesses observable classroom behaviors, but not as useful for assessing beliefs, feelings, intentions, or out-of-classroom activities.

Method	Reliability and validity	Acceptability to stakeholders	Cost	Usefulness to teachers and administrators
Classroom observation (continued)	There is a lack of research on observation protocols as used in context for teacher evaluation. Evaluation instruments used without proper training or regard for their intended purpose will impair validity. Principals may not be qualified to evaluate teachers on measures highly specialized for certain subjects or contexts			
Analysis of classroom artifacts	Pilot research has linked artifact ratings to observed measures of practice, quality of student work, and student achievement gains. More research is needed to establish scoring reliability and to determine the ideal amount of work to sample. Lack of research exists on the use of structured artifact analysis in practice. More validity and reliability research is needed. Artifacts analysis may be a compromise in terms of feasibility and validity between full observation and less direct measures, such as self-report.	Unknown.	Training knowledgeable scorers can be costly, but is necessary to ensure validity. Is practical and feasible because artifacts have already been created for the classroom.	Can be a useful measure of instructional quality if a validated protocol is used, if raters are well trained for reliability, and if assignments show sufficient variation in quality.

Method	Reliability and validity	Acceptability to stakeholders	Cost	Usefulness to teachers and administrators
Teacher portfolios	Research on validity and reliability for portfolios is ongoing, and concerns have been raised about consistency of scoring. Stability of scores may not be high enough to use for high-stakes assessment. There is a lack of research linking portfolios to observed changes in teaching practice or student achievement. Some studies have linked NBPTS certification (which includes a portfolio) to student achievement, but other studies have found no relationship. Scorers should have content knowledge of the portfolios they score. Portfolios are difficult to standardize (compare across teachers or schools).	Has a high level of credibility among stakeholders.	This measure is time-consuming for teachers and scorers.	Is comprehensive; can measure aspects of teaching that are not readily observable in the classroom. Portfolios represent teachers' exemplary work but may not reflect everyday classroom activities. Can be used with teachers of all content areas, grade levels, or student populations. Is a useful tool for teacher reflection and improvement.
Self-report of practice	Studies on the validity of teacher self-report measures present mixed results. Highly detailed measures of practice may be better able to capture actual teaching practices but may be more difficult to establish reliability or may result in narrowly focused measures. Reliability and validity of self-report has not been fully established and depends on the instrument used. This measure should not be used as the sole or primary measure in teacher evaluation.	Unknown.	Is feasible and cost-efficient; can collect large amounts of information at once [through a survey]. Using or creating a well-developed and validated instrument will decrease cost-efficiency but will increase accuracy of findings.	Can measure unobservable factors that may affect teaching, such as knowledge, intentions, expectations, and beliefs. Provides the unique perspective of the teacher.

Method	Reliability and validity	Acceptability to stakeholders	Cost	Usefulness to teachers and administrators
Student evaluation	Several studies show that student ratings of teachers may be as valid as judgments made by college students and other groups and, in some cases, may correlate with measures of student achievement; thus, students can provide useful information about teaching. Validity is dependent on the instrument used and its administration and is generally recommended for formative use only. Can potentially provide ratings as accurate as those provided by adult raters. Student ratings have not been validated for use in summative assessment and should not be used as the sole or primary measure of teacher evaluation.	Unknown.	Student ratings are cost-efficient and time-efficient, and can be collected unobtrusively.[9]	Provides perspective of students, who have the most experience with teachers. Can provide formative information to help teachers improve practice in a way that will connect with students. Students cannot provide information on aspects of teaching such as a teacher's content knowledge, curriculum fulfillment, or professional activities.

Note: This matrix does not include a particular instrument for measuring Teacher Collaboration or Contributions to School Community. However, a discussion can be found in the question 1 section of 1D about the ways in which some teacher evaluation rubrics, especially those based on Danielson's *Framework for Teaching*, measure those two components.

