



# POLICY BRIEF

DECEMBER 2013

## Making Observation Count: Key Design Elements for Meaningful Teacher Observation

Jennifer Goldstein,  
California State University  
Fullerton



Policy Brief 13-5

**T**eacher evaluation has become a linchpin in current state and federal policies intended to improve public education in the United States. An emphasis on “test score” accountability has led many to overlook the opportunity that high quality evaluation can provide for improving teachers’ practice as well as removing ineffective teachers from classrooms. To reach that level of quality, however, teacher evaluation needs to include a system of teacher observation far more sophisticated than what is typically practiced in schools and districts today.

Historically, teacher evaluation has relied on the observation of teaching performance. This approach has proven largely unsuccessful, failing to support teachers in improving their practice and simultaneously failing to identify teachers whose performance is sub-par, who lack the ability or will to improve that performance, and who need to be removed from teaching. A confluence of influential reports on the role of teacher effectiveness and on the failure of teacher evaluation, together with the development of new statistical methods, spurred the relatively recent shift to teacher evaluation as a policy lever. Teacher evaluation reform then shot to the policy forefront with the federal government’s Race to the Top

### Executive Summary

Teacher evaluation has emerged as a potentially powerful policy lever in state and federal debates about how to improve public education. The role of student test scores and “value-added” measures in teacher evaluation has generated intense public controversy, but other approaches to evaluation including especially classroom observations of teaching are certain to remain as essential features of any evaluation system.

In this policy brief Jennifer Goldstein lays out four key design principles that should guide the observation-based assessment of teaching:

- Use standards-based instruments for data collection;
- Rely on observers/assessors other than building administrators, ideally master teachers, to conduct observations;
- Support observers by establishing shared responsibility and accountability for evaluations and employment decisions; and
- Partner with the teachers union.

Goldstein concludes that a robust teacher observation system can

*Continued on page 2.*

Jennifer Goldstein is Professor of Educational Leadership, California State University Fullerton. She is the author of *Peer Review and Teacher Leadership: Linking Professionalism and Accountability* (Teachers College Press, 2010).

### Executive Summary (Cont.)

contribute to policymakers' and the public's need for accountability, while also providing a powerful tool for improving instructional practice. California already has policies in place and practices in use that can help pave the way toward successful teacher evaluation reform in our state's schools.

initiative, which has rewarded states for linking teacher evaluation to student growth on test scores. The notion that we need “multiple measures” to assess teaching has become ubiquitous, and new policy reports on teacher evaluation reform appear regularly.

Despite its unsuccessful history, however, observation-based assessment of teaching is not going away. The observation of teaching performance will, except in rare cases, continue to be one of the multiple measures that districts use to assess teaching. Observations can be conducted for all teachers, not merely those teaching tested grades and subjects. Observations continue to hold more legitimacy in the eyes of teachers than statistical analyses of student test scores (perhaps ironically, given the historically low opinion teachers have had of their observations). Significantly, classroom observations done well provide not merely a piece of the teacher's overall performance assessment, but can provide immediate feedback to the teacher for improvement. Districts and states need to resist giving teacher observa-

tion short shrift as the measure with which we are already most familiar. It is absolutely imperative that we radically improve upon our historic systems of teacher observation.

In this brief, I first lay out key design principles for improved teacher observation systems. Second, I present an example of one approach that is aligned to these principles, peer assistance and review, which is commonly known as PAR. Third, I address critical issues that policymakers will have to address in order to put such redesigned systems into place.

### Design Principles for High Quality Teacher Observation

Quality teacher observation is challenging for several reasons. Doing it well requires a lot of time, and not merely time to conduct the observations themselves. An investment of time is needed to ensure an observer's skill in diagnosing performance and providing accessible feedback to teachers. In addition, time is needed to ensure calibration across observers. Beyond time, high quality observation requires appropriate grade level and content area expertise on the part of observers, matched to the teaching they are observing and assessing. All of this is needed to greatly improve the reliability and validity of observation-based assessments.

Doing observation well, however, is possible. There are four key principles that should guide the observation-based assessment of teachers:

- Use standards-based instruments for data collection;
- Rely on observers/assessors other than building administrators, ideally master teachers, to conduct observations;
- Support observers by establishing shared responsibility and accountability for evaluations and employment decisions; and
- Partner with the teachers union.

### *Using standards-based instruments for data collection: What is assessed?*

Standards are an integral part of any robust observation-based assessment system. Richard Elmore has argued eloquently that a main reason teaching is considered a “semi-profession” is the lack of agreed upon protocols of practice. Such protocols should not be static, but must evolve as the knowledge base expands with ongoing research.

Performance standards can provide this sort of protocol. For one thing, standards increase consistency between observers. In addition, focusing feedback to teachers on standards provides a framework for professional learning conversations, rather than feedback based on individuals and personality. The history of vague, unspecific, unhelpful observation comments and assessments has generated the current desire to identify valid and reliable observation instruments that can be used effectively for both teacher growth

and personnel decisions. This is a positive development for the field.

California participated in this conversation early, with the California Standards for the Teaching Profession (CSTP) in development since the 1980s. Charlotte Danielson’s Framework for Teaching has been hugely influential nationwide. (For examples, see Tables 1A and 1B.) The CLASS (Classroom Assessment Scoring System) is also widely used. Content-specific performance standards have emerged, including PLATO (Protocol for Language Arts Teaching Observation) and MQI (Mathematical Quality of Instruction). The Milken Foundation’s TAP (a placeholder for “The System for Teacher and Student Advancement”) has internal research demonstrating correlation to

value-added scores. Teach for America developed the Teaching as Leadership (TAL) framework, focused on teaching in low-income settings, for corps member assessment.

The critical point for any system of standards is to create enough differentiation to capture a wide range of performance levels and the nuances of teaching quality, as compared to the historic and woefully insufficient “Needs Improvement/Satisfactory” of many current school district teacher evaluation systems. Some recent federal initiatives have gone so far as to specify the necessary number of levels on performance standards (i.e., columns on a rubric) such as unsatisfactory, developing, meets standards, and exceeds standards. Beyond this, some

educators have long argued for different performance standards altogether for novices and experienced teachers.

Standards-based observation and follow-up is very time consuming, and many principals balk at their ability to conduct such time-intensive teacher assessment. Principals who are already not completing their evaluations are going to be slow to embrace a much more time-consuming model. In addition, evaluating to standards-based observation tools with validity and reliability requires extensive evaluator training and calibration across observers. It is therefore essential to reconsider who is responsible for teacher observation.

