



Envisioning New Directions in Teacher Evaluation

HIGHLIGHTS

- Teacher evaluations have come under criticism 2
- The Stull Act and local collective bargaining agreements address teacher evaluations 4
- California teachers' groups say current evaluations are not helpful 7
- State law describes steps for dismissing teachers for unsatisfactory performance 8
- Teachers, administrators, and researchers generally agree on how teacher evaluations should change 10
- Districts throughout the state are revising their evaluation systems 14
- A group of charter management organizations is developing a new evaluation system 15
- Use of student test scores in teacher evaluations is gaining momentum despite some researchers' concerns 17
- What role should state policy play in helping to bring about reform in evaluation systems? 19

EdSource thanks the **Walter & Elise Haas Fund** for supporting the development and dissemination of this report.

Research confirms what many educators, students, and parents know intuitively—the most important thing a school can do to help students learn is to put them in a class with a highly effective teacher.

State and local policymakers have therefore become increasingly focused on ways to ensure instructional quality. They face vexing questions—What constitutes good teaching? How can a teacher's effectiveness be evaluated, measured, and strengthened? To what extent should evaluations be more tightly linked with high-stakes decisions?

Teaching is a “very complex endeavor” that is “both an art and a science” says Randi Weingarten, the president of the American Federation of Teachers, the nation's second-largest teachers' union. Good instruction begins with solid subject-matter knowledge. In addition, teachers are expected to employ a variety of strategies to help young people who vary greatly in their readiness to meet increasingly high academic expectations. Good teaching also requires the ability to adapt when lesson plans go awry, generate data from student work and analyze it, work as part of a team of teachers and administrators, and communicate well with parents.

Current systems related to teaching quality result from an amalgam of state laws—credentialing requirements, recruitment and

retention incentives, professional development programs, and dismissal procedures—as well as locally determined policies and practices. To help existing teachers improve, policymakers are focusing on ways to boost the quality of teacher evaluations. There is broad agreement that few districts' evaluation systems foster improvement in teaching.

Organizations of different stripes have sharply criticized the “typical” teacher evaluation system in place today. The American Federation of Teachers (AFT) has stated that “with rare exceptions, teacher evaluation procedures are broken—cursory, perfunctory, superficial, and inconsistent.” Additionally, the National Council on Teacher Quality, which advocates for reforms in a broad range of teacher policies, gave the nation a grade of D- in identifying effective teachers in its 2009 ratings of state-level teacher policies. And the New Teacher Project—which describes itself as a national nonprofit dedicated to closing the achievement gap by ensuring that high-need students get outstanding teachers—has asserted that “most teacher evaluation systems suffer from a slew of design flaws.”



Several aspects of teacher evaluation systems have come under criticism

Critics agree on many—though certainly not all—points in their assessment of the problems with teacher evaluation systems and the reforms needed. Discussions tend to focus on five aspects of evaluation—their frequency, content, differentiation, helpfulness, and attachment to consequences. Common criticisms and discussions about those five aspects are summarized below.

■ Frequency—Teachers are not evaluated often enough.

According to the National Comprehensive Center for Teacher Quality, probationary teachers across the country are typically evaluated twice a year, while permanent teachers are generally evaluated once every three to five years unless they receive an unsatisfactory rating, which triggers more frequent evaluation.

Some groups believe that all teachers should be evaluated at least annually. Others want to maintain the status quo or have experienced teachers direct their own improvement efforts in years they are not being formally evaluated. In California, the recent belt-tightening that districts have had to do makes increasing the frequency of evaluations unthinkable to some.

■ Content—Evaluations often involve superficial judgments about behaviors and practices—and too seldom take into consideration student academic progress.

Stakeholders agree that student learning should be the focus, but groups differ on the role that standardized test scores should play in teacher evaluations.

■ Differentiation—Few systems distinguish between poor, fair, good, and excellent teaching.

Some critics, such as the New Teacher Project, assert that strong performers do not receive the distinction they deserve and weak performers do not get a signal that they need additional support. However, leaders of the California Federation of Teachers (an affiliate of AFT) and the National Education Association (the country's largest teachers' union) see some reform efforts as overly focused on rewards and dismissals based on ratings.

Probationary versus permanent status and “tenure”

Teachers generally have probationary status for the first two to four years in the profession, depending on the state, and then transition to permanent status, which means they have more procedural safeguards protecting them from unjust dismissals. In common parlance and this report, permanent status is used interchangeably with “tenure,” though some stakeholders, including the California Teachers Association, define the two terms differently. (See *Evaluation: Key to Excellence* at www.cta.org.)

■ Helpfulness—Most teachers do not get useful feedback on their performance.

Stakeholders tend to agree that the primary goal of evaluations should be to help teachers improve so they can advance student achievement. Teachers especially say that most evaluations do not facilitate improvement.

■ Consequences Attached to Evaluations—Results rarely inform decisions about individual teacher's professional development or promotion, much less compensation, tenure, or dismissal.

All major groups agree that teachers should undergo periodic assessments that affect whether they may continue working in the classroom. However, national unions stress the need to protect the rights of permanent employees and provide opportunities for improvement before dismissal.

The federal government creates an incentive for states to reform evaluation policies

As part of the Race to the Top competitive grant program that began in 2009, the federal Education Department encouraged states to strengthen their policies to address these weaknesses in teacher evaluation systems (along with reforms in other areas of education policy). Despite offering relatively small grants, the Race to the Top generated a great deal of reform effort in nearly every state.

The program helped put teacher evaluation systems in the spotlight.

To be considered for a grant, a state had to ensure that its participating local education agencies would:

- conduct annual evaluations of teachers and principals;
- establish a clear approach to measuring student growth and incorporate that growth as a significant factor in evaluations;
- differentiate educator effectiveness using multiple rating categories;
- provide timely and constructive feedback; and
- use evaluations to inform decisions regarding professional development, compensation, promotion, retention, tenure, full certification, and dismissal.

Although ultimately unsuccessful, California applied in both rounds of the Race to the Top grant competition. In round one, California indicated that, if it received a grant, it would convene an advisory group to develop models of evaluation systems that met the federal criteria.

In round two, California's application was not from the state as a whole but from a consortium of seven school districts. The consortium pledged to develop an evaluation system using multiple measures of teacher

effectiveness, with 30% of a teacher's overall rating to be based on growth in student achievement.

Without the impetus of a Race to the Top grant, California lawmakers did not change

state law on evaluations. But the increased attention to the topic has helped prompt legislative efforts to revamp evaluation requirements. As this report went to press, the Legislature was considering Assembly Bills

5 and 48, which set out different visions for the evaluation of teachers. If one becomes law, it would be the first time in several years that evaluation policy has been substantially amended.

Many groups are working to improve teacher evaluation systems

This report refers to several organizations or networks—described below—that are involved in efforts to improve teacher evaluation. These organizations are funded in different ways. Some are membership organizations; some receive support from foundations—in many cases from the Bill & Melinda Gates Foundation; and some are funded by a combination of sources.

Accomplished California Teachers (ACT) formed in January 2008 to provide an educator's perspective on policy issues facing the state. Organized under the work of the National Board Resource Center at Stanford University, the teachers have achieved distinction in a number of ways, such as being selected as Milken* award winners or named teachers of the year; assuming leadership positions in their schools or districts; and earning certification from the National Board for Professional Teaching Standards. ACT is funded by the Stuart Foundation and The William and Flora Hewlett Foundation. <http://nbrc.stanford.edu/act>

American Federation of Teachers (AFT), an AFL-CIO affiliate, is the nation's second-largest teachers' union, representing 1.5 million workers, including 850,000 prekindergarten through grade 12 public school teachers. The state affiliate, the **California Federation of Teachers (CFT)**, advocates for 120,000 education employees in public and private schools and colleges. www.aft.org

Association of California School Administrators (ACSA) is the largest umbrella organization for school leaders in the nation, serving 16,000 administrators. ACSA offers training programs on school leadership and works as an advocate on education policy issues at the local, state, and federal levels. www.acsa.org

California Office to Reform Education (CORE) is a not-for-profit organization created by seven California school districts: Clovis, Fresno, Long Beach, Los Angeles, Sacramento, Sanger, and San Francisco. The districts first came together in October 2010 to develop California's application in the second round of Race to the Top federal funding. Although California was not chosen, CORE is working to put into practice some of the reform proposals from that application such as talent development, which includes teacher evaluation.

California Teachers Association (CTA) is the state's largest teachers' union, with about 325,000 members, including teachers, counselors, school librarians, social workers, psychologists, and nurses. The union also has affiliates that represent community college faculty, California State University faculty, and education support professionals. Besides acting as an advocate for educators, CTA also provides training on a variety of education-related topics. CTA is an affiliate of the **National Education Association (NEA)**, which has 3.2 million members. www.cta.org

* Milken award winners are early- to mid-career teachers who receive \$25,000 from the Milken Family Foundation "for what they have achieved and for the promise of what they will accomplish in the future."

Center for the Future of Teaching and Learning (CFTL) is a public, not-for-profit organization dedicated to strengthening teacher development policy and practice. CFTL guides and sponsors collaborative initiatives, including research, that focus on improving teacher quality and makes the information available to education policy stakeholders. www.cftl.org

The College-Ready Promise (TCRP) is a project that involves five charter management organizations (CMOs). The CMOs work together to create innovative approaches to recruit, train, evaluate, and compensate teachers and principals. A total of 90 schools serving about 30,000 students belong to the member CMOs—Alliance College-Ready Public Schools, Aspire Public Schools, Green Dot Public Schools, Inner City Education Foundation, and Partnerships to Uplift Communities. The College-Ready Promise is one of four recipients of a Bill & Melinda Gates Foundation grant program called Intensive Partnerships for Effective Teaching. www.thecollegereadypromise.org

National Board for Professional Teaching Standards. The National Board has developed teaching standards that describe expectations for accomplished teachers. For most subject and developmental levels, the board offers certificates that teachers can earn by successfully completing a set of rigorous assessments. Candidates may qualify for financial aid from federal, state, private, or school district sources. www.nbpts.org

National Comprehensive Center for Teacher Quality (TQ Center) was launched in October 2005 to serve as "the premier national resource" to which research and technical assistance organizations, states, and other education stakeholders turn for strengthening the quality of teaching. The federally funded TQ Center focuses in particular on strengthening teacher quality in high-poverty, low-performing, and hard-to-staff schools. www.tqsource.org

National Council on Teacher Quality (NCTQ) is a nonpartisan research and advocacy group that is funded by private foundations and is based in Washington, D.C. NCTQ advocates for reforms in a broad range of teacher policies at the federal, state, and local levels. The group does policy-oriented research and focuses on increasing public awareness about "the four sets of institutions that have the greatest impact on teacher quality: states, teacher preparation programs, school districts, and teachers unions." www.nctq.org

The New Teacher Project (TNTP) is a national nonprofit that partners with school districts and states "to implement scalable responses to their most acute teacher quality challenges." Since its inception, TNTP has established more than 75 programs and initiatives in 31 states, and published four studies on urban teacher hiring and school staffing. The majority of its revenue comes from its work with clients on a fee-for-service basis, though it does receive some federal and private funding. <http://tntp.org>

California law and local collective bargaining agreements each address aspects of teacher evaluation



The Stull Act of 1971 forms the basis of current state policy on the evaluation of certificated personnel (teachers, principals, counselors, and others). Stull Act provisions—which have been amended somewhat since the original enactment—plus collective bargaining agreements between employee unions and school districts—determine the outlines of teacher evaluation systems in California.