Source: Adapted from Little, O., Goe, L., and Bell, C. (2009). *Practical guide to evaluating teacher effectiveness* (available online at http://www.tqsource.org/publications/practicalGuide.pdf). Copyright 2009 National Comprehensive Center for Teacher Quality. Used with permission. The text in this table was taken directly from Appendix C in *A Practical Guide,* but it has been reorganized, with different column headings, and was re-titled to use terminology consistent with the rest of this REL West Reference Desk response.

---

[9] Worrell, F. C., and Kuterbach, L. D. (2001), as cited in Little et al. (2009).

3. How could these methods be used as part of a teacher evaluation system?

Little et al. recommend taking a deliberate approach to deciding which among the various evaluation methods will best serve a school's or district's needs. They suggest starting with a process to identify the priorities of the school or district and the purpose of the evaluation, followed by strategic selection of specific evaluation methods to create a system that accomplishes the identified goals.

Little and colleagues also suggest that, in designing a teacher evaluation system, the following points be considered:

- Incorporate measures that take into account the different contexts across content areas, grades, subgroups of students, varied student populations between schools (a consideration when building a teacher evaluation system for districtwide use), school culture, and local circumstances (Campbell, Kyriakides, Muijs, and Robinson 2003, as cited in Little et al. 2009).

- Integrate summative and formative processes so that summative results lead to teacher assistance, if needed, and not directly to teacher dismissal.

- To ensure validity, decide what needs to be measured and choose instruments/methods designed for that purpose.

- Select multiple measures that, collectively, capture different aspects of teacher effectiveness and, thus, yield a rich understanding of teacher practice.

The following table provides a side-by-side comparison of six evaluation measures that can be used in different combinations to build a teacher evaluation system.

Table 2. Matching methods to specific purposes

Purpose of evaluation of teacher effectiveness	Value-added	Classroom observation	Artifacts analysis	Portfolio	Teacher self-report	Student survey
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		X			
Gather information in order to provide new teachers with guidance related to identified strengths and shortcomings.		X	X	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	X	X		
Examine the effectiveness of teachers in nonacademic subjects (e.g., art, music, and physical education).		X		X		X
Determine whether a new teacher is meeting performance expectations in the classroom.		X	X	X		X
Determine the types of assistance and support a struggling teacher may need.		X	X		X	X
Gather information to determine what professional development opportunities are needed for individual teachers, instructional teams, grade-level teams, etc.	X	X			X	
Gather evidence for making contract renewal and tenure decisions.	X	X				
Determine whether a teacher's performance qualifies him or her for additional compensation or incentive awards.	X	X				
Gather information on a teacher's ability to work collaboratively with colleagues to evaluate needs of and determine appropriate instruction for at-risk or struggling students.				X	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.
Source: Little, O., Goe, L., and Bell, C. (2009). *A practical guide to evaluating teacher effectiveness.* Chicago: National Comprehensive Center for Teacher Quality. Copyright 2009 National Comprehensive Center for Teacher Quality. Adapted with permission. The title of the matrix from which this is adapted is "Matching measures to specific purposes."

4. **What are some examples of comprehensive teacher evaluation systems around the nation?**

This section provides brief descriptions of three comprehensive teacher evaluation systems that combine various methods for evaluating teacher effectiveness.[10] These three are identified, with others, on the NCCTQ website as examples of combination models, that is, those that combine multiple evaluation methods. They were selected for inclusion here because of their applicability to teachers of all grade levels and subjects, whereas other models on the NCCTQ site may have limitations due to a particular focus (e.g., on pre-service teachers, new teachers and induction, a specific content area) or are in use only in small districts. The Toledo Plan, though not in the NCCTQ's *Guide to Teacher Evaluation Products*, is also included here because it is frequently cited as the first peer evaluation and mentoring program negotiated collaboratively between union and management and is the foundation for many other peer evaluation programs.