**TABLE 1A.** Framework For Teaching – Overview

<b>DOMAIN 1: Planning and Preparation</b>	<b>DOMAIN 2: The Classroom Environment</b>	<b>DOMAIN 3: Instruction</b>	<b>DOMAIN 4: Professional Responsibilities</b>
<p><b>1a Demonstrating Knowledge of Content and Pedagogy</b></p> <ul style="list-style-type: none"> <li>Content knowledge • Prerequisite relationships • Content pedagogy</li> </ul> <p><b>1b Demonstrating Knowledge of Students</b></p> <ul style="list-style-type: none"> <li>Child development • Learning process • Special needs • Student skills, knowledge, and proficiency • Interests and cultural heritage</li> </ul> <p><b>1c Setting Instructional Outcomes</b></p> <ul style="list-style-type: none"> <li>Value, sequence, and alignment • Clarity • Balance • Suitability for diverse learners</li> </ul> <p><b>1d Demonstrating Knowledge of Resources</b></p> <ul style="list-style-type: none"> <li>For classroom • To extend content knowledge • For students</li> </ul> <p><b>1e Designing Coherent Instruction</b></p> <ul style="list-style-type: none"> <li>Learning activities • Instructional materials and resources • Instructional groups</li> </ul> <p><b>1f Designing Student Assessments</b></p> <ul style="list-style-type: none"> <li>Congruence with outcomes • Criteria and standards • Formative assessments • Use for planning</li> </ul>	<p><b>2a Creating an Environment of Respect and Rapport</b></p> <ul style="list-style-type: none"> <li>Teacher interaction with students • Student interaction with students</li> </ul> <p><b>2b Establishing a Culture for Learning</b></p> <ul style="list-style-type: none"> <li>Importance of content • Expectations for learning and achievement • Student pride in work</li> </ul> <p><b>2c Managing Classroom Procedures</b></p> <ul style="list-style-type: none"> <li>Instructional groups • Transitions • Materials and supplies • Non-instructional duties • Supervision of volunteers and paraprofessionals</li> </ul> <p><b>2d Managing Student Behavior</b></p> <ul style="list-style-type: none"> <li>Expectations • Monitoring behavior • Response to misbehavior</li> </ul> <p><b>2e Organizing Physical Space</b></p> <ul style="list-style-type: none"> <li>Safety and accessibility • Arrangement of furniture and resources</li> </ul>	<p><b>3a Communicating With Students</b></p> <ul style="list-style-type: none"> <li>Expectations for learning • Directions and procedures • Explanations of content • Use of oral and written language</li> </ul> <p><b>3b Using Questioning and Discussion Techniques</b></p> <ul style="list-style-type: none"> <li>Quality of questions • Discussion techniques • Student participation</li> </ul> <p><b>3c Engaging Students in Learning</b></p> <ul style="list-style-type: none"> <li>Activities and assignments • Student groups • Instructional materials and resources • Structure and pacing</li> </ul> <p><b>3d Using Assessment in Instruction</b></p> <ul style="list-style-type: none"> <li>Assessment criteria • Monitoring of student learning • Feedback to students • Student self-assessment and monitoring</li> </ul> <p><b>3e Demonstrating Flexibility and Responsiveness</b></p> <ul style="list-style-type: none"> <li>Lesson adjustment • Response to students • Persistence</li> </ul>	<p><b>4a Reflecting on Teaching</b></p> <ul style="list-style-type: none"> <li>Accuracy • Use in future teaching</li> </ul> <p><b>4b Maintaining Accurate Records</b></p> <ul style="list-style-type: none"> <li>Student completion of assignments • Student progress in learning • Non-instructional records</li> </ul> <p><b>4c Communicating with Families</b></p> <ul style="list-style-type: none"> <li>About instructional program • About individual students • Engagement of families in instructional program</li> </ul> <p><b>4d Participating in a Professional Community</b></p> <ul style="list-style-type: none"> <li>Relationships with colleagues • Participation in school projects • Involvement in culture of professional inquiry • Service to school</li> </ul> <p><b>4e Growing and Developing Professionally</b></p> <ul style="list-style-type: none"> <li>Enhancement of content knowledge / pedagogical skill • Receptivity to feedback from colleagues • Service to the profession</li> </ul> <p><b>4f Showing Professionalism</b></p> <ul style="list-style-type: none"> <li>Integrity/ethical conduct • Service to students • Advocacy • Decision-making • Compliance with school/district regulation</li> </ul>

Copyright, 2013, Charlotte Danielson, All rights reserved.

**TABLE 1B.** Framework For Teaching – Example Detail Component 3b “Using Questioning and Discussion Techniques”

<b>Unsatisfactory</b>	<b>Basic</b>	<b>Proficient</b>	<b>Distinguished</b>
The teacher’s questions are of low cognitive challenge, with single correct responses, and are asked in rapid succession. Interaction between the teacher and students is predominantly recitation style, with the teacher mediating all questions and answers; the teacher accepts all contributions without asking students to explain their reasoning. Only a few students participate in the discussion.	The teacher’s questions lead students through a single path of inquiry, with answers seemingly determined in advance. Alternatively, the teacher attempts to ask some questions designed to engage students in thinking, but only a few students are involved. The teacher attempts to engage all students in the discussion, to encourage them to respond to one another, and to explain their thinking, with uneven results.	While the teacher may use some low-level questions, he poses questions designed to promote student thinking and understanding. The teacher creates a genuine discussion among students, providing adequate time for students to respond and stepping aside when doing so is appropriate. The teacher challenges students to justify their thinking and successfully engages most students in the discussion, employing a range of strategies to ensure that most students are heard.	The teacher uses a variety of series of questions or prompts to challenge students cognitively, advance high-level thinking and discourse, and promote metacognition. Students formulate many questions, initiate topics, challenge one another’s thinking, and make unsolicited contributions. Students themselves ensure that all voices are heard in the discussion.

Copyright, 2013, Charlotte Danielson, All rights reserved.

***Relying on observers/assessors other than building administrators: Who is assessing?***

In thinking about this question, it is crucial to distinguish between individual shortcomings and broader systemic limitations. Some site administrators may be ill-suited to the task of assessing classroom performance, while others do an excellent job. The argument here is that the historic *system* of observation-based assessment has not served anyone well – principals, teachers, or students.

At the center of designing new observation practices is an expanded conception of instructional leadership. Leadership by educational administrators is crucial, including leadership in teacher evaluation by principals. Rather than assessing and supporting all of their teachers, however, administrators can support expert teachers to work directly with classroom teachers. This reflects a move from individual instructional leadership (based in the principal) toward strong instructional systems, a distributed approach in which administrators rely in part on the leadership of those around them. Among other leadership tasks, the jobs of teacher leaders can be designed to focus specifically on the observation and assessment of teaching. Creating a designated observer role for teacher evaluation is a crucial element in strong observation systems.