The Stull Act balances the state's interest in having teachers of a certain quality with employees' rights—namely the right to respond to their evaluations and to work without fear of capricious dismissal. The act does not speak to all the important details involved in teacher evaluation.

Collective bargaining agreements generally include more specific requirements, such as the amount of advance notice, if any, that teachers must receive before an administrator may observe their classes for evaluation. Further details may get worked out informally at the school level, with schools not always following even their own district policies. But this is not to say that the general approach to evaluation varies extensively throughout the state; researchers have found relative similarity in evaluation practice among schools in California. The research also indicates that many of the national criticisms related to the five aspects of teacher evaluations—frequency, content, differentiation, helpfulness, and consequences—apply generally to this state.

Limited funding and lean administrative staffs affect California school districts' approach to evaluations, including their frequency and content. In addition, districts tend to use relatively simple rating systems and channel their improvement efforts through means other than evaluations per se. State law attaches few consequences to teacher evaluations, but district and school administrators use them to inform decisions about teachers' careers, particularly those of probationary teachers.

State law specifies a minimum frequency for evaluations of teachers

Under the current version of the Stull Act, teacher evaluation must occur on a regular

basis, with the specifics depending on the employee's professional status. Probationary teachers must be evaluated at least once every school year. Permanent employees may be evaluated every other year, or less often if specific conditions are met. Those with permanent status who are "highly qualified," who have been employed at least 10 years in the same district, and whose previous evaluation was at least satisfactory, may be evaluated once every five years.¹ Any teacher who receives an unsatisfactory evaluation must be evaluated annually until a satisfactory evaluation is achieved or dismissal occurs.

An analysis of collective bargaining agreements by Katharine Strunk, an assistant professor at the University of Southern California, provides some insight into local practice as it relates to these state minimums. Her findings are based on the collective bargaining agreements in place in the summer of 2006 from 464 California districts with four or more schools. Those 464 districts represented 82% of districts that size, which served approximately 85% of California students. Policy Analysis for California Education (PACE) published her analysis in January 2009 in a report titled *Collective Bargaining Agreements in California School Districts: Moving Beyond the Stereotype*.

Strunk found that most collective bargaining agreements established the frequency of evaluations at the state minimum. However, 16% of districts had agreements calling for more than one evaluation per year for nontenured teachers, and 6% required more than one evaluation every two years for tenured teachers.

Strunk also looked at the relationship between districts' evaluation practices and their students' socioeconomic characteristics. Urban districts were more likely to require probationary teachers to be evaluated beyond the state minimum. In addition, large and urban districts had a strong tendency to have administrators spend more time on each observation than was the case in smaller and suburban districts.

California law specifies the content of teacher evaluations, but local practices vary somewhat

California's Education Code requires local school boards to establish standards of student achievement at each grade in each subject and to evaluate certificated personnel in the following four areas:

1. the progress of students toward reaching the district's standards and, if applicable, state content standards as measured by state-adopted assessments;²
2. instructional technique and strategies;
3. adherence to curricular objectives; and
4. the establishment and maintenance of a suitable learning environment.

Although the Education Code is clear about what must be assessed, the state does not actively monitor and enforce compliance. Nor does California law specify how districts must evaluate teachers and what sources of information they must use (other than state tests in applicable grades and subjects).

Schools can use a variety of methods and information sources for evaluations. The National Comprehensive Center for Teacher Quality (TQ Center) has grouped them into eight categories. (See the box on page 5.)

The TQ Center has created eight categories of information sources for teacher evaluations

Classroom observations, usually conducted by school administrators but sometimes by veteran teachers. Observations cover specific teacher practices and interactions between teachers and students.

Instructional artifacts, such as lesson plans, teacher assignments, scoring rubrics, and student work.

Portfolios, which can be used to evaluate a large range of teaching behaviors and responsibilities. Many states use them for assessing the performance of teacher candidates and beginning teachers. For example, the Performance Assessment for California Teachers (PACT), one of three approved assessments for teacher candidates in this state, includes a review of video clips, examples of student work, and daily reflections.

Teacher self-report measures, which may:

- consist of straightforward checklists of easily observable behaviors and practices;
- contain rating scales that assess the extent to which certain practices are used or aligned with certain standards;
- require teachers to indicate the precise frequency of use of practices or standards;
- take the form of surveys, instructional logs, or interviews.

Student surveys about teachers' practices. According to the TQ Center, several studies have shown that student ratings of teachers can be useful in providing information about teaching.

Value-added models, which summarize growth in student achievement in order to estimate the incremental effect of a teacher's instruction on students' learning. The statistical models can be complex; but in essence, value-added modeling looks at how students perform in one year relative to how they would have been predicted to perform, based on a host of factors including previous test scores, school and classroom characteristics, and specific student characteristics such as English fluency and poverty status.

Student performance measures, which allow users to examine student progress through goal setting, objectives, or testing.

Combination models, which may include a suite of measures of student and/or teacher performance such as those described above.

These information sources can be used for either of the two types of evaluations—formative or summative. Both types measure teaching performance, and both ideally include feedback to the teacher. Formative evaluations stress ways to improve, while summative evaluations are more geared toward informing decisions about the teacher's career advancement.

In California, formative evaluations are a particular focus during the required induction process for teachers in their first two years in the profession. The most common induction program, Beginning Teacher Support and Assessment (BTSA), assigns experienced teachers to provide ongoing, individualized support to newly credentialed teachers. But BTSA evaluations are conducted independent of any decision about a new teacher being given permanent status.

Researchers describe actual evaluation practice in California

A report by The Center for the Future of Teaching and Learning (CFTL) sheds light on how teacher evaluations are often done in California and their focus. *The Status of the Teaching Profession 2007* included a report of survey responses from principals in about 300 schools representing a range of performance and grade-span levels, plus case studies of 21 schools in seven districts varying in size, geography, and population density. CFTL found that the typical performance review consists of three steps, all revolving around a classroom observation by a school administrator:

1. a pre-observation meeting between the evaluating administrator and teacher, at which the teacher discusses the goals and background of the lesson to be observed.

2. a classroom observation at a predetermined time, during or after which the evaluator completes an evaluation form.
3. a meeting between the evaluator and teacher a few days after the observation to discuss the evaluation.

According to principal-survey data, 73% held a pre-observation meeting, 81% conducted announced observations, 84% had a post-observation meeting, and 91% provided the teacher with a copy of the completed observation form. Sometimes principals diverged from stated district policies—for example, not holding meetings before observations. In some cases, principals reported that this was done to avoid having the teachers prepare for a lesson differently than they would normally. In other cases, administrators deemed the pre-observation meeting unnecessary because they were evaluating veteran teachers.

In theory, teaching frameworks or standards guide the content of evaluations

When conducting a classroom observation, administrators typically bring a written rubric to document their assessment of a teacher's performance. The rubric is often based on teaching frameworks, or descriptions of instructional performance at multiple levels of competence.

Among the best known frameworks is Charlotte Danielson's Framework for Teaching, which she developed in 1996.

California's Commission on Teacher Credentialing adopted a somewhat similar document—the California Standards for the Teaching Profession (CSTP)—in January 1997. (See the box on page 6.) The standards, which were revised in 2009, are intended to prompt teachers' self-reflection about student learning and teaching practice; help them formulate professional goals; and guide, monitor, and assess progress toward their goals and professionally accepted benchmarks.

Unlike Danielson's framework, the CSTP do not include descriptions of levels of competence and were not developed expressly for evaluators to use. However, most districts in California use a rubric

based on the CSTP to evaluate teachers' performance. Some base their rubrics on California's Continuum of Teaching Practice, which describes five levels of performance on the CSTP—emerging, exploring, applying, integrating, and innovating.

The National Board for Professional Teaching Standards are also influential in California. State law allows school districts to base their evaluation systems on the National Board's standards or the CSTP, though neither is required. The National Board has standards for a broad range of subjects and developmental levels of students, such as early childhood and early adolescence, and the board certifies teachers in each of its 25 subject/developmental level combinations. The standards describe expectations for accomplished teachers in each certificate area, and certification is generally regarded as quite rigorous.

Administrators limit their focus when evaluating

School leaders have demanding jobs, especially in California where they are responsible for relatively large numbers of staff and students. This undoubtedly limits the time they can spend on teacher evaluations. Survey data from the CFTL study suggest that it may also mean that administrators have to focus on just a few elements of instructional practice, rather than student outcomes, when conducting observations. According to survey responses, almost all principals view good classroom management as a very important aspect of teaching quality. A strong majority—approximately eight of every 10—saw teacher knowledge of curriculum and content as very important. But just two of 10 rated student-related measures, including test performance and attendance, as very important aspects of teaching quality. (See Figure 1 on page 7.)

Many teachers and administrators see evaluations as lacking substance

A 2010 report by a group of distinguished California teachers echoes many of the points made by national organizations and CFTL about the content of evaluations. The report, *A Quality Teacher in Every Classroom: Creating a Teacher Evaluation System that Works for California* by Accomplished California

Danielson's Framework for Teaching and the California Standards for the Teaching Profession cover similar topics

In Danielson's framework, the complex activity of teaching is broken into 22 components clustered into four domains:

- Planning and Preparation
- Classroom Environment
- Instruction
- Professional Responsibilities (for example, reflecting on one's performance, professional growth, and communicating with families)

The CSTP includes six standards, each with a narrative description and a series of questions that teachers are expected to ask themselves to foster improvement. The six standards include:

- Engaging and Supporting All Students in Learning
- Creating and Maintaining Effective Environments
- Understanding and Organizing Subject Matter
- Planning Instruction and Designing Learning Experiences
- Assessing Student Learning
- Developing as a Professional Educator

What qualifications are necessary to become and remain a teacher in California?

To become a fully qualified teacher in a K-12 public school, a person must earn a *preliminary* credential. This requires having at least a bachelor's degree; passing a test of basic skills in reading, writing, and math; demonstrating subject-matter knowledge in the subject(s) one plans to teach; and participating in a state-approved teacher preparation program. Such programs generally take one year and include coursework, supervised teaching in a public school classroom, and passing a Teaching Performance Assessment. A preliminary credential is valid for only five years.