Teacher Advancement Program (TAP)
http://www.tapsystem.org/

The TAP system evaluates teachers in all content areas, grade levels, and stages of teacher development, using a combination of measures, including, but not limited to, classroom observations and value-added measures. Teachers are evaluated four to six times each year by trained peer evaluators and school administrators, all of whom go through a recalibration process annually. In conducting classroom observations, evaluators map what they see against the standards in the TAP Instructional Rubric, which is a Danielson-like framework. Afterward, a post-observation conference is conducted with the teacher by the TAP evaluator to discuss potential needs for teacher support. TAP evaluators are also responsible for analyzing data to determine student growth in each teacher's classroom. Evaluations are both formative and summative. Teachers who become evaluators, or master teachers, participate in the TAP Multiple Career Path and are released part-time from classroom instruction. Teachers who move up the TAP career ladder are compensated for their additional teacher leadership responsibilities, and teachers in a TAP school are eligible for performance bonuses based on observation evaluations and student achievement value-added measures. TAP is currently not used in California, but is being implemented widely in schools or districts throughout South Carolina, Louisiana, and Texas, as well as in individual schools in Minneapolis, Chicago, and Philadelphia, among others.[11]

Washington, DC IMPACT
http://dcps.dc.gov/DCPS/In+the+Classroom/Ensuring+Teacher+Success/IMPACT+%28Performance+Assessment%29

IMPACT is a performance assessment system designed for Washington, DC, public schools, which was first implemented in fall 2009. Differentiated evaluations exist for teachers of

---

[10] Descriptions are based on information culled from each evaluation program's home website.
[11] Location information of where TAP is currently being implemented was downloaded from http://www.tapsystem.org/action/action.taf?page=where on May 12, 2010.

general education, English language learner, and special education students; counselors; mentor teachers; and other support staff. Teacher evaluations use a combination of methods, each weighted differently, depending on the type of teacher being assessed. For example, for a teacher of English or math in grades four through eight, student value-added (as measured by the DC Comprehensive Assessment System, or CAS) currently accounts for 50 percent of the evaluation, classroom observation accounts for 40 percent, commitment to the school community accounts for 5 percent, and the school's overall value-added accounts for the final 5 percent. For the majority of DC teachers, no DC Comprehensive Assessment System (CAS) data exist. For these teachers, classroom observation accounts for 80 percent of the teacher evaluation, and non-value-added student achievement growth (as determined by high-quality, non-CAS assessments) account for 10 percent, with commitment to the school community and the school's overall value-added making up the final 10 percent. Five formal classroom observations are conducted each year — three by a school administrator and two by a "master educator," that is, a trained peer with subject area knowledge. All teachers are introduced to the *Teaching and Learning Framework,* the rubric used for evaluating instruction. The combined measures lead to formative and summative evaluations, and targeted professional development is provided for teachers needing improvement.

North Carolina Educator Evaluator System (NCEES)
http://www.ncptsc.org/

NCEES was developed for teacher and principal evaluation in North Carolina to identify high-performing teachers and principals to serve as mentors or leaders, as well as to identify low-performing teachers and principals in need of support or possible dismissal. The evaluation system is used for teachers in all subjects, grade levels, and levels of development. Teacher evaluations involve a self-assessment, a professional development plan, and three or four classroom observations — depending on tenure status — that may be conducted by a school administrator or a "designated evaluator." For probationary teachers, one of the four evaluations is a peer observation. Evaluators use a rubric to measure teachers as *Developing, Proficient, Accomplished,* or *Distinguished* in 25 elements within five standard areas: leadership, respect for diversity, content knowledge, instruction, and reflection. Post-observation conferences are designed to provide formative information to the teacher.