Drawing on observers other than active administrators carries two main benefits, each of which addresses some of the challenges to quality observation

outlined above. First, the observer can be substantively matched to the teaching content of the observee. Second, the observer can devote far more time to observation and evaluation than can an administrator responsible for all aspects of running a school. Teacher leaders can commit the time necessary for ongoing professional learning in diagnosing performance and providing feedback, calibration with other observers, and the actual conduct of observations.

***Grade/subject match.*** Except in the rarest of cases principals do not possess substantive expertise in all of the grade levels or subject areas in their building. Relying on a single evaluator (the principal) reinforces idiosyncratic beliefs – “I know good teaching when I see it” – and undermines the creation of a professional knowledge base for teaching grounded in common standards. Especially with the advent of the Common Core State Standards, a belief in generic evaluation is increasingly untenable. Differentiating observers by grade level and/or subject area values the complexity and specificity of teaching practice. It values the role of content knowledge and pedagogical content knowledge—both for the purpose of targeted assistance to teachers to improve their practice, and for the purpose of appropriate summative decisions about the quality of practice.

***Time.*** Countless researchers have documented principals’ lack of time and their sense of being overwhelmed by the demands placed on them. Time

is perhaps the initial building block on which any effort to improve instructional quality rests, and—without substantial job redesign—principals are limited in the attention they can bring to teacher evaluation. They are often unable to find the time to complete and document observations thoroughly, because evaluations are only one of an endless parade of responsibilities. Principals admit that they cut corners with their evaluations, typically doing fewer than desired or even required on teachers perceived to be performing acceptably. For teachers whose practice does not meet expectations, principals have learned that it is far easier to engage in the “dance of the lemons,” eliminating a position and consolidating a teacher out of their school (and into someone else’s) than to produce the documentation required to pursue dismissal.

***Creating support systems in which the observers work: Move from individuals to teams***

We have an unfortunate habit of replicating with administrators the bad practices we establish for teachers. We widely recognize that teachers need ongoing feedback and formative assessment to improve their practice. Administrators, coaches, and others who may serve as observers also need feedback and opportunities to improve. If strong evaluation systems rely on strong evaluators (observers), then the system must also be designed to engage those observers in strengthening their own practice.

The notion of evaluator accountability is fairly novel in K-12 education. The Measures of Effective Teaching (MET) Project’s 2012 Report, “Gathering Feedback for Teaching,” begins to raise the issue, arguing that districts and states should “systematically track the reliability of their classroom observation procedures.”<sup>1</sup> Designing systems to ensure inter-rater reliability is a step in the right direction, to be sure. Acting as critical friend to observers by engaging in ongoing conversations about teaching and assessment practice is a different matter entirely.

Structures that support observers and hold them accountable can improve the assessment of teachers in a number of ways. First, creating some form of cross-district body that oversees teacher assessment can provide much-needed support to those conducting observations. Second, this body can hold observers accountable for their assessments. Typically, principals are not required to defend their assessments of teachers in any meaningful

way. The obligation to defend one’s assessments with evidence, however, encourages greater care in conducting and documenting observations. This likely sounds simple or obvious, but it is an absolutely crucial distinction between the systems of observation that we have now, and those we need. Third, such a cross-district body can keep a finger on the pulse of teaching quality across the district, and not just in individual schools. As such, it can play a role in teacher quality *equity*, ensuring that standards do not vary across schools serving different student populations. Fourth, when decisions are made by a team the weight of a negative evaluation does not fall solely on a single observer. Giving negative evaluations is personally difficult for many people, in all employment sectors. Sharing this responsibility with a team may therefore lead to more honest and effective evaluations.

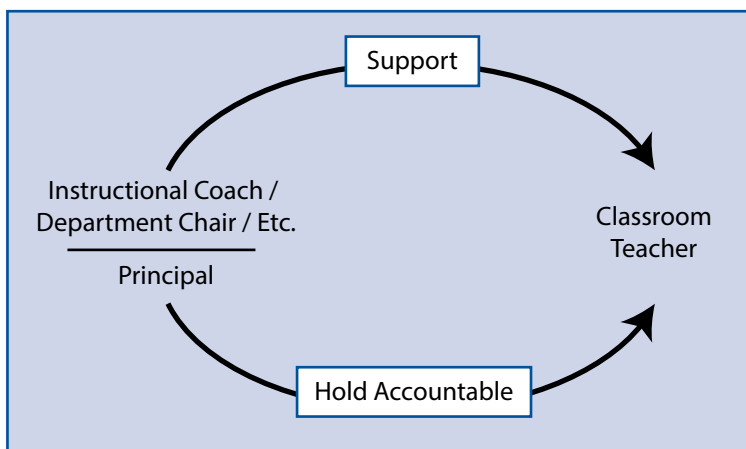
Figure 1 shows the relationship between support and evaluation in a typical teacher observation: a support

provider observes for the purpose of formative assessment (professional growth), while a “supervisor” observes for the purpose of summative assessment (personnel evaluation). In many cases, such as California’s Beginning Teacher Support and Assessment (BTSA) program, there is a so-called firewall preventing communication between these two people. By contrast, Figure 2 shows an altered relationship between support and evaluation, in which the two functions are linked and duplicated. Elmore has argued that accountability must be reciprocal; for every unit of performance I require of you, I owe you a unit of capacity to produce that result.<sup>2</sup> In Figure 2, observers are responsible for both supporting and evaluating classroom teachers. In turn, the district—through a designated team—is responsible for supporting observers and holding them accountable. This involves both supporting observers to support and evaluate the classroom teachers, *and* holding observers accountable for supporting and evaluating the classroom teachers.