To continue teaching beyond the initial five years, an individual must obtain a *clear* credential by either completing an induction program or earning a certificate from the National Board of Professional Teaching Standards. In induction programs, beginning teachers receive ongoing, individualized support and formative assessments. National Board Certification requires passing 10 rigorous assessments—four that feature teaching practice and six essay tests of content knowledge. A clear credential must be renewed every five years thereafter.

For more detailed information, go to www.edsource.org/iss_capacity_teacher_credentials.html.

Teachers (ACT), says that evaluations too often focus on easy-to-observe practices, such as classroom management and whether students are on task, rather than looking for evidence that students are actually mastering learning goals set for them. Despite the often superficial nature of evaluations, teachers are concerned that they may be misjudged if they have an "off" day when they are observed.

In addition, the ACT report argues that the amount of time principals have to conduct effective evaluations is seriously limited, especially in large and/or high-need schools where administrative duties are extensive. Particularly in high schools, rarely is one evaluator able to do substantive evaluations for a large number of teachers across a range of subjects. (See the box on page 7.)

These accomplished teachers also find the California Standards for the Teaching Profession wanting. They say the standards are good at identifying the elements of effective teaching, but “there is little agreement that they are a force in the work of teachers in classrooms or the reference points that drive conversations about teaching practice.” In the eyes of this group, the CSTP do not provide enough specificity for teachers and evaluators to know whether a standard of good instruction is being met.

Some districts help teachers improve through different kinds of evaluations

CFTL found a few schools and districts that diverge from the typical teacher-evaluation model of a single, brief classroom observation. Some would call their activities formative rather than summative evaluations, or might not label them as evaluations at all because of their ongoing and team-oriented nature. Regardless of labels, they are activities that both school leaders and teachers find beneficial.

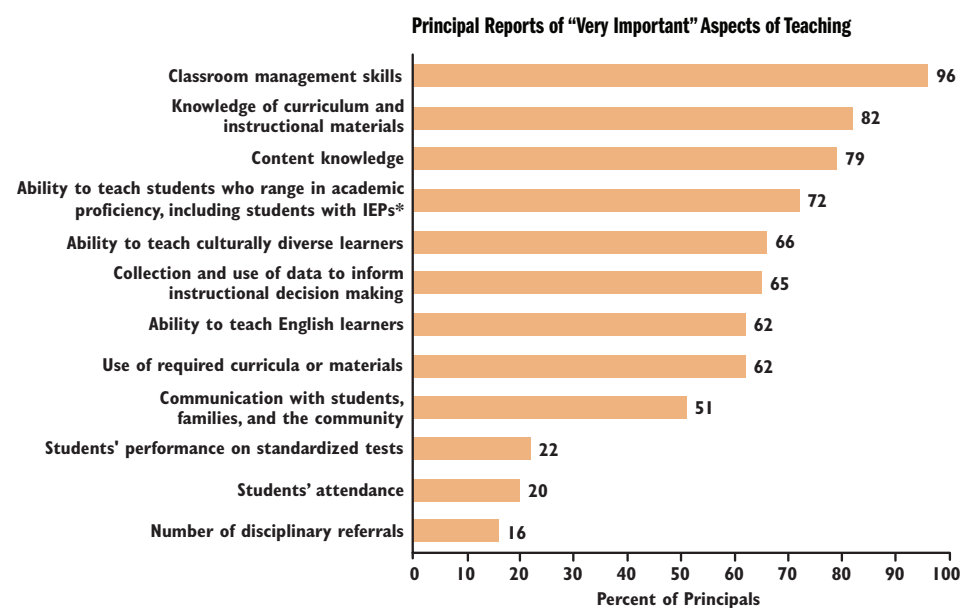
For instance, one district supplements the formal, pre-arranged classroom observation with shorter, more frequent “walk-through” visits. Prior to the walk-through, administrators discuss with all teachers an area of focus. An instructional coach then accompanies administrators on brief classroom visits, observing practices related to the area of focus. After all classes are observed, the full faculty discusses what was seen. A week later, the faculty reconvenes to discuss ways to improve.

In another district, veteran teachers are evaluated based on their ability to lead a small number of colleagues in setting and meeting goals for improving teaching. According to CFTL, many experienced teachers value opportunities to become leaders and share their knowledge this way.

California law is silent on how teacher performance should be differentiated

How evaluation results are summarized—for example on a rating scale—is decided locally. Strunk, in ongoing analyses of 2008–09 collective bargaining agreements, has found that the majority of districts include two or three levels of competence in their evaluation forms.

figure 1 | School administrators prioritize instructional practices over student outcomes when evaluating teachers



*IEPs are individualized education programs for students with disabilities.

DATA: THE CENTER FOR THE FUTURE OF TEACHING AND LEARNING (CFTL)

EDSOURCE 6/11

School administrators are spread thinly in California

Data on staffing ratios from the National Center for Education Statistics offer some insight into the workload that school administrators face in California relative to their counterparts in the rest of the country. This state ranked 48th among the 50 states plus Washington, D.C., in the number of principals and assistant principals per 1,000 students in 2008–09. While the average in California is 2.3 school administrators per 1,000 students, the average for the country as a whole is 3.2.

The largest district in the state uses two levels. An April 2010 report by a task force on teacher effectiveness in Los Angeles Unified School District states that teachers can receive an overall rating of “meets standard performance” or “below standard performance.” And 99.3% of the district’s teachers received the favorable rating, according to a 2009 report by the New Teacher Project. The task force recommended increasing the number of rating categories available to allow for the identification of exemplary teachers and those needing guidance, but did not specify a number.

California teachers' groups say current evaluations are not helpful

The state’s Education Code does not prescribe what many teachers want from evaluations—

affirmation of what they are doing well and suggestions for ways to improve.

California law makes some attempt to encourage these evaluation characteristics. It requires that evaluations include recommendations for improvement as needed. Teachers must be notified in writing if their performance is deemed unsatisfactory and provided a description of their performance. Specific recommendations for improvement must be made in conjunction with district support for teachers to meet them. The law also dictates that evaluations must be in written form and provided to the teacher at least 30 days before the end of the school year. A teacher may provide a written response to the evaluation. A meeting between the teacher and the evaluator must

be held before the last day of the school year to discuss the evaluation.

CFTL and ACT indicate that even these requirements are sometimes not met. Recall that CFTL reported that less than 100% of principals surveyed have a post-observation meeting with teachers (84%), or provide the teacher with a completed observation form (91%). And ACT asserts that substantive discussions focused on improvement before or after an observation are rare. One experienced teacher interviewed by ACT told of receiving a completed evaluation form in her mailbox without undergoing a classroom observation or meeting with an administrator. In cases such as this, teachers miss out on suggestions for improvement and formal recognition of their hard work and successes.

The ACT report asserts that evaluations are rarely well-timed or linked with professional development opportunities. This is particularly true for tenured teachers, who feel that evaluations are often pro forma. In CFTL’s survey, only half of the principals who responded said that the formal evaluation was very important in determining teachers’ professional goals or professional development plans.

“The current system focuses on a few, very small snapshots in time and isn’t really geared to improving practice,” says Robert Ellis, who chairs the Teacher Evaluation and Academic Freedom Committee for the California Teachers Association, the state’s largest teachers’ union. “We’d like to see an evaluation model that is truly helpful to teachers, one where they can learn and build on what they already know. Evaluation should support good teaching.” Educators have also indicated that they would like the people evaluating them to have experience teaching the same subject to similar student populations.

Teachers have a role to play along with administrators in making evaluations constructive. However, CFTL’s research revealed that not all teachers view evaluations as an opportunity to help them continually sharpen their practice. When CFTL asked experienced teachers whether the

results of evaluations were used to help them improve, several teachers responded that they had received positive reviews so they did not have specific areas to work on.

State law ties few consequences to teacher evaluations

California law has few provisions that specifically link teacher evaluations to consequences such as professional development, tenure, salary, and dismissal. However, in practice, schools often attach consequences to job performance generally and evaluations specifically.

An unsatisfactory rating can lead to participation in Peer Assistance and Review (PAR)

Under existing state law, any teacher who receives an unsatisfactory evaluation must participate in Peer Assistance and Review (PAR) if the teacher’s district runs such a program. However, experts say that few districts actually run one.³

Under PAR, a consulting teacher works closely with his struggling peers, supporting and assessing them for a year or two. The consulting teacher reports on their progress to a joint management-union board. The panel includes district administrators (often high-level

Under state law, districts must follow these steps for dismissing a teacher for unsatisfactory performance

Probationary Teachers*	Permanent Teachers
District provides written notice of intention to dismiss. <ul style="list-style-type: none"> ■ Must be given 30 days prior to dismissal. ■ For second-year employees, this can come no later than March 15. ■ Must include reasons for dismissal and a copy of performance evaluation. 	District provides written notice of intention to dismiss. <ul style="list-style-type: none"> ■ Must generally be given at least 90 days in advance of “filing charges” (see next step). ■ Must be given between Sept. 15 and May 15. ■ Performance evaluation must accompany notice. District “files charges” based on a majority vote of the school board. District must specify the problems with the teacher’s performance.
Employee has 15 days to request a hearing.	Employee has 30 days to request a hearing. If the employee does not request one, the district can dismiss the teacher.
If parties hold a hearing, it can be conducted according to procedures established by the district, including the involvement of an administrative law judge.	Hearing by a three-member Commission on Professional Competence. <ul style="list-style-type: none"> ■ Hearing must begin within 60 days of request. ■ The employee selects one member of the commission, and the district selects one. The commissioners must be certificated educators and must not be related to the employee or employed by the district. The third commissioner is an administrative law judge. Commission decides, by majority vote, for or against dismissal.
	Either party can appeal the decision in court. This process can last several years.

* For districts with average daily attendance of less than 250, the requirements are slightly different.

ones) and teachers and union officials appointed by the local teachers' union. Panel members engage the consulting teacher in discussions about the teacher's performance and eventually decide whether to recommend that the district retain or dismiss the teacher. The state's Education Code allows both probationary and permanent teachers to be dismissed, though the processes and the criteria for each differ. (See the box on page 8 for a summary of the dismissal process, and see the box on this page for a brief description of Poway Unified School District's PAR program.)

Tenure or permanent status is based on experience, but evaluations can play a role

State law does not explicitly link evaluation with a teacher's transition from probationary to permanent status. However, most principals effectively tie them to each other, according to survey data from CFTL.

California's Education Code establishes that teachers' first two years on the job are a probationary period. During this time, a district may choose not to rehire a teacher without providing a reason as long as the action is legal and does not violate civil rights. At the start of their third year of full-time employment in a district, teachers are given permanent status.⁴ According to CFTL, 87% of principals reported using performance review data to inform decisions to dismiss or retain beginning teachers.

Teachers' training and years of experience, not their evaluation results, determine their salaries

California law does not explicitly tie teachers' performance evaluations to their career growth. However, a teacher's performance in the classroom and with fellow teachers matters from a practical perspective. First, as previously mentioned, poor performance can lead to termination—though this rarely happens for permanent teachers. Second, strong performance can bring about leadership opportunities, such as designing curriculum, mentoring less experienced teachers, and serving on districtwide committees. In addition, honors such as being named teacher of the year generally reward an educator's ability to help her students, and the school as a whole, succeed.