The Toledo Plan
http://www.tft250.org/
http://www.tft250.org/Toledo_Plan_Guide.doc

The Toledo Plan is a peer-evaluation and -assistance program aimed at improving teacher performance for beginning (first- and second-year) teachers and underperforming veteran teachers. Consulting teachers — experienced teachers who undergo a rigorous qualifying process — evaluate and mentor peers for up to three years. During this time, they are released full-time from classroom duties and are paid an augmented salary. Consulting teachers are assigned a caseload of 10–12 first-year teachers for whom they provide ongoing classroom observations and coaching for an average of 40 hours per year per teacher. Consulting teachers are solely responsible for evaluating first-year teachers, while

principals evaluate all returning second-year teachers. All classroom observations are conducted using a standards-based rubric. Consulting teachers mentor new teachers and may offer targeted support by demonstrating a classroom lesson, suggesting an instructional video, or arranging to have the new teacher observe instruction in another class. Underperforming veteran teachers may be referred to the "intervention program," through which they receive targeted professional assistance from consulting teachers. Recommendations of renewal, non-renewal, or termination are made to the district's Intern Board of Review. The program is in place in Chicago; Minneapolis; Cincinnati; Columbus, OH; and Montgomery County, MD.

5. **What innovative teacher evaluation practices exist within California school districts?**

We were unable to find any systematic documentation of evaluation systems in California; however, the four evaluation systems described below often are cited anecdotally as being innovative.

Long Beach Unified School District

Long Beach Unified's districtwide reform strategy, referred to as the "Long Beach Way," includes implementation of common pedagogical approaches that support consistency in instruction and learning for all district students in all grades and content areas. Teachers are expected to use a "common lesson plan" that directly addresses state standards and exhibits the Essential Elements of Effective Instruction (EEEI) — a range of research-based strategies all teachers are expected to implement in their lesson planning and classroom instruction.[12]

Long Beach Unified's teacher evaluation process involves three formal walkthroughs per year by the principal, who evaluates each teacher against standards similar to those in Danielson's *Framework of Teaching*. During a classroom observation, a principal may even document the details of a teacher's lesson, including student responses, and then review the transcript with the teacher in their post-observation conference.[13] The evaluation process also includes a "lesson analysis," using a protocol linked to teacher proficiency in the EEEI.

Separate from the evaluative walkthroughs are the Focused Instructional Walkthroughs, an integral component of the district's strategic and comprehensive professional development program.[14] Teams of district administrators or school-based leaders conduct informal walkthroughs at a school site to gather information on teaching and learning that will be used to identify professional development needs and provide feedback to the school. These walkthroughs, along with student achievement data, identify any needs for teacher instructional support that, in turn, may be provided through professional development,

---

[12] Long Beach Unified School District Fact Sheet, The Broad Foundation, 2009.
[13] *Promising Practices in the Long Beach Unified School District*, a letter from Long Beach Unified School District Superintendent Christopher J. Steinhauser, dated December 2007.
[14] Email communication from Cheri Haas for Bobbie Montes, Program Administrator for Long Beach Unified School District Human Resource Services – Teacher Quality and Retention, April 30, 2010.

coaching, or a specialist working with the teacher(s). Struggling teachers may also voluntarily participate in the district's Peer Assistance Review program.

San Gabriel Unified School District

San Gabriel Unified School District has begun using student growth data to inform its personnel practices in a pilot. Last year (2009/10), the district established a contract with Dr. William Sanders and Dr. June Rivers, pioneers in value-added modeling, to analyze student growth data using their value-added model. These researchers are producing district-, school-, grade-, and teacher-level reports based on its value-added analyses. The district has begun training teachers and principals on the effective use of these data, though data are not yet formally used as a part of the evaluation process. Superintendent Susan Parks has spearheaded this work. She was previously the superintendent in Baldwin Park, where she led a similar effort for several years, in which the local union actively participated. Since her departure that district has discontinued these efforts.