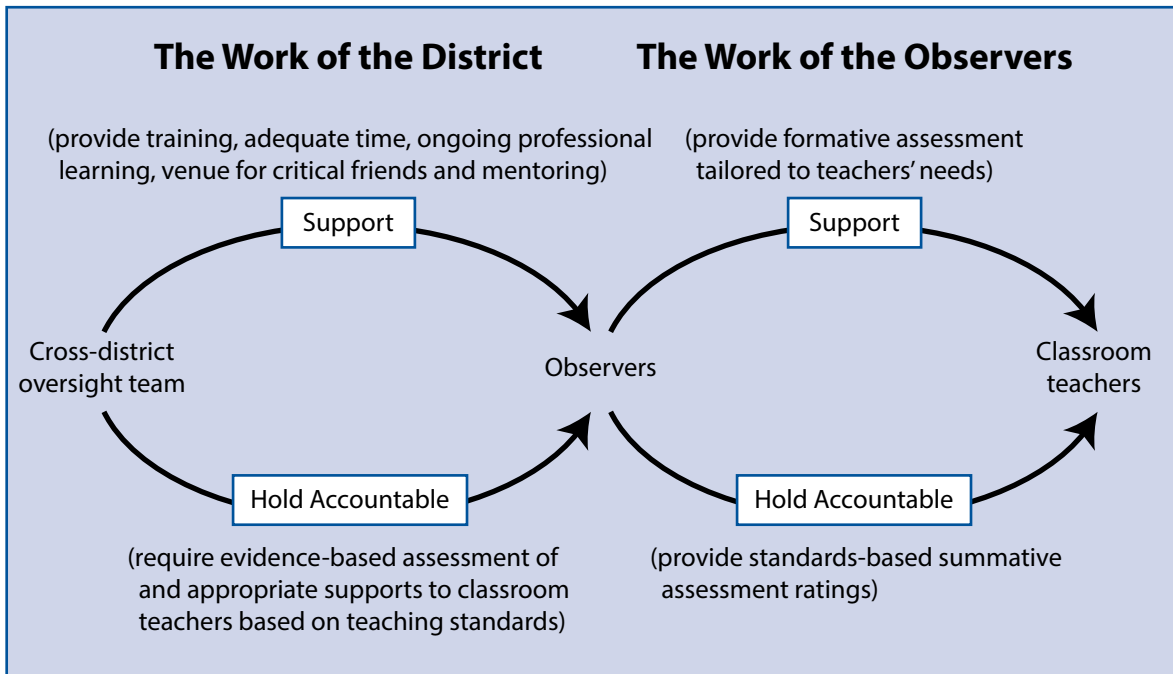
***The tenor of labor relations: Shifting from conflict to partnership***

Teachers unions have long been the object of widespread political animus. Research has documented that teachers union contracts are far less restrictive on the whole than they are often made out to be, and responsibility for selecting teachers, tenuring teachers, and documenting unsatisfactory performance are all currently responsibilities of administrators, not unions.<sup>3</sup> It is nonetheless certainly fair

**FIGURE 1.** Standard Teacher Evaluation’s Model of Support and Accountability



**FIGURE 2.** Teacher Observation Requires Double Reciprocal Accountability



to say that teachers unions have often acted as a major obstacle in efforts to dismiss underperforming teachers. In an unfortunate symbiotic relationship, principals become less likely to engage in rigorous observation and documentation of underperforming teachers, because they anticipate that union obstructionism will render their efforts meaningless. Unions, meanwhile, become accustomed to a pattern of principals trying to dismiss teachers without adequately documented evidence, and are consequently motivated to defend the removal of even ineffective teachers.

A critical step toward reforming teacher evaluation, including the practice of teacher observation, is to bring teachers unions to the table as serious players in the conversation, both to draw on their perspective, and

to remove them as an obstacle to dismissals when dismissals are warranted. Certainly union stances vary widely from local to local, and reform-minded positions of the national unions, where they exist, have always been slow to make their way to the local level. But many teachers union leaders are already reorienting their organizations to defend the quality and integrity of teaching, and to smooth the removal of poorly performing teachers.

**Existence Proof: Peer Assistance and Review**

Each of these design principles can be seen at work in a handful of districts nationwide that have implemented teacher peer assistance and review (PAR) programs. These programs vary widely, and different programs therefore result in very different out-

comes. My purpose here is to highlight the promising design elements that are included in many PAR programs, because research strongly suggests that these elements produce better outcomes than traditional systems of teacher observation and evaluation. In addition to the research on existing PAR programs around the country, I draw heavily on my own research, conducted over six years, of one urban California district’s implementation of a PAR program.

In the classic “Toledo model” of peer review and its variations, replicated in a variety of locations around the country,<sup>4</sup> expert teachers are released from classroom teaching duties full-time for two to three years in order to provide two types of support: mentoring for teachers who are new to the district or the profession, and inter-

---

vention for identified veteran teachers who are experiencing difficulty. These expert teachers (whom I will call PAR coaches but who are often called consulting teachers) typically work with a total of 12-15 teachers at a time, matched by grade level and/or content area. Notably, they also conduct the formal personnel evaluations of the teachers in the program, instead of (or in some cases, together with) the principal.<sup>5</sup> They report those evaluations to an oversight panel composed of teachers and administrators from across the district. The panel in turn determines the employment recommendation to be passed to the superintendent. Classroom teachers must meet specified quality standards within a set period of time, usually one year, or face removal from the classroom. PAR is a joint endeavor by a school district and its teachers union, and the panel is typically co-chaired by the teachers union president and a high-ranking district administrator such as the head of human resources. Once a teacher successfully exits PAR, he or she is on the principal's caseload for evaluation.

In studies of PAR programs across the country, the rate of removing underperforming veteran teachers from the classroom (whether dismissed or counseled out of teaching prior to dismissal proceedings) typically ranges from 40 to 70 percent, and can be as high as 100 percent at the outset of a program when those teachers long considered “the worst” are most likely addressed. The rate of removing underperforming beginning teachers from the classroom (not renewing them) can

average about 10 percent. These figures are for teachers *placed in PAR*, not all teachers. Principals nationally report that approximately 5 percent of their teachers are below standards, while only dismissing approximately 0.1 percent of them. PAR's figures therefore deserve attention—or rather, the design principles that lead to these higher removal rates deserve attention.

### ***Standards-based observation***

PAR programs exist that are not standards-based, and nothing intrinsic to PAR or to observation by expert teachers more generally requires the use of standards. Because PAR coaches are dedicated full-time to teacher observation and assessment, however, they have the time to conduct high-quality standards-based assessments. As a result, one key characteristic of the strongest PAR programs around the country is that they are standards-based. In California, some strong programs have been grounded in the CSTP. In the district I studied, for example, panel members and principals were impressed with coaches' expertise with the CSTP; the coaches became adept at using these standards for diagnostic purposes, referencing them line and verse. They could link intensive support to a teacher's diagnosed weak areas on the standards, and then link summative assessments to the teacher's progress in the areas of the focused support. Significantly, the coaches could document all of this work: regular observations; written feedback to teachers on what the coaches observed, with recommendations for improvements; growth or lack thereof over time; and ultimately

summative ratings based on their body of work with these classroom teachers. It is not uncommon for conscientious principals to see the observation documents and evaluations completed by a PAR coach and ask the coach to train them how to conduct similar evaluations.<sup>6</sup>

### ***PAR coaches***

PAR coaches are typically released from classroom duties to support their caseload of classroom teachers full-time. They can engage in frequent, ongoing announced and unannounced observations, once every one to two weeks, breaking down the privatization of teaching practice. Observations can be followed with specific feedback and other assistance, strengthened by the substantive grade/subject expertise the coach brings to the relationship. Ultimately, they can produce summative assessments that are the product of their accumulated formative assessments.

