



## **Creating a Comprehensive System for Evaluating and Supporting Effective Teaching**

By Linda Darling-Hammond  
With the assistance of Channa Cook,  
Ann Jaquith, and Madlene Hamilton

Stanford University

This study was conducted by the Stanford Center for Opportunity Policy in Education (SCOPE) with support from the Ford Foundation and the Sandler Foundation.

© 2012 Stanford Center for Opportunity Policy in Education. All rights reserved.

The Stanford Center for Opportunity Policy in Education (SCOPE) supports cross-disciplinary research, policy analysis, and practice that address issues of educational opportunity, access, equity, and diversity in the United States and internationally.

Suggested citation: Darling-Hammond, L. (2012). *Creating a comprehensive system for evaluating and supporting effective teaching*. Stanford, CA. Stanford Center for Opportunity Policy in Education.



**Stanford Center for Opportunity Policy in Education**

Barnum Center, 505 Lasuen Mall

Stanford, California 94305

Phone: 650.725.8600

[scope@stanford.edu](mailto:scope@stanford.edu)

<http://edpolicy.stanford.edu>

## Acknowledgements

This initial draft of this paper was written for and benefited from feedback offered by participants in the Good Schools Seminar Series, supported by the Albert Shanker Institute, a nonprofit, nonpartisan organization endowed by the American Federation of Teachers to promote excellence in public education. The Good Schools Seminar is an effort to build a network of union leaders, district superintendents, and researchers to work collaboratively on improving public education through a focus on teaching.

We are particularly thankful for the assistance of: Barbara Byrd Bennett, former Chief Academic and Accountability Auditor (Superintendent), Detroit Public Schools; Larry Carter, Jr., President, United Teachers of New Orleans; Leo Casey, Vice President for Academic High Schools, United Federation of Teachers; Tom Forkner, President, Anderson Federation of Teachers; Dal Lawrence, former President, Toledo Federation of Teachers; Francine Lawrence, Executive Vice President, American Federation of Teachers; Daniel J. Montgomery, President and Chief Operating Officer, Illinois Federation of Teachers; Jody Papini, Douglas County Federation of Teachers; Mary Cathryn Ricker, President, St. Paul Federation of Teachers; Mary Ronan, Superintendent, Cincinnati Public Schools; Nathan Saunders, President, Washington Teachers Unions; Julie Sellers, Cincinnati Federation of Teachers; Rod Sherman, President, Plattsburgh Teachers Association; Brenda Smith, President, Douglas County Federation of Teachers; and Zelda Smith, New Orleans Public Schools. Eugenia Kemble, executive director of the Shanker Institute and Joan Baratz Snowden, consultant to the American Federation of Teachers, also offered helpful comments on the document.

In addition, we are grateful for useful feedback and advice from Barnett Berry, president of the Center for Teaching Quality; Gail Connelly, executive director of the National Association of Elementary School Principals; Dan Domenech, executive director of the American Association of School Administrators; Adriane Dorrington, senior policy analyst, and Rebecca Pringle, secretary-treasurer of the National Education Association; Jim Kohlmoos, executive director of the National Association of State Boards of Education (NASBE), and NASBE staff members Liz Ross and Patty Yoo; Janice Poda, education workforce strategic initiative director, and Kathleen Paliokas, InTASC program director, at the Council for Chief State School Officers; Adam Urbanski, director of the Teacher Union Reform Network; and Arthur Wise, president emeritus of the National Council for Accreditation of Teacher Education.

Support for this work was provided by the Ford Foundation and the Sandler Foundation through grants to the Stanford Center for Opportunity Policy in Education. We are grateful for their investment in this work.

# Table of Contents

Executive Summary .....	i
Introduction .....	1
Understanding <i>Teacher Quality</i> and <i>Teaching Quality</i> .....	2
A Systemic Approach to Teacher Evaluation and the Support of Effective Teaching.....	4
1. <i>Start with Standards</i> .....	5
2. <i>Create Performance-Based Assessments</i> .....	6
3. <i>Build a Standards-Based System of Local