Poway Unified School District has been an innovator with respect to teacher evaluations

Poway Unified School District, near San Diego, has developed uncommon approaches to teacher evaluation for more than 24 years. The Poway Professional Assistance Program (PPAP) operates the district's Beginning Teacher Support and Assessment (BTSA) program, Peer Review, and Permanent Teacher Intervention Program. Experienced teachers play a key consulting role in all three.

In most districts' BTSA programs, the mentor teacher does not evaluate the new teacher so that new teachers do not have to worry about exposing weaknesses. However, in Poway's BTSA program, the "teacher consultant" provides support for beginning teachers and conducts formal performance evaluations as part of Peer Review. Poway's former coordinator of PPAP says this approach works because the teacher consultant understands the developing needs of a beginning teacher, provides individualized support, and fosters a supportive, trustful relationship. The teacher consultant informs the principal and the PPAP governance board about the progress of the new teacher. The governance board—which includes the president of the teachers' union, two high-level district administrators, and two classroom teachers—guides teacher consultants' work, provides resources, and addresses challenging situations.

Poway runs a similar program for experienced teachers in professional jeopardy called the Permanent Teacher Intervention Program in which the teacher consultant performs the same functions. However, under this program the principal is considered the official evaluator.

And since the late 1980s, Poway has had an alternative evaluation system for successful veteran teachers. Under this system, the veteran teachers set their own professional goals, create their own professional development plans, and choose how their progress will be evaluated.

One element that contributes greatly to the success of PPAP is a set of specific teaching standards, according to *Peer Assistance and Review: Working Models Across the Country*, a March 2000 report by the Institute for Education Reform. The Poway Continuum of Teaching Standards is similar to the CSTP, but it gives examples of unsatisfactory, basic, proficient, and distinguished levels of performance.

In addition, a strong working relationship between the district's administration and teachers' union makes it possible to develop these progressive programs, according to a report by Accomplished California Teachers (ACT).

Where evaluations seem to have little consequence is in teachers' paychecks. State law creates an incentive for districts to set a minimum annual salary of \$34,000 for a teacher holding a bachelor's degree and a valid California teaching credential. Beyond that, teacher salaries are set by individual districts through the collective bargaining process, and evaluation results do not appear to play much of a role. The law requires districts to create a salary schedule on which all teachers are classified by their years of training and experience, or by other criteria if the district and local teachers' union agree.

Strunk's review of local collective bargaining agreements reveals that it is fairly common for districts to offer compensation

incentives for graduate degrees, but somewhat less common to boost pay for those certified to work with English learners or students with disabilities. Teachers used to receive \$20,000 from the state for becoming National Board certified and agreeing to teach in a low-performing school for at least four years, but the state stopped providing new awards in April 2009.

Budget-related layoffs are generally based on seniority

If a school district decides to reduce the number of certificated employees (for example, to balance its budget), state law specifies that layoffs must generally be done based on seniority, with the most recently hired employees being the first laid off. In cases where teachers

Recent court settlement affirms that factors other than seniority must be considered when districts implement budget-based layoffs

In February 2010, students at three Los Angeles Unified School District (LAUSD) middle schools, represented by the American Civil Liberties Union of Southern California, Public Counsel Law Center, and the law firm Morrison and Foerster, filed a lawsuit against the state and LAUSD seeking to stop budget-based teacher layoffs at their schools. The plaintiffs' attorneys argued that disproportionate layoffs at the three schools led to turmoil and the over-use of temporary replacements and rotating substitutes, violating the students' fundamental right to equal educational opportunity under the California Constitution. In May 2010, the Los Angeles Superior Court judge in the case issued a preliminary injunction, halting budget-based layoffs at the three schools.

Five months later, LAUSD reached an agreement with the student plaintiffs to settle the lawsuit. The district agreed not to lay off teachers for budgetary reasons at up to 45 targeted schools.

Twenty-five of these schools are in the bottom 30% of statewide academic performance rankings, suffer the highest rates of chronic teacher turnover, and yet demonstrate some academic improvement. LAUSD has identified 20 more schools that would be negatively and disproportionately affected by teacher turnover and is offering employees in those schools special protection.

The settlement also requires LAUSD to provide additional support to the targeted schools, including priority assistance with filling teacher vacancies and recruitment and retention incentives for teachers and administrators. To ensure that protecting the targeted schools from layoffs does not shift the burden to students at other schools, the settlement prohibits LAUSD from "redirecting" layoffs that would have occurred at the targeted schools to any school that will experience a higher

percentage of layoffs than the districtwide average for that year.

In February 2011, the trial court approved the settlement. The court's ruling was based on California Education Code provisions allowing a district to deviate from seniority as the basis for layoffs if it is necessary to maintain or achieve equal protection of the law. The United Teachers of Los Angeles, the local teachers' union, argued that those provisions were intended to protect teachers, but the court held that they were intended to ensure students' equal protection rights.

UTLA filed an appeal and sought a stay of the court's approval of the settlement pending the outcome, which the trial court and Court of Appeal denied. Accordingly, while the appeal is pending, LAUSD is implementing the terms of the settlement.

The statewide implications of the recent legal action and settlement are as yet unclear.

were hired on the same day, evaluations can help determine who is retained. In addition, the state's Education Code allows a deviation from seniority-based layoffs for either of two reasons: 1) a district has a specific need to maintain specialized services, such as those provided by a school nurse or Special Education teacher; or 2) to maintain or achieve equal protection of the laws. A recent legal decision

may give the latter provision greater salience going forward. (See the box above.)

First-year teachers can be laid off at the end of the school year, but other certificated employees are guaranteed more notice. Preliminary layoff notifications ("pink slips") must be issued by March 15; these are not an actual layoff notification, but rather warning that an individual is on a list of potential

layoffs. If teachers do not receive the preliminary notice, the district cannot lay them off. Districts then have until May 15 to issue final layoff notices or rescind preliminary notifications. For the most part, no new layoffs can occur after May 15; though during revenue declines that meet specific criteria, districts may be able to lay off certificated employees up to Aug. 15.

Teachers, administrators, and researchers recommend a new direction for teacher evaluations



Not only do many research, educator, and policymaking groups make similar criticisms about the current weaknesses in teacher evaluations, but they are often in general agreement regarding ways to make them stronger. Examined through the framework of frequency, content, differentiation, helpfulness, and consequences, their recommendations coalesce around several general goals. Not surprisingly, the various groups differ more in the specifics they emphasize and the concerns they raise.

All groups want teachers to receive frequent feedback

All of the interested stakeholders agree that teachers, especially those who are new or struggling, should receive frequent feedback on their practice. This would occur

mostly through formative evaluations and other instructional-improvement activities. In some visions of reform, teachers would receive such feedback monthly or even weekly. Instructional coaches and experienced colleagues could offer this support

along with administrators. Not all agree that teachers should play this role, however. In addition, although some organizations acknowledge that even formative evaluations require resources, none suggests what administrative or instructional activities

current staffs should give up in order to increase the frequency of formative evaluations in these fiscally lean times.

Groups diverge on the desired frequency of *summative* evaluations. The New Teacher Project asserts that school leaders should evaluate every teacher at least once a year and that these annual evaluations would allow schools to make important employment decisions based on up-to-date information. Race to the Top echoed this call for annual evaluations.

In contrast, leaders of some teachers' unions such as the National Education Association (NEA), the country's largest, say that permanent teachers do not need to be evaluated as frequently as probationary teachers. And the former president of the California Federation of Teachers (CFT) believes that experienced teachers do not need evaluations every year or even every other year but do need access to expert help when in a particularly challenging situation.

The Association of California School Administrators (ACSA) also sees room for flexibility, stating that the frequency should be determined at the local level as long as the state minimum is met. That approach is one in a list of recommended criteria for effective teacher evaluations that ACSA approved in October 2010.

Broad agreement also exists on the need for evaluations with richer content

Stakeholder groups agree that the typical evaluation needs to be more substantive—more focused on helping teachers improve and on evidence of student learning. Basing assessments on specific expectations embodied in rigorous teaching standards, and involving teachers in the design of evaluation systems, are part of the answer for these groups. In addition, multiple measures of teaching effectiveness have gained broad support. Finally, teacher groups see instructional quality as a collective responsibility of the entire school and even the broader community, and would like to see assessments of teacher performance reflect that.

Educators want evaluations to be based on specific expectations

Several national and California-based organizations believe that specific teaching

standards should be the foundation of both formative and summative evaluations.

However, the former president of the CFT is more skeptical of the efficacy of teaching standards. In correspondence with EdSource, Martin Hittelman stated, “We are not convinced that state standards for teaching are possible or realistically helpful.”

Accomplished California Teachers (ACT) believes in teaching standards but would like to see California's strengthened. The ACT report recommends that California use its teaching standards to create a continuum of specific expectations from entry in the profession to accomplished practice. Such an approach would be more purposeful and geared to improvement than is currently the case.

ACT would also add two summative performance assessments to the one that teacher candidates must pass.⁵ The first would occur during the induction phase (the first two years in the profession, generally speaking) and would help guide that process. The next would be administered a few years into a teacher's career. As teachers gain more experience, schools could use National Board assessments to encourage teachers to continue growing professionally.

Teachers say they should help design and implement evaluations

Educators say it is appropriate for them to play a role in designing evaluation systems, in part through collective bargaining. For example, the NEA leadership believes that individual teachers should help determine the set of practices and student learning objectives they are assessed on. In New York and Rhode Island, AFT has helped shape all aspects of teacher evaluation frameworks in 10 school districts.

In California, the law sets parameters on teachers' involvement, but practical considerations matter also. The state's Government Code specifies that the scope of collective bargaining includes “procedures to be used for the evaluation of employees.” Evaluation details that are not spelled out in law are determined locally. For example, state law requires that evaluations cover four elements (see page 4), but it does not specify the amount of emphasis that each should receive. In addition, local school boards may add elements.

Although teachers or their representatives are not explicitly guaranteed a place at the table when making decisions about such details, such an approach could make implementation smoother and more productive. As the ACT report states, teachers who do not share some power over decisions made about their work will resort to “the power of resistance.”

Evaluation systems will vary from district to district, but teacher groups assert that all systems should have certain elements. They advocate for strong training for evaluators, and they see a role for expert teachers in conducting evaluations. ACT goes further, saying evaluators should understand how to teach the relevant subject and be trained to recognize and develop teaching quality. The group of distinguished educators also believes that final recommendations from evaluators should be subject to review by an oversight team.

Both ACSA and the TQ Center agree that evaluators need training. ACSA says that professional development for principals should include training on evaluating teachers and that principals should be assessed partly on the quality of their teacher evaluations. Researchers from the TQ Center say that training can reduce evaluators' personal biases when assessing a teacher's effectiveness.