PAR coaches are not school-based but are instead matched to classroom teachers across the district by grade and subject. In other words, an expert high school math teacher might provide formative and summative assessment to 12 math teachers across five schools. As a result, their support and assessment is targeted by content. They can also bring a broader, cross-site perspective regarding practice and quality, enhanced by their ongoing conversations with other coaches and panel members.

One reasonable criticism of models like PAR is that the coaches lack a deep understanding of site context. This is

clearly a trade-off, and wise district leaders can compensate by providing new teachers in PAR with some form of site-based mentor for non-instructional support.

### ***District panel***

PAR involves oversight panels that hold hearings several times a year, during which coaches provide reports about classroom teachers' progress. The presence of the PAR panel (which goes by different names in different districts) is a crucial aspect of the PAR design.

PAR coaches are supported in their work not only by their fellow coaches, but by the PAR panel members. Panel hearings and ongoing communication between coaches and panel members serve as a sort of "Individual Education Plan" meeting for classroom teachers, with the group brainstorming how the coach can best support each teacher. At the spring hearing, and sometimes sooner, the PAR coaches make recommendations about the continued employment of each teacher, and when necessary, the panel challenges the coaches on the evidence they have provided to support their recommendations. Employment decisions are determined by the panel based on the recommendation of the PAR coach, sometimes together with the principal, at the panel hearing. Coaches by and large report appreciation for the professional input they receive from the panel.

The purpose of the panel is to support evaluators and to hold them accountable. PAR hearings bring more eyes

to teachers' practice (teaching) and to evaluators' practice (assessment). The panel, in its role holding the observer accountable, must study the evidence presented (such as artifacts from the classroom) and prevent coaches from making unwarranted employment recommendations. The effect of this diffused responsibility across the coach-panel-principal design appears to lead, perhaps counter-intuitively, to more accountability, not less. PAR coaches, who carry the biggest weight of making a negative recommendation about a teacher, can go back to that teacher following a panel hearing and say, "the panel decided not to renew you." Part of the reason higher rates of dismissal are typically seen with PAR than with traditional observation by a principal is this sense of shared responsibility.

### ***District-union collaboration***

In 1999, the California Assembly passed PAR legislation AB 1X (see Box 1). The law required union sign-off on districts' PAR programs. This gave teachers unions leverage to be a key player in teacher quality control, as did teachers union presidents serving as panel co-chairs in partnership with district administrators.

When they first attend panel hearings, many principals are surprised to see a teachers union president arguing to dismiss teachers. With this sort of district/union collaboration, the district in effect removes the union barrier that principals blame for their inability to remove ineffective teachers. As a result, districts are able to hold principals more accountable for

their role in the process of teacher evaluation. Anecdotal reports across PAR programs nationally suggest that principals start referring more teachers into the program once they see that supports and consequences exist for identified underperforming teachers.

This joint union/district, teacher/administrator work is a form of professional learning community that creates social capital and a collaborative environment over time. It is worth noting that while some districts across the country have started PAR programs through so-called trust agreements outside the contract, others have negotiated every last program detail as part of the contract. In other words, a collaborative district/union relationship is not a prerequisite to the creation of a PAR program; rather, partnership can grow over time as a product of joint work.

### **Box 1: PAR in California**

In 1999, California became the first state to institute PAR statewide; nationally, it was the first time a major district had implemented the policy in over a decade. Soon after Gray Davis became governor of California, the California legislature passed Assembly Bill 1X (AB 1X), which phased out the California Mentor Teacher Program and established peer review in its place. The legislation allocated approximately \$100 million for PAR—\$83.2 million in money previously attached to the mentor



program, plus \$16.8 million in new money. The bill created a de facto mandate for peer review. The legislation required school districts to have a peer review program in place by 2000 for veteran teachers who had received an unsatisfactory evaluation by their principals, or lose the state mentor money that the districts were already receiving. District leaders could decide for themselves whether new teachers would also participate in the program.<sup>7</sup>

Due to enormous policy flexibility and varying opinions about the wisdom of PAR among educators, school districts across the state created PAR programs that looked quite different from one another. Many California districts did not include new teachers in their PAR programs, since the state law only required the program for veterans. In addition, many programs did not create full-time, out-of-classroom positions for master teachers, and many did not involve the master teachers in summative evaluation. The legislation required “maximum local flexibility” for program details. The legislation did, however, require that school districts negotiate the development and implementation of their programs with local teachers unions, including a requirement that the union sign off on a district’s policy. AB 1X specified oversight panels with teachers in the majority (nine members comprising five teachers and four administrators).<sup>8</sup>

In the fall of 1998, Senate Bill 2042 introduced sweeping changes to

teacher credentialing in California. SB 1422, which had established the BTSA program for new teacher induction at the state level in 1992, required that a panel be formed to study teacher credentialing in the state. SB 2042 was the result of that panel’s findings. In turn, SB 2042 provided for the appointment of an advisory panel, which spent three years (1998 to 2001) developing program standards that would flesh out the new credentialing legislation. The result was a two-tiered credentialing program, which included induction as a formal second tier of teacher credentialing. Since AB 1X, legislating peer review, was passed in 1999, the SB 2042 advisory panel developed the credentialing program standards simultaneously with—though completely separate from—the implementation of AB 1X. Notably, districts had previously had discretion to use the \$83.2 million in state funds for the existing mentor program (transferred to PAR with AB IX) for their Beginning Teacher Support and Assessment (BTSA) programs.

The resulting SB 2042 program standards, presented to the California Commission on Teacher Credentialing in September 2001 (to be implemented by 2003), included a clause that prohibited *any* of the formative assessments generated for induction and credentialing purposes from being used for summative evaluation purposes. In other words, all new teachers would be required to participate in induction in order to earn a professional clear credential, and none of the documentation generated as part of

that induction process could be used for summative personnel evaluation (though one could perhaps argue that determining whether or not someone earns a credential is itself summative in nature). This meant that either districts had to run two separate parallel programs, with credentialed new teachers simultaneously participating in both—not likely at a time when money was drying up, nor a parsimonious way for new teachers to receive their induction support—or these new teachers could not be evaluated by PAR coaches.

These legislative developments had immediate and concrete results for PAR implementation in districts where new teachers had been included in the program. Coaches’ caseloads could still be made up of both new and veteran teachers, but coaches could no longer conduct summative evaluations of new credentialed teachers. (They could, however, continue to conduct evaluations of *uncredentialed* new teachers, as uncredentialed teachers were not included in the SB 2042 provisions.) In addition to these changes, at the same time as the SB 2042 standards were put in place, the state budget for PAR was cut by 75 percent.<sup>9</sup> Districts were then allowed to roll leftover money originally allocated for PAR into the general fund.<sup>10</sup> What remained, and what remains today—still called “PAR”—is certainly narrower in scope than its original incarnation.