San Juan Unified School District

San Juan Unified School District, particularly known for its Peer Assistance and Review program, has three models in its teacher evaluation program. "Option A" is the traditional supervisory model for all new teachers, transferred teachers, and experienced teachers who desire the formative assessment model. Under this model, the principal conducts formal and informal observations, gathering evidence of effective teaching practice through the use of a rubric based on the California Standards for the Teaching Profession. The principal then reviews and discusses the evaluation with the teacher before the evaluations are turned in. Teachers who have worked in San Juan schools for four years and who have met the evaluation criteria in Option A, may choose "Option B," which recognizes that teachers grow professionally. Under Option B, a teacher identifies individual goals and pursues appropriate professional development or collaborates with other professionals to meet the goals. Throughout the year, the teacher documents and records notes of professional activity, while the principal conducts informal observations. The teacher writes a self-evaluation, and the principal writes an evaluative summary. The third model is the Peer Assistance and Review (PAR) program, which is available to beginning, referred, and volunteer teachers. In the PAR program, full-time teachers are released from the classroom and become consulting teachers, responsible for providing ongoing support to new and struggling veteran teachers, as well as conducting their PAR evaluations.

Poway Unified School District

Teachers in Poway Unified School District may be evaluated under four different programs: traditional evaluation using Poway's Standards and Continuum; Alternative Evaluation; Poway Professional Assistance Program for beginning teachers; and Permanent Teacher Intervention Program. The district is particularly known for the latter two, which are Peer Assistance and Review programs. Under the traditional program, tenured teachers are evaluated twice a year against a set of standards based on Danielson's *Framework for Teaching*. Teachers who have consistently received satisfactory evaluations for five years may choose to participate in the Alternative Evaluation program, in which teachers identify

individual professional goals, and then collaborate and reflect on how their area of focus made a difference for students in the classroom. In addition to being evaluated on measurable student outcomes through California Standards Test (CST) scores,[15] teacher performance in these first two programs is evaluated by teacher reflection, observation, documentation, and conference. Teachers in their first and second year in the Poway district receive support through the Poway Professional Assistance Program (PPAP), in which teacher consultants conduct frequent classroom observations, provide feedback and coaching, and assume responsibility for the new teachers' first-year evaluations. A separate program, Permanent Teacher Intervention Program (PTIP), is designed for struggling veteran teachers who are recommended to PTIP by a site administrator or who might volunteer for the needed support. In the PTIP, teachers receive coaching and assistance from teaching consultants. In Level 1 Assistance, the site administrator conducts the evaluation, and in Level 2 Assistance, an evaluation team is composed of the site evaluator, a district administrator, and a third person chosen by the PPAP Governance Board.

Suggested resources

Of interest may be a new online tool, *Guide to Teacher Evaluation Products* found on the National Comprehensive Center for Teacher Quality website. The tool reviews teacher evaluation measures and provides links to districts currently utilizing that particular measure or to teacher evaluation programs that utilize the measure and are available for purchase. http://www3.learningpt.org/tqsource/GEP/

References

In addition, it may be of interest to follow the development of The College-Ready Promise—a coalition of five charter-management groups in Los Angeles—that was recently awarded (November 2009) a $60 million grant from the Bill & Melinda Gates Foundation to develop new teacher effectiveness plans.

Borko, H., Stecher, B. M., Alonzo, A. C., Moncure, S., and McClam, S. (2005). Artifact packages for characterizing classroom practice: A pilot study. *Educational Assessment. 10*(2), 73–104.

Campbell, R. J., Kyriakides, L., Muijis, D., and Robinson, W. (2003). Differential teacher effectiveness: Towards a model for research and teacher appraisal. *Oxford Review of Education, 29*(3), 347–362.

Goe, L., Bell, C., and Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis.* National Comprehensive Center for Teacher Quality.

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

---

[15] Details on how measureable outcomes on the CSTs are used were not available through sources reviewed for this document.

Kane, T. J., Taylor, E. S., Tyler, J. H., and Wooten, A. L. (2010). *Identifying effective classroom practices using student achievement data.* NBER Working Paper No. 15803. *http://www.nber.org/papers/w15803*

Little, O., Goe, L., and Bell, C. (2009). *A practical guide to evaluating teacher effectiveness.* National Comprehensive Center for Teacher Quality.

Worrell, F. C., and Kuterbach, L. D. (2001). The use of student ratings of teacher behaviors with academically talented high school students. *Journal of Secondary Gifted Education, 14*(4), 236–47.