Reformers and researchers are interested in multiple measures of teaching effectiveness

Advocates of teacher evaluation reform believe that teachers should be assessed based on a range of evidence, such as classroom observations, lesson plans, and multiple student outcomes.

Regarding classroom observations in particular, a consensus is growing among teacher and administrator groups about using multiple observations to inform each evaluation. Although research on the ideal number is limited, so far it suggests that evaluations be based on three to five observations.

When it comes to using student outcomes to assess teacher effectiveness, many groups call for a variety of indicators. For example, in its criteria for effective teacher evaluations, ACSA calls for evidence of student academic growth based on multiple measures such as local and state

academic assessments, classroom participation, and student presentations, projects, and portfolios. ACSA also believes that teacher performance assessments should look for evidence that the teacher sets high expectations, engages students, and tailors instruction to students' needs.

Leaders of teacher groups have voiced great skepticism about a heavy reliance on students' standardized test scores. For example, Hittelman of CFT believes that student test scores "don't describe the whole of what a student knows nor do they indicate how good a teacher they currently have." Groups such as AFT and ACT are open to using test scores in conjunction with classwork, enrollment in advanced courses, graduation rates, pursuit of higher education, and success at work. In addition, they believe contributing indicators, such as student attendance, should be incorporated.

The National Education Association's board of directors has recently shown more openness toward using test scores in evaluations. In May 2011, the NEA board approved a policy statement on teacher evaluation and accountability that will go before the organization's policymaking body for approval in July. The statement calls for regular evaluations of all teachers based on multiple indicators—including the limited use of students' scores on standardized tests that are valid, reliable, and high quality measures of student learning. The policymaking body has not always supported the board's proposals, so the outcome of the July meeting is uncertain as this report goes to press. In addition, national policy statements do not limit state and local affiliates at the bargaining table. (For more on the use of student test scores in teacher evaluations, see the discussion of value-added modeling on page 17.)

The TQ Center has summarized the research on some of the possible measures for evaluations. A particular concern is that several would require substantial time and/or money to implement. Districts that are considering including them in an evaluation system would need to weigh the personnel hours and funding required against the potential benefits to instruction and student learning.

These choices are particularly constrained in California, both because of lean

administrative staffing levels at school sites and recent funding reductions that have led to cutbacks in numerous areas, particularly personnel. (See the table discussing potential evaluation tools on page 13.)

Teachers say that instructional quality is a collective responsibility

Some groups, such as the American Federation of Teachers and Accomplished California Teachers, believe that teacher evaluations should take into account factors beyond what happens in an individual educator's classroom.

In AFT's view, evaluations should not only measure the outputs that teachers help create such as test scores and student work, but also the inputs that teachers have to work with such as decent and safe facilities, professional growth opportunities, resources, and good school leadership. According to AFT, "accountability and responsibility for quality lie with teachers, administrators, other school staff, and other community members."

In a draft of principles on teacher evaluation that is still in progress, the California Teachers Association sounds a similar note, stating that "any evaluation system must consider the complexities of teaching and student learning that are outside of the teacher's control and beyond the classroom walls."

The ACT report argues that teachers should be evaluated based not only on success in their own classroom, but also on the success of their peers and the school as a whole. Including such measures would formalize a sense of collective responsibility that these experienced teachers already feel. The group states that "the genuine accountability that we feel to our students and to one another, when we work as part of a functional collaborative community, dwarfs any sense of accountability that can be imposed by test scores, site administrators, or state oversight."

Reformers have varying views of differentiation

The New Teacher Project (NTP) is the main proponent of creating evaluations that include multiple performance ratings. In *Teacher Evaluation 2.0*, NTP argues that:

Each teacher should earn one of four or five summative ratings at the end of each school year: for

example, "highly effective," "effective," "needs improvement" or "ineffective." This number of categories is large enough to give teachers a clear picture of their current performance, but small enough to allow for clear, consistent distinctions between each level and meaningful differentiation of teacher performance within schools and across the district.

Hittelman of CFT has a very different view, asserting that the emphasis should be on improvement in individual teacher behavior—not rating teachers.

The National Education Association bridges those two viewpoints. It does so by treating the feedback from formative and summative evaluations differently. In the NEA's vision of reform, as expressed in a December 2010 document, formative evaluations would provide feedback that is more nuanced than a rating.⁶ Formative assessments would allow peers, mentors, and coaches to offer constructive criticism and engage the teacher in a discussion, without employment-related decisions on the line.

In contrast, the outcome of a summative evaluation would be an up-or-down decision regarding a teacher's career, for example, to receive tenure or a promotion. It could also lead to an intensive improvement plan for a struggling teacher, or even dismissal if an intervention has not brought about the necessary improvement. Summative evaluations would be led by an administrator or supervisor and would adhere to prescribed schedules and rules.

Reform efforts are geared toward making evaluations more helpful

For organizations working on the improvement of teacher evaluations, the main goal of the reforms described above is to help educators hone their craft. Ideally, evaluations provide clear and actionable feedback based on established expectations plus evidence of impact on student learning.

Formative evaluations are, by definition, intended to promote improvement. They can occur in many forms—for example, in one-on-one mentoring sessions as part of an induction program, or in conjunction with schoolwide or department-wide examinations of instruction.

The general consensus is that summative evaluations should help teachers as well. First, the outcome of a summative evaluation is not just a decision, but also feedback. The assessment can lead to professional development opportunities designed to help

teachers improve as well as chances to lead other teachers. In addition, depending on how a district structures teaching careers and salary schedules, summative evaluations could lead to distinctions, promotions, and even pay increases.

Reformers agree that summative evaluations should have consequences

Stakeholder groups differ in how readily they would attach consequences to summative evaluations. The New Teacher Project states that the primary purpose of

Research on the strengths and limitations of potential evaluation tools

Below is a summary of several possible evaluation tools and their strengths and limitations, as described by the TQ Center in *Improving Instruction Through Effective Teacher Evaluation: Options for States and Districts*, published in February 2008. The table is not an exhaustive list. For example, some schools assess a teacher's performance with walk-through visits and surveys from students, parents, and a teacher's peers.

Evaluation Tool	Strengths	Limitations
Review Teachers' Lesson Plans	<ul style="list-style-type: none"> -Lesson plans show how well prepared teachers are to deliver content, develop student skills, and manage the classroom. -The level of planning has been shown to correlate with student learning. 	<ul style="list-style-type: none"> -Lesson plans are often adjusted as the lesson is taught; thus, the effectiveness of a lesson cannot be evaluated simply by looking at the plan.
Classroom Observations	<ul style="list-style-type: none"> -This is the most commonly used tool because it is able to capture information about instructional practices. -This can be used as both a formative and as a summative assessment tool. When used in formative evaluations, the observer can track a teacher's growth and suggest needed professional development and then later observe whether changes in teaching have been made. 	<ul style="list-style-type: none"> -Poorly trained observers and/or inconsistent, brief observations can lead to biased or inaccurate results. However, when observations occur more frequently, their reliability improves. -Observers often are not aware of the teacher's lesson plan. If, for example, the plan requires student accommodations, it would be difficult for the evaluator to know if the accommodations were implemented appropriately.
Self-Assessments	<ul style="list-style-type: none"> -Self-reflection during grade- or subject-area meetings, debriefings, or developing a portfolio or individual professional development plan may encourage teachers to continue to learn and grow. Videotaping class sessions allows teachers to review their performance. 	<ul style="list-style-type: none"> -Requires large amounts of time from the teacher.
Portfolio Assessments	<ul style="list-style-type: none"> -Combines the usefulness of a variety of other evaluative tools, such as review of lesson plans, a video of classroom teaching, reflection, and examples of student work and teacher feedback. -Promotes the active participation of teachers in the evaluation process. -Allows evaluators to review nonclassroom aspects of instruction. 	<ul style="list-style-type: none"> -No conclusive findings exist on the reliability of portfolios as part of an objective evaluation system. -Time consuming for both teachers and administrators.
Student Work-Sample Reviews	<ul style="list-style-type: none"> -May be able to identify which elements of teaching have a positive effect on learning better than standardized test scores. 	<ul style="list-style-type: none"> -Reviewing samples can be time consuming. -More prone to issues of validity and reliability than test items that have been validated for similar comparisons across different students in different schools answering similar test items. However, a way to reduce such subjectivity would be to develop a research-informed scoring rubric and train those who use it.

evaluations should not be punitive, and that good evaluations identify excellent teachers and help teachers of all skill levels improve. But the organization also states that an effective evaluation system “must be fully integrated with other district systems and policies and a primary factor in decisions such as which teachers receive tenure, how teachers are assigned and retained, how teachers are compensated and advanced, what professional development teachers receive, and when and how teachers are dismissed.”

As described earlier, the federal government used the Race to the Top program to encourage states to make those same linkages between evaluations and other personnel policies.

Teachers’ unions at the national level and in California agree that summative evaluations should have consequences, but they believe that it is fair and appropriate to attach consequences only if their members have the resources and time necessary to succeed. For example, AFT argues that states should establish standards for the environment in which teachers work and assess schools regularly for whether the conditions are conducive to teaching and learning. The union also believes that after “a valid and comprehensive system of teacher development and evaluation is in place, districts can formulate a fair process for tenure, career ladders, and, when necessary, removal of ineffective teachers who do not improve.” In addition,

the NEA says that novice teachers should have less demanding assignments and more time for planning than their experienced colleagues.

Accomplished California Teachers focus on tenure as a potential consequence of evaluations. The group proposes that tenure be granted to teachers only upon successful completion of a summative evaluation, with a school’s veteran teachers having a role in conducting that evaluation. This assessment would be the culmination of a substantive induction process. Beginning teachers who did not pass the evaluation would get additional support and have another opportunity to be assessed.

Designing the future of teacher evaluations



Several players inside and outside California are beginning efforts to improve how teacher effectiveness is defined and evaluated. For example, a sizable number of school districts throughout the state are developing new teacher evaluation systems. In addition, a group of five California-based charter management organizations are preparing to launch a new system of teacher development and evaluation in fall 2011 with significant financial support from the Bill & Melinda Gates Foundation.

The Gates Foundation is also funding an ambitious multistate research project called Measures of Effective Teaching (MET). One of the measures that project is analyzing is student test score improvement. The idea of including standardized test scores in teacher evaluations has generated substantial debate, and work by prominent researchers reveals some of the complexities involved.

Districts throughout the state are revising their evaluation systems

In many parts of California, school districts are beginning to design and implement new teacher evaluation systems. For example, approximately 20 small- to medium-sized districts in the northern part of the San Francisco Bay Area have created the North Bay Collaborative. Pivot Learning Partners, a statewide nonprofit, is helping guide the work.