## Key Issues Involved in Redesigning Teacher Observation

The cost of implementing stronger models of teacher observation, the implications for giving teachers more authority in the quality control of their peers, and the current reality of labor statutes are all issues for districts and states to address.

### Cost

A common concern about improving teacher evaluation, and specifically observation, is cost. This is not an unjustified concern; time costs money. The main cost involved with PAR, for example, is the replacement cost of teachers who leave the classroom to become coaches, with additional minor costs including stipends and substitutes for teachers on the PAR panel and perhaps release days for classroom teachers to observe other teachers. Districts interested in having two observers—the principal and a second observer—face similar additional costs if the second observer is a teacher. This cost may be reduced if the second observer is a retired principal or district administrator, but this strategy will likely mean lower legitimacy in the eyes of those being observed.

The financial cost of a redesigned teacher observation system must be compared to the cost of teacher observation as currently conducted. At present teacher observation typically takes one-half to one full day of principal time for each probationary teacher each year, as well as an additional half

day of clerical time. There is a corollary figure, albeit lower, for satisfactory tenured teachers. Factor in any current expenses for induction and mentoring programs that could be terminated or folded into revamped systems of observation. Beyond this, the legal costs for removing an unsatisfactory veteran teacher typically amount to hundreds of thousands of dollars, depending on the state, and it usually takes three to six years for the litigation to run its course, which often dissuades administrators from pursuing dismissals.

A comprehensive cost comparison between different types of observation programs is challenging, given the potential for such programs to affect various aspects of human capital management. Effective peer review programs, however, have been found to reduce litigation costs associated with terminating tenured teachers, because the teachers union is involved in the process and a copious amount of data has been collected. Programs that weed out weak teachers while they are probationary avoid the expense of termination later, after the teachers become tenured. Peer review and other strong mentoring programs have been shown to improve retention, avoiding the expense of recruiting, hiring, and orienting even more new teachers. These cost savings are real, even if they are hard to measure.<sup>11</sup>

### *Teacher professionalization*

Including teachers in the evaluation of other teachers holds the potential to fundamentally shift the prevailing hierarchical organization of K-12

public education. Admittedly, this may not be the goal of policymakers and so-called reformers. Given the enormous current policy attention on teacher evaluation, however, *not* considering the larger context of teaching work in which teacher evaluation sits would be to miss a profound opportunity. We have an organization of schools created during the industrial era of the early 20<sup>th</sup> century and modeled on factories. If we really want to transform the quality of teaching in the United States, what we need is not merely stronger quality control (more accurate ratings and perhaps more dismissals), but fundamental reform of teaching work in order to create professional organizations. Responsibility for defining, measuring, and enforcing professional standards is central to the way all fields of work are organized into occupations or professions and to the current debate about teacher evaluation as well.

In a national survey that asked principals how much control they have in regard to a number of the most important decisions in schools, the principals reported holding a high level of control over teacher evaluation decisions more often than over any other decisions.<sup>12</sup> Administrator control over teacher evaluation is widely taken for granted, but it is in fact a legacy of the industrial era's efficiency movement. In the early 1900s, the architects of our current public education system largely viewed teaching to be a routine activity involving only the knowledge needed to follow a textbook, rather than complex work involving judgment. They standardized and routin-

---

ized education, compartmentalizing it into grades and subjects, and created layers of (mostly male) management to supervise (mostly female) teachers.

The limitations of this view of teaching work—and the subsequent way of organizing the work—lead directly to our present problems with low teaching quality. Not only does this particular way of organizing teaching work grow out of a conception of teaching as low-skilled, but it supports and reinforces it. Teachers in such a system typically spend all day with children and lack meaningful professional opportunities to build their knowledge and skills. Moreover, a view of teaching as low-skilled, routinized work leads directly to routinized *assessment* of that work, focused on observations and ratings using the “proper” instrument and the “proper” number of classroom visits. If we were instead to view teaching as professional work involving complex judgment, we would be forced to engage the observer in ongoing discussion with the classroom teacher in order to make determinations about the teacher’s quality of practice, the teacher’s reflection on that practice, and his/her likelihood of improvement.

Control of gatekeeping (who enters) and quality (who stays) is the *key factor* that distinguishes professions from occupations. Expanding the role that teachers play in the evaluation of their peers has the potential to professionalize a career in teaching.<sup>13</sup>

As the states and districts that were successful in the “Race to the Top”

competition design new evaluation systems, one common feature has been the inclusion of peer observers. Recognizing that an increased number of observations improves reliability and that principals have limited time, the use of teachers as additional observers makes sense from a utilitarian perspective. Many of these new systems are missing a rich opportunity, however, by not giving teachers a more professional role in the process. By merely using teachers as observers who report a rating to administrators (who then determine outcomes), rather than creating the collective responsibility for professional standards that is central to professions, accountability for quality is not truly shared. These systems are distributing the task of observation but not the consequential decision-making of leadership. As a result, they are likely not fully leveraging the opportunity to re-organize teaching work, along with schools as organizations.

Master teachers have the grade level and content area expertise to conduct observations and provide useful feedback. They can be given adequate time for observation far more easily than principals. They can be trained in observation as easily as (if not more so than) principals. Admittedly, however, expert teachers have not often been eager to engage in the quality control of their peers. Far more educators—teachers and administrators alike—are willing to assign responsibility to teachers for assessing whether teaching standards are being met than to assign them responsibility for removing teachers not meeting

standards from the classroom. In fact, however, these opinions shift radically once people experience programs like PAR in practice.

### **Labor statutes**

Given the potential benefits of district/union collaboration, one key task for state policymakers is to clarify a lingering labor relations issue for school boards and local unions. In 1980, the United States Supreme Court ruled in favor of Yeshiva University in New York and against the National Labor Relations Board (NLRB) that faculty members at the university did not have the right to bargain collectively because they were, in effect, managers who set policy. Under the National Labor Relations Act (NLRA) exceptions are made for “managerial” and “supervisory” behavior. In *Yeshiva*, the NLRB argued that faculty could not be considered managerial because the University’s Board of Trustees held ultimate authority for policy decisions. The Court, however, held that the University’s Board upheld the faculty’s recommendations on matters of hiring and tenure, curriculum, and so forth in the “overwhelming majority” of cases, and therefore that the faculty acted in a managerial capacity.<sup>14</sup>

Partly as a result, K–12 public school teachers and their unions have often been reluctant to extend their leadership into such realms as, for example, the evaluation of other teachers, for fear of having their collective bargaining rights taken away. It is not entirely clear why this should be so. The National Labor Relations Act

pertains to private sector, not public, employees. Labor relations between school districts and their employees are governed not by the NLRA but by state and local statutes. Although local lawmakers could choose to be influenced by the Court’s rationale in *Yeshiva*, they need not be.

Schools are well served when leadership is shared by teachers and administrators. Yet the *Yeshiva* decision—whether relevant or not for public elementary and secondary school teachers—leaves some teachers worried that they must choose between their right to bargain collectively and their desire to assume leadership roles in educational improvement efforts.

For the reforms described here to flourish—namely the reliance on expert teachers for observation and assessment of other teachers—the statutes governing district labor relations must address these concerns. They must clearly protect teachers’ rights to bargain collectively even as teachers engage more substantively in local policy decisions and implementation.

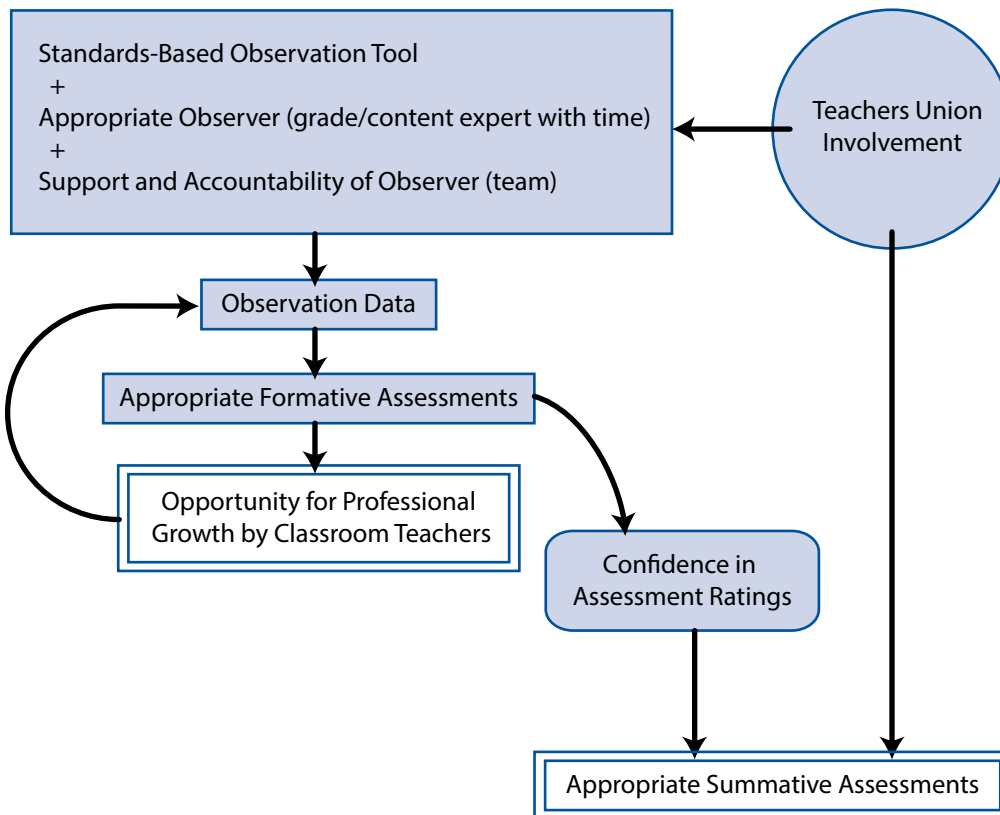
This is, to some degree, what happened in Ohio and California with those states’ forays into peer review. Ohio changed the relevant teacher-bargaining statutes to legalize the evaluation of members of a bargaining unit by other members of the same unit after Toledo, Cincinnati, and Columbus had

implemented peer review. California, rather than alter this aspect of collective bargaining law, created a separate law focused on peer review, which held that “a member of a bargaining unit who evaluated another member of that same unit remained a unit member and could not be declared a supervisor.”<sup>15</sup>

### Conclusion

No single policy or reform approach is a panacea for systemic weaknesses in teacher evaluation. The design elements presented here are no exception. They are, however, concrete steps that can be taken in a relatively short amount of time to radically improve teacher observation—one key com-

**FIGURE 3.** Key Design Elements of Observation-Based Assessment of Teaching Practice



---

ponent of reformed teacher evaluation systems.

Figure 3 provides an overview of the key components of district systems for observation-based assessment outlined in this brief.

As displayed in Figure 3 and discussed earlier, the four key design elements of effective teacher observation are:

- Use standards-based instruments for data collection;
- Rely on observers/assessors other than building administrators, ideally master teachers, to conduct observations;
- Support observers by establishing shared responsibility and accountability for evaluations and employment decisions; and
- Partner with the teachers union.

I have argued that these design elements, implemented well, can lead to:

- quality observation data, appropriate formative assessments, and opportunities for professional growth by classroom teachers, and
- confidence in assessment ratings and appropriate summative assessments.

In the realm of teacher observation, good instruments alone do not generate good data and good assessments. If we do not also pick the right observers who bring the right expertise, train the observers, give them the time they need to conduct good observation-based

evaluations, put a team in place to support them and hold them accountable, and involve the teachers union, what we will have are good instruments and not good systems of teacher observation. In other words, we must design effective systems of teacher observation in which good observation instruments operate.

The good news has long been that California has a strong infrastructure for observation based on BTSA and the CSTP. California's teaching standards are not perfect, but the concept of teaching standards is not foreign in California, and this puts us ahead of many states. We can build on this foundation by expecting more targeted and personalized diagnostic feedback, and by breaking down the firewall between formative and summative assessment, giving the master teachers who observe the teaching and provide the formative assessments more authority in the summative assessment process.<sup>16</sup> No doubt, such reforms will be unattractive to some teachers currently serving as mentors who have no interest in being involved in summative assessment. Normative shifts take time.

The unprecedented current policy attention paid to teacher evaluation has created a sense of urgency. This is good news for students and teachers alike, provided that urgency does not lead to the implementation of half-baked ideas and programs. Done well, a robust teacher observation system can contribute to policymakers' and the public's need for accountability,

while also providing a powerful tool for improving instructional practice. Successful teacher evaluation reform must ultimately increase the capacity of teachers while holding them accountable for performance. Teacher observation is a crucial piece of that puzzle, and we have policies in place and practices in use that can show the way forward.