Each district has joined the collaborative with the agreement of the local board, superintendent, and teachers’ union. Through a series of four day-long workshops, district teams consisting of central office staff, principals, and teachers will:

- conduct an internal scan of the strengths and weaknesses of their current systems;
- review and discuss literature on best practices in teacher evaluation;
- develop a framework for teaching and learning;
- create and align evaluation tools to the framework; and
- begin to establish processes for connecting evaluation to professional development, leadership opportunities, and interventions for teachers who are consistently failing to implement effective practice and demonstrate sufficient student growth.

The federal School Improvement Grant program is prompting districts to change their evaluations for teachers and principals

More than two dozen California school districts are also modifying their principal and teacher evaluation systems as part of their participation in the federal School Improvement Grant (SIG) program. SIG provides at least \$500,000 per year for three years to the lowest-achieving 5% of Title I schools that have also repeatedly missed academic performance targets. School districts and county offices of education with schools in the program also receive \$50,000 to \$500,000 per participating school for central office work.

The governing agencies for participating schools must implement one of four intervention approaches in those schools. The intervention approach relevant to this discussion is called transformation. Of the 92 schools receiving SIG funding in California, 57 are implementing a transformation. These

57 schools are spread among 29 districts. As part of a transformation, schools must establish evaluation systems for teachers and principals that make student academic growth a significant factor along with such components as multiple observations of performance and ongoing collections of evidence of educators' practice.

According to the legislation, teachers and principals must be involved in the design of the evaluation systems, though application timelines may have constrained these districts' ability to involve all stakeholders fully in a comprehensive reform effort. Many of the districts—for example, Twin Rivers Unified in Sacramento—view their new systems as pilot projects and plan to implement them throughout the district.

A consortium of seven districts is working to improve teacher development

California Office to Reform Education (CORE) is a not-for-profit organization created by seven California school districts in October 2010 to foster collaboration and learning. The seven unified districts are the same that applied for Race to the Top funding in round two of that grant competition—Clovis, Fresno, Long Beach, Los Angeles, Sacramento, Sanger, and San Francisco. CORE is working to actualize some of the reform proposals from that application.

One of CORE's areas of focus is strategies for teacher training, support, and evaluation. In addition to facilitating discussions among representatives of the participating districts, CORE plans to develop an open-source web portal so that they can share teacher and leader evaluation tools, trainings, and policies. Other projects are yet to be determined.

The inception of CORE did not mark the beginning of work on teacher evaluation reform for these districts. For example, both San Francisco and Los Angeles have been participating in SIG.

In addition, Los Angeles has been working on the design of a new teacher evaluation system since its teacher effectiveness task force issued a report in April 2010. The task force recommended changes in district

policies on evaluation, compensation, career pathways, tenure, and support for educators' professional development. On April 28, 2011, LAUSD superintendent John Deasy sent a letter to employees outlining a proposal for a new evaluation system and listing incentives for teachers to volunteer to be evaluated under the new system. In response, the teachers' union mounted a legal challenge, asserting that the district violated its legal obligation to negotiate in good faith with the union. As this report went to press, the dispute had not been resolved.

A group of charter management organizations is developing a new teacher development and evaluation system

The College-Ready Promise (TCRP) is a project with five charter management organizations (CMOs) collaborating to prepare students to succeed in college by creating innovative approaches to recruit, train, evaluate, and compensate teachers and principals. The member CMOs include Alliance College-Ready Public Schools, Aspire Public Schools, Green Dot Public Schools, Inner City Education Foundation, and Partnerships to Uplift Communities. Only Green Dot has a teachers' union. A total of 90 schools, which are mostly in the Los Angeles area and serve about 30,000 students, belong to the member CMOs. The College-Ready Promise is one of four recipients of an Intensive Partnership grant from the Bill & Melinda Gates Foundation, with the others being in Hillsborough County (Florida), Memphis, and Pittsburgh.

TCRP staff spent much of 2009–10 researching and designing a teacher development and evaluation system that reflects many of the reform ideas discussed above. In 2010–11, the group piloted a prototype after gathering significant input from teacher focus groups and advisory panels. Four of the CMOs will fully implement the new system in 2011–12, and the fifth, Green Dot, will implement beginning the following year. Green Dot's administrators and teachers' union will take an additional year to negotiate the details before ratifying a final system. Details of the new system might vary slightly among the CMOs when it is fully implemented.

The lessons that TCRP learns as it sets up, evaluates, and refines its ambitious teacher development and evaluation system could benefit California's larger K–12 education community. That said, the vast majority of the state's school districts would need to adapt those lessons to their own circumstances, which would likely include a teachers' union and no significant philanthropic support.

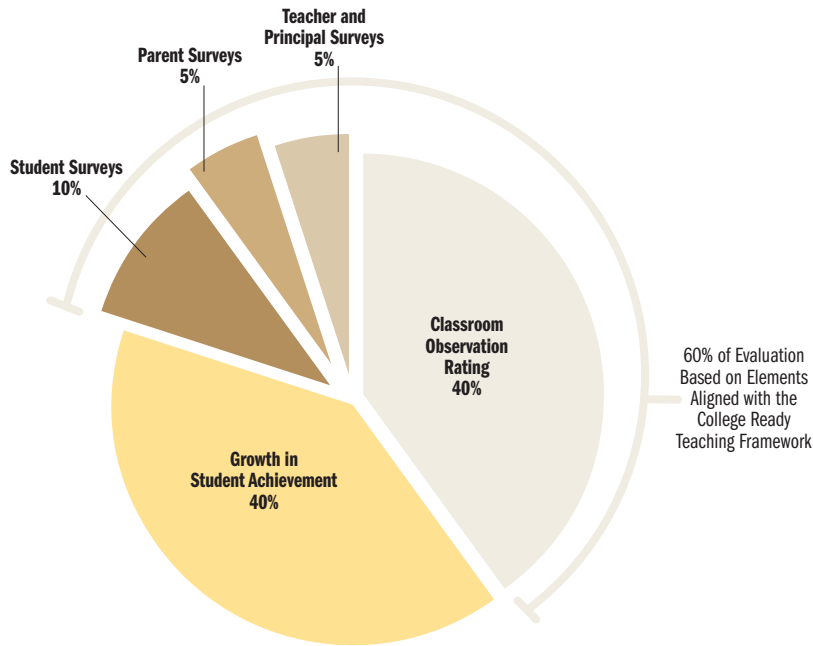
Multiple measures will inform a teacher's evaluation

Each teacher will be evaluated based on several factors, including classroom observation ratings by administrators and other certified educators, growth in student achievement, and survey results from students, parents, and fellow teachers. Each factor will have a specific weight. The weights that Aspire, the largest and oldest of the five CMOs, is using for the pilot project are shown in figure 2 on page 16.

Classroom observation will cover items such as planning, instructional practice, data-driven instruction, and classroom environment. To guide these ratings, the team has developed a teaching framework and rating rubric. The new College Ready Teaching Framework is based loosely on Danielson's but expanded and reflective of TCRP's values—such as promoting college-going for their students, collaboration among teachers, and “customer service” to families. The new framework has 50 indicators of teacher effectiveness. Some CMOs may try to focus on a few indicators each year, while others may try to incorporate all of the indicators into teacher evaluations every year.

The team spent substantial time working with the principals and instructional coaches doing the evaluations so that teachers will receive similar ratings no matter who the evaluator is. In addition to conducting classroom observations, evaluators will look at related lesson plans, student work, and teachers' review of their own practice. Teachers will have a chance to discuss any differences between the self-review and evaluators' assessments. TCRP's goal is to create both a snapshot measure of teacher effectiveness and an improvement measure.

figure 2 | Proposed evaluation components and weights in Aspire’s overall evaluation rating



DATA: THE COLLEGE-READY PROMISE (TCRP)

EDSOURCE 6/11

Improvement in student achievement will be a significant factor in a teacher’s evaluation. TCRP also plans to use “student growth percentiles” (SGP) in its evaluations. The measure compares a student’s improvement in California Standards Test scores to that of other students with the same prior score. Thus, a student whose achievement growth was greater than 64% of his peers would have a growth percentile of 64. The pool of comparison is the entire Los Angeles Unified School District. Teachers’ SGP is the median of their students’ growth percentiles.

In general, student achievement will represent 40% of the teacher’s evaluation, but this factor will not be based just on the achievement of students from the teacher’s class. The achievement of all classes in a department or grade, as well as the school as a whole, will be figured into the calculation. The exact distribution will require additional study as the designers want to create individual accountability for teachers and at the same time incentivize collaboration among colleagues. For teachers of subjects and/or grades for which the state does not have standardized tests, the achievement growth component will either be based on schoolwide performance or a teacher- or CMO-created test.

Survey responses from students, parents, fellow teachers, and principals will comprise the remainder of the composite measure. The surveys will be aligned to the College Ready Teaching Framework. Surveys will ask parents about the school as a whole and ask the other respondents about individual teachers.⁷

Evaluations will be tied to consequences

Evaluation scores will affect teachers’ compensation and professional growth. For example, they will help determine teachers’ eligibility for promotions, distinctions, and additional training and responsibilities.

Teachers in one of the CMOs—Aspire Public Schools—are used to having compensation tied partly to growth in student achievement, but the emphasis has been placed on schoolwide growth up to this point. Under Aspire’s previous system, teachers were comfortable sharing scarce resources such as mentor teachers or instructional aides. However, Aspire officials say the new evaluation system puts more weight on the improvement of individual classes, which is making some teachers nervous about whether they will get all the support they

need. In addition, teachers wonder about the sustainability of performance pay given the state’s recent history of budget cuts and the limited lifespan of foundation support.

The Gates Foundation is trying to determine the best measures of teaching effectiveness

Another ambitious project is called Measures of Effective Teaching (MET). Funded by the Bill & Melinda Gates Foundation, it has two goals. The first is to develop a set of measures that serves as an accurate indicator of a teacher’s impact on student achievement. Among several measures being examined are scores on state tests and supplemental tests designed to measure higher-order conceptual thinking, as well as student perceptions of the classroom instructional environment. The second goal is to help school districts determine whether the teacher-observation rubric they use accurately measures teacher effectiveness. The tool being developed will be called a “validation engine.” (See the box on page 17.)

A report of initial findings appears to bolster the case for using test score gains and student surveys

In December 2010, the project team issued initial findings based on student test score gains and student perception data from 2009–10. Subsequent reports, due out in 2011 and 2012, will cover the team’s findings after it has analyzed the other information it has collected. From initial analyses, the researchers concluded that:

- teachers’ past success in raising test scores is one of the strongest predictors of their ability to do so again;
- the teachers with the greatest gains on state tests also tend to help students on supplemental, higher-order tests;
- average students know effective teaching when they experience it—in other words, students’ ratings of teachers tend to align with test score gains; and
- different sources of information, in combination, can provide diagnostic, targeted feedback to teachers who are eager to improve.

A UC-Berkeley professor strongly criticized the report

Soon after the release of the initial findings, Jesse Rothstein, an associate professor of

public policy at the University of California–Berkeley, issued a critical response. Rothstein’s paper is relatively technical, but two of his major points are summarized here.

First, Rothstein questions whether the MET researchers predetermined their conclusions by their stated premises, rather than engaging in an open inquiry of the relationship between the data collected and teacher effectiveness. The project states upfront that teachers’ evaluations should depend to a significant extent on their students’ achievement gains, and any additional components of the evaluation (such as classroom observations) should be valid predictors of student achievement gains. For Rothstein, those premises rule out exploration of other questions—such as whether test score gains add substantively to other available information on teacher performance, are only loosely related with good teacher practice, or are a poor measure of all that students are supposed to learn in school.

Second, the evidence for the MET researchers’ conclusions is weak, in Rothstein’s view. For example, when the MET report states that teachers with the greatest gains on state tests also tend to help students on supplemental, higher-order tests, Rothstein says that is technically true, but the tendency is “shockingly weak.” A math teacher who placed at the 80th percentile in terms of student test score gains, for example, would have about a 30% chance of being below average on the supplemental, higher-order test, according to Rothstein.

Use of student test scores in teacher evaluations is gaining momentum despite some researchers’ concerns

The MET project is just one example of the growing interest in using standardized test scores as one measure of instructional quality. Forty states are beginning to incorporate student achievement gains into teacher evaluations, according to a February 2011 report by the Center on Education Policy.

Score growth can be calculated in several ways. Three types of measures have gained currency in recent years. One is *student growth percentiles*, which The College-Ready

The Measures of Effective Teaching Project is working with a large amount of data

In order to develop measures of effectiveness, the project began gathering and analyzing a large amount of data in 2009–10 and continued through 2010–11. The MET research team is working with a total of 3,000 teachers spread across several school districts throughout the country (but not including any California districts). Researchers are focused on math and English classes in grades 4–8, English in grade 9, and algebra and biology in high schools.

MET researchers are collecting five types of data on these teachers and classes:

1. Student achievement gains on state standardized tests and on supplemental tests designed to measure higher-order conceptual thinking.
2. Classroom observations and teacher self-assessments. Each teacher is videotaped four times per year, and teachers provide written commentary and supporting contextual materials related to the videotaped lesson. Trained experts review the videos based on the following teaching standards:
 - The Classroom Assessment Scoring System (CLASS) measure developed at the University of Virginia.
 - The Framework for Teaching developed by Charlotte Danielson.
 - The Mathematical Quality of Instruction developed at the University of Michigan and Harvard University.
 - The Protocol for Language Arts Teaching Observation (PLATO) developed at Stanford University.
3. Teachers’ pedagogical content knowledge. Teachers are tested on their ability to recognize and diagnose common student misperceptions.
4. Student perceptions of the classroom instructional environment. Students complete surveys about their experience in the classroom and teachers’ ability to engage them in the material.
5. Teachers’ perceptions of working conditions and instructional support at their schools. These data are also collected through surveys.

Promise uses. Another measure is *growth to standard*. That metric uses students’ current and past achievement to predict whether future achievement will meet a standard such as proficiency on a state test.

The third major method of computing achievement growth is value-added modeling (VAM). With this method, analysts try to attribute improvement in test scores to a particular program or teacher. To determine a teacher’s “value added,” analysts look at how his students score in one year relative to their predicted scores. Those predictions are based on a host of factors including previous test scores, school and classroom characteristics, and specific student characteristics such as poverty and English fluency. The difference between predicted and actual scores is attributed to the teacher. Among the three analytic methods, VAM is currently

receiving the most attention and occupies the center of the discussion below.

Value-added modeling and its uses have engendered vigorous discussions

The *Los Angeles Times* brought VAM into the public arena in August 2010 and again in May 2011, when it gave readers access to value-added data that linked student test score gains to thousands of individual teachers, by name, in LAUSD. That generated great controversy, with opinions differing on the importance of employees’ right to privacy versus parents’ right to know, and on what test scores can indicate about learning and teaching.

Some stakeholders think it is wholly inappropriate to even include test scores in teacher evaluations, much less rate them publicly. They provide many arguments for their position, but perhaps the two main assertions are that

it is impossible to isolate an individual teacher's impact from all of the factors that contribute to student learning and that test scores do not adequately capture the range and depth of content that students learn in school.

Others are willing to consider including test scores in teacher evaluations. These individuals agree on a few points:

- Analyses should focus on students' improvement, not on scores from a single testing period. Otherwise, a teacher would be inappropriately held accountable for the achievement level with which students entered the class.
- Measures of test score growth can be calculated for individual teachers in only a minority of cases because students do not take state tests in all subjects and grades.
- Even for that minority of teachers, growth measures have a substantial margin of uncertainty and should not form the only basis for a teacher's evaluation.

Where agreement ends is on the acceptable amount of uncertainty and potential negative side effects. Some researchers have sounded strong notes of caution regarding the inclusion of test score gains in teacher evaluations. Others acknowledge the problems but believe the benefits of inclusion outweigh the costs.

Some researchers have expressed strong reservations about VAM

Described below are several reasons that some researchers urge caution in using test scores and VAM in teacher evaluations. One of the most widely cited critiques is an August 2010 paper entitled *Problems with the Use of Student Test Scores to Evaluate Teachers*, published by the Economic Policy Institute.⁸

- The quality of VAM analyses depend greatly on the underlying tests. In some states, standardized tests are not sufficiently nuanced or comprehensive to give an accurate indication of the effects of a program or particular teacher. Furthermore, many state testing systems are not designed to measure growth from grade to grade.
- VAM does not make all necessary adjustments for the impact of student background on test scores. For example, a VAM analysis comparing results from one year to the next

does not necessarily neutralize differences in summer learning loss among student subgroups. In addition, it is possible that not just initial achievement, but also rates of progress vary by socioeconomic status. VAM would need to correct for that in order to be a fair measure of individual teacher effectiveness. The imperfect adjustment for student background characteristics would matter less if students were randomly assigned to teachers and teachers were randomly assigned to schools; but assignments are often deliberate, not random.

- Attributing students' results to a particular teacher is difficult. Education is complex and cumulative, with many factors influencing how much or how well students learn. For example, English teachers are not the only educators who can affect a student's writing skill. Furthermore, a student's progress in a given year depends on the preparation received from teachers in prior years.
- VAM results can be imprecise and unstable. One study using standard VAM techniques indicated that if the goal is to distinguish relatively high- or low-performing teachers from average ones, the error rate is about 26% when three years of test data are used for each teacher. This means that one in four teachers of average quality would be misclassified as outstanding or poor teachers, and that a quarter of those who should be singled out for special praise or support would be deemed average. To reduce the error rate to 12% would require 10 years of data for each teacher. In addition, teachers' scores at the high and low ends of the scale—where decisions of rewards and potential dismissal are most likely to occur—are the most unstable.
- Emphasizing test scores can have negative side effects. Research has shown that test-based accountability can lead to a narrowing of the curriculum and teacher attrition and demoralization. It can also create an incentive to work only with students likely to show growth.

Other researchers acknowledge VAM's weaknesses but encourage its use

Other researchers say that VAM should not be held to a standard of perfection. They say

that VAM should be used in conjunction with other evaluation tools and is a far better complement than any available alternative.

A November 2010 report published by the Brookings Institution, *Evaluating Teachers: The Important Role of Value Added*, states that if student test achievement is the desired measure of teacher effectiveness, VAM is a far superior predictor than other measurable teacher characteristics.⁹ The authors compare VAM to scores on teacher licensing tests, certification by the National Board for Professional Teaching Standards, years of teaching experience, and quality of undergraduate institution, among other measures.

This group recognizes VAM's instability but says it is on par with measures used in other important decisions. For example, the correlation between the average teacher's VAM scores from year to year is similar to the correlation between college admissions tests and freshman grade point average. They argue that VAM scores should inform decisions about teachers for the same reason that tests such as the SAT should play a role in college admissions—because they are one of the best available predictors of performance.

In addition, other industries make important decisions based on similarly unstable measures, according to authors of the Brookings report. For example, individuals and large organizations decide on health care providers based on metrics that are only modestly correlated with patient outcomes. In addition, an analysis of 22 studies of objective performance measures used for highly complex jobs found that year-to-year correlations were consistent with those of VAM for teachers.

However, the more cautious VAM researchers respond that it may not be appropriate to compare teaching with other industries. Random error may be more responsible for the instability of the measures in the other industries mentioned than in teaching because teachers and students are not randomly assigned to schools or to each other. Thus, any comparison between teaching and those industries may not be appropriate.

What role should state policy play in helping to bring about reform in evaluation systems?



Many factors influence teacher effectiveness directly or indirectly, from credentialing requirements to compensation and professional development.

In recent years, various stakeholders have zeroed in on another factor—teachers' performance evaluations. They agree that most evaluations are weak, and they are offering ideas to help school districts improve in this area. For their part, the federal government and private foundations have provided funding to help districts and charter schools design and implement new evaluation systems. In California, state officials want to contribute to these improvement efforts, and they may best be able to help by enforcing existing statutes and disseminating information about successful evaluation practices.

Teacher groups, administrators, and researchers in California agree that in most school districts, teacher evaluation systems are inadequate. The critique has several points:

- Teachers do not receive feedback on their practice frequently enough.
- The administrators charged with doing the evaluations often receive little training and may not have the same subject-matter background as the teacher being assessed.
- Evaluators typically focus on readily apparent teaching practices at the expense of thorough reviews of a teacher's impact on student learning.
- Nearly all teachers get the same satisfactory rating despite varying substantially in their skill levels.
- At least one study indicates that in about half of the state's schools, evaluations are not tied to teachers' professional development plans and many teachers say that evaluations are rarely helpful.
- Especially for veteran teachers, evaluations often amount to pro forma exercises with little or no meaningful consequences.

Various stakeholder groups are also relatively aligned on the changes they would like to see in evaluation systems. Reform proposals from different quarters call for frequent feedback through low-stakes formative

assessments as well as periodic summative appraisals to determine whether a teacher may continue in the classroom, with dismissals occurring only after a struggling teacher receives additional support and substantial time to come up to par. These groups want evaluations focused on evidence of student learning, measured in multiple ways, though opinions differ on the appropriate role of student test score gains. All believe that the primary goal of evaluations should be to affirm what teachers are doing well and help them continually improve.

What role should state policy versus local school districts play in strengthening teacher evaluations? To a considerable extent, the local educators who would actually implement new systems are best positioned to decide how evaluations should be conducted—in part because systems need to be designed to fit local circumstances.

However, state policy can set reasonable parameters, balancing the interests of various groups. Students have an interest in receiving good instruction because of the opportunities that a solid education affords an individual. Society as a whole has an interest in making sure its members are well-educated because of the benefits to civic and economic activity. And teachers, as employees, have an interest in knowing their employers' perception of their performance and having a chance to improve if poor performance is jeopardizing their job.