## Endnotes

- 1 Kane & Staiger, 2012: 16.
- 2 Elmore, 2003.
- 3 Strunk, 2009; Koski & Tang, 2011.
- 4 Kerchner & Koppich, 1993.
- 5 In Toledo, consulting teachers provide summative assessments for beginning but not veteran teachers, while in other districts they provide it for both. In some districts with “PAR-light,” however, master teachers do not engage in summative assessments at all.
- 6 Goldstein, 2007a, 2010.
- 7 In most districts that implemented PAR prior to California’s legislation, the program began with beginning teachers as the less controversial part of the policy, and later expanded to veterans once the idea of teachers conducting teacher evaluations was established in a district. California’s Assembly Bill 1X, however, focused on veteran teachers while allowing for the inclusion of beginning teachers in the program.
- 8 Villaraigosa et al., 1999.
- 9 This reduction was part of a broader reduction in state funding for professional development programs that resulted from an increasingly constrained state fiscal situation. Four of the five main state-funded professional development programs were reduced from \$222 million in 2000–2001 to roughly \$62 million in 2003–2004 (Esch et al, 2005).
- 10 In the urban district that I studied over six years, the superintendent chose not to do so, in large part because administrators in the specially-designated high-needs schools being served by PAR reported on surveys that PAR was invaluable and should not be eliminated.
- 11 Kaboolian & Sutherland, 2005.
- 12 Ingersoll, 2003.
- 13 Goldstein, 2004, 2008.
- 14 Iorio, 1987: 100.
- 15 Koppich, 2006: 226.
- 16 See Goldstein, 2007b for a discussion of linking formative and summative assessment.

---

## References

Elmore, R. (2003). *Doing the right things, knowing the right thing to do: The problem of failing schools and performance-based accountability*. Cambridge, MA: Harvard Graduate School of Education and Consortium for Policy Research in Education.

Goldstein, J. (2004). Making sense of distributed leadership: The case of peer assistance and review. *Educational Evaluation & Policy Analysis*, 26(2), 173-197.

Goldstein, J. (2007a). Easy to dance to: Solving the problems of teacher evaluation with peer assistance and review. *American Journal of Education*, 113(3), 479-508.

Goldstein, J. (2007b). Debunking the fear of peer review: Combining supervision and evaluation and living to tell about it. *Journal of Personnel Evaluation in Education*, 18(4), 235-252.

Goldstein, J. (2008). Taking the lead: With peer assistance and review, the teaching profession can be in teachers' hands. *American Educator*, 32(3), 4-11, 36-38.

Goldstein, J. (2010). *Peer review and teacher leadership: Linking professionalism and accountability*. New York, NY: Teachers College Press.

Goldstein, J., & Noguera, P. (2006). A thoughtful approach to teacher evaluation. *Educational Leadership*, 63(6), 31-37.

Ingersoll, R. (2003). *Who controls teachers' work? Power and accountability in America's schools*. Cambridge, MA: Harvard University Press.

Iorio, J. E. (1987). The relevance of Yeshiva to public education. *Journal of Law and Education*, 16, 95-103.

Kaboolian, L., & Sutherland, P. (2005). *Evaluation of Toledo Public School District Peer Assistance and Review Plan*. Unpublished Report, Harvard University, John F. Kennedy School of Government, Cambridge, MA.

Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Seattle, WA: Bill & Melinda Gates Foundation Measures of Effective Teaching (MET) Project.

Kerchner, C. T., & Koppich, J. E. (1993). *A union of professionals: Labor relations and education reform*. New York, NY: Teachers College Press.

Koppich, J. E. (2006). The as-yet unfulfilled promise of reform bargaining: Forging a better match between the labor relations system we have and the education system we want. In J. Hannaway & A. J. Rotherham (Eds.), *Collective bargaining in education: Negotiating change in today's schools* (pp. 203-227). Cambridge, MA: Harvard Education Press.

Koski, W. S., & Tang, A. (2011). *Teacher employment and collective bargaining laws in California: Structuring school district discretion over teacher employment*. (Policy Brief 11-2). Stanford,

CA: Policy Analysis for California Education.

Strunk, K. (2009). *Collective bargaining agreements in California school districts: Moving beyond the stereotype*. (Policy Brief 09-1). Stanford, CA: Policy Analysis for California Education.

Villaraigosa, A., Strom-Martin, V., & Alquist, E. K. (1999). California Assembly Bill 1X. Introduced January 19. Amended in Assembly February 12 and 18.

Policy Analysis for California Education  
Stanford Graduate School of Education  
520 Galvez Mall, CERAS Rm. 401  
Stanford, CA 94305-3001  
(650) 724-2832  
<http://www.edpolicyinca.org>



We would like to thank the California Education Policy Fund (a sponsored project of Rockefeller Philanthropy Advisors), the Dirk and Charlene Kabcenell Foundation, the Noyce Foundation, and the Stuart Foundation for financial support for the publication of this policy brief. The views expressed are those of the author, and do not necessarily reflect the views of PACE or its funders.

---

## Recent PACE Publications

---

- Heather J. Hough and Susanna Loeb. Can a District-Level Teacher Salary Incentive Policy Improve Teacher Recruitment and Retention? Policy Brief 13-4, August 2013.
- Charles Taylor Kerchner. Education Technology Policy for a 21st Century Learning System. Policy Brief 13-3, May 2013.
- Mike Kirst. The Common Core Meets State Policy: This Changes Almost Everything. Policy Memorandum, March 2013.
- Svetlana Darche and David Stern. Making it Real: How High Schools Can Be Held Accountable for Developing Students' Career Readiness. Policy Brief 13-2, February 2013.
- Mary Perry. School Finance Reform – Can It Support California's College- and Career-Ready Goal? Report 2, February 2013.
- Morgan S. Polikoff and Andrew McEachin. Fixing the Academic Performance Index. Policy Brief 13-1, January 2013.
- Dominic J. Brewer, David N. Plank, Michelle Hall. How Californians Feel about Public Education: Results from the PACE/USC Rossier August 2012 Poll. September 2012.
- William Welsh, Erin Coghlan, Bruce Fuller, Luke Dauter. New Schools, Overcrowding Relief, and Achievement Gains in Los Angeles – Strong Returns from a \$19.5 Billion Investment. Policy Brief 12-2, August 2012.
- Robert Linqunti and Kenji Hakuta. How Next-Generation Standards and Assessments Can Foster Success for California's English Learners. Policy Brief 12-1, July 2012.
- Mary Perry. School Finance Reform – A Weighted Pupil Formula for California. Report 1, May 2012.
- Getting Down to Facts: Five Years Later. May 2012.
- Hilary McLean. California's Early Assessment Program: Its Effectiveness and the Obstacles to Successful Program Implementation. March 2012.
- Michal Kurlaender, Eric Grodsky, Samuel J. Agronow, Catherine L. Horn. State Standards, the SAT, and Admission to the University of California. Policy Brief 11-3, November 2011.
- The Road Ahead for State Assessments. Research Report, May 2011. PACE and Rennie Center for Education Research & Policy.