Enforcing current laws costs less than passing new ones

Indeed, California law sets out some basic principles for evaluations. It even shares some elements with evaluation-reform proposals. For example, it requires:

- formative evaluations for beginning teachers through the required induction period and for struggling teachers through peer assistance and review;

- periodic summative evaluations and meetings between evaluators and teachers to discuss the assessment;
- evaluators to assess teachers (and other certified employees) based on four substantive elements, including student achievement; and
- that evaluations include recommendations for improvement as needed and that specific recommendations for improvement be made in conjunction with district support for teachers to meet them.

According to reports by CFTL and ACT, actual evaluation systems do not live up to what is envisioned in state law. One could argue that state policymakers could best help improve districts' evaluation systems by focusing on ways to enforce current law. The state has a few enforcement mechanisms at its disposal, including audits, certification processes, and periodic monitoring (as it does with categorical program compliance monitoring). Policymakers face fewer constraints in enforcing current law than in creating new requirements, in part because California's constitutional and fiscal realities substantially limit policy creation.

The California Constitution requires the state to reimburse school districts for the cost of implementing mandated new programs or increased levels of service. Thus, state lawmakers are reluctant to impose new requirements—such as more frequent evaluations or additional factors to be evaluated—unless they can dedicate funding for that purpose upfront. In a time of massive budget deficits, lawmakers are unlikely to institute new requirements.

The state could support an exchange of research findings and success stories

The state could play a low-cost but useful role by disseminating research findings. For example, if researchers could discern an association between specific evaluation practices and increased student achievement, state

officials would do a great service to the field by distributing information about those findings. Or if the state published information about effective evaluation systems or made it easy for local practitioners to share their success stories, those efforts could be of great benefit.

For some districts, descriptions of strong evaluation systems would provide models to

emulate. However, for other districts, the details of the actual system would not be as important as a description of the process used to develop it. What may matter most about an evaluation system is that it reinforces a culture in which all members continually work together to assess and improve their performance in order to advance student learning. ■■

ENDNOTES

1 Under the No Child Left Behind Act, the federal government ties substantial funding to states' ensuring that their teachers are "highly qualified." Federal legislation lays out parameters for the definition of highly qualified but gives states some latitude to tailor the criteria. California's definition overlaps substantially with the credentialing requirements described in the box on page 6.

2 A 1999 law added the provision regarding state content standards as measured by state-adopted criterion-referenced tests.

3 According to data from the California Department of Education, nearly all local agencies received PAR funding in 2010–11. However, lawmakers allowed PAR funding (along with funding from about 40 other state programs) to be used for any educational purpose beginning in 2008–09.

4 If a teacher with permanent status changes districts, the new district may employ the teacher as a permanent employee without requiring a probationary period.

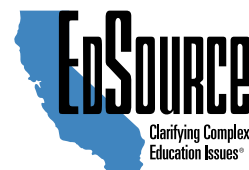
5 There are three approved pre-service assessments in California—the California Teaching Performance Assessment (CalTPA) designed by ETS, the Performance Assessment for California Teachers (PACT) designed by a collaborative including Stanford University and the University of California, and a third alternative recently approved by the Commission on Teacher Credentialing.

6 See: *Teacher Assessment and Evaluation: The National Education Association's Framework for Transforming Education Systems to Support Effective Teaching and Improve Student Learning* at www.nea.org/home/41858.htm.

7 In 2010, California lawmakers enacted a bill related to student surveys. Senate Bill 22 authorizes student governments in high schools to form a committee of students and teachers to create student surveys that teachers may use to gather feedback on aspects of a class and their effectiveness. Responses from this type of survey belong to the teacher and may be viewed by administrators only with permission of the teacher. They cannot be used as part of official teacher evaluations or collective bargaining.

8 The authors are Eva Baker and Robert Linn, National Center for Research on Evaluation, Standards, and Student Testing; Paul Barton, education writer and consultant; Linda Darling-Hammond, Edward Haertel, and Richard Shavelson, Stanford University; Helen Ladd, Duke University; Diane Ravitch, New York University; Richard Rothstein, Economic Policy Institute; and Lorrie Sheppard, University of Colorado at Boulder.

9 The authors are Steven Glazerman, Mathematica Policy Research; Dan Goldhaber, University of Washington; Susanna Loeb, Stanford University; Stephen Raudenbush, University of Chicago; Douglas Staiger, Dartmouth College; and Grover Whitehurst, Brookings Institution.



Trish Williams

EdSource Executive Director

2010–11 EdSource Board of Directors

Carl A. Cohn, President

*Co-Director, Urban Leadership Program,
Claremont Graduate University*

Lawrence O. Picus, Vice President

*Professor, Rossier School of Education,
University of Southern California*

Kenneth F. Hall, Fiscal Officer

*Executive in Residence,
University of Southern California*

Susan K. Burr, Secretary

Executive Director, California State Board of Education

Davis Campbell, Immediate Past President

Senior Fellow, UC Davis School of Education

Brice W. Harris

Chancellor, Los Rios Community College District

Reed Hastings

CEO, Netflix, Inc.

Gerald C. Hayward

*Co-Director (Retired),
Policy Analysis for California Education*

Barbara Inatsugu

*Program Director for Education,
League of Women Voters of California*

Santiago G. Jackson

*President, Santiago G. Jackson, Inc.
Education Consulting*

Loren Kaye

*President, California Foundation
for Commerce and Education*

Carol Kocivar

President-Elect, California State PTA

Kelvin K. Lee

*Superintendent (Retired),
Dry Creek Joint Elementary School District*

John B. Mockler

President, John Mockler & Associates, Inc.

Amado M. Padilla

Professor, School of Education, Stanford University

Don Shalvey

*Deputy Director, U.S. Program, Education,
Bill & Melinda Gates Foundation*

Michael Stone

Teacher, Aliso Viejo Middle School

Gloria Taylor

*Co-President, American Association
of University Women—California*

To Learn More

The key references used for this report include:

Accomplished California Teachers, *A Quality Teacher in Every Classroom: An Evaluation System that Works for California*. May 2010.

Center for the Future of Teaching and Learning, *Teaching and California's Future: The Status of the Teaching Profession 2007*. December 2007.

National Comprehensive Center for Teacher Quality, *Improving Instruction Through Effective Teacher Evaluation: Options for States and Districts*. February 2008.

New Teacher Project: *The Widget Effect*, June 2009; *Teacher Evaluation 2.0*, October 2010.

Policy Analysis for California Education, *Collective Bargaining Agreements in California School Districts: Moving Beyond the Stereotype*. January 2009.

For links to these materials and a complete list of references and related resources, go to www.edsource.org/pub11-teacher-evaluation-resources.html.

Acknowledgments

This report was researched and written by:

Brian Edwards

With research support from:

Barbara Wright, Liz Terry, Sue Frey

Edited by:

Mary Perry

Related Resources

- Accomplished California Teachers, *A Quality Teacher in Every Classroom: An Evaluation System that Works for California*. May 2010.
- American Federation of Teachers, *A Continuous Improvement Model for Teacher Evaluation*. January 2010.
- Association for California Administrators, *Effective Teacher Evaluations* (October 2010) and *Effective Principal Evaluations* (November 2010).
- Beginning Teacher Support and Assessment (BTSA) is the most common induction program for beginning teachers in California.
- To learn more about bills referenced in this report—Assembly Bills 5 and 48 (2011) and Senate Bill 22 (2010)—go to the California Legislature's [bill information](#) website.
- The Brookings Brown Center Task Group on Teacher Quality, *Evaluating Teachers: The Important Role of Value-Added*. November 2010.
- The [California Standards for the Teaching Profession](#) are intended to prompt teachers' self-reflection about student learning and teaching practice; help them formulate professional goals; and guide, monitor, and assess progress toward their goals and professionally accepted benchmarks.
- California Teachers Association, *Evaluation: Key to Excellence*. 2005.
- Center for the Future of Teaching and Learning, *Teaching and California's Future: The Status of the Teaching Profession 2007*. December 2007.
- Center on Education Policy, *More To Do, But Less Capacity To Do It: States' Progress In Implementing the Recovery Act Education Reforms*. February 2011.
- California's [Commission on Teacher Credentialing](#) works to “ensure integrity and high quality in the preparation, conduct, and professional growth of the educators who serve California's public schools.”
- California's [Continuum of Teaching Practice](#) describes five levels of performance on the California Standards for the Teaching Profession.
- The Danielson Group, *The Framework for Teaching: Components of Professional Practice*.
- Economic Policy Institute, *Problems with the Use of Student Test Scores to Evaluate Teachers*. August 2010.
- EdSource, *How state laws and collective bargaining shape the way teachers are evaluated, paid, and dismissed in California*. June 2011.
- Institute for Education Reform, *Peer Assistance and Review: Working Models Across the Country*. March 2000.
- Los Angeles Times* articles on value-added results for teachers in Los Angeles Unified School District:
- [“Who's Teaching L.A.'s Kids?”](#) Aug. 14, 2010.
 - [“Times updates and expands value-added ratings for Los Angeles elementary school teachers.”](#) May 7, 2011.
- Measures of Effective Teaching Project, *Learning about Teaching: Initial Findings from the Measures of Effective Teaching Project*. December 2010.
- The [National Comprehensive Center for Teacher Quality \(TQ Center\)](#) website has a wealth of information. Two resources that were particularly helpful in the preparation of this report follow:
- [Improving Instruction Through Effective Teacher Evaluation: Options for States and Districts](#). February 2008.
 - [Guide to Teacher Evaluation Products](#), an online resource.
- National Council on Teacher Quality, *Blueprint for Change: National Summary* (2010 State Teacher Policy Yearbook) and *2009 State Teacher Policy Yearbook*. (To see the 2009 Yearbook, click on the 2009 tab.)
- National Education Association, *Proposed Policy Statement on Teacher Evaluation and Accountability*. May 2011.
- National Education Association, *Teacher Assessment and Evaluation: The National Education Association's Framework for Transforming Education Systems to Support Effective Teaching and Improve Student Learning*. December 2010.
- New Teacher Project, *The Widget Effect* (June 2009) and *Teacher Evaluation 2.0* (October 2010).
- Two documents published by Policy Analysis for California Education (PACE) were helpful in the preparation of this report:
- [Collective Bargaining Agreements in California School Districts: Moving Beyond the Stereotype](#). January 2009. This is most easily found on the [Education Resources Information Center \(ERIC\)](#) website.
 - [Teacher Employment and Collective Bargaining Laws in California: Structuring School District Discretion over Teacher Employment](#). February 2011.
- [Pivot Learning Partners](#) is a San Francisco-based nonprofit organization that works “with education leaders in both schools and districts to develop, assess, and use the knowledge needed for schools to engage in a systematic and sustainable improvement process.”
- Rothstein, Jesse, *Review of Learning about Teaching*. January 2011.
- Teacher Effectiveness Task Force, Los Angeles Unified School District, *Final Report*. April 2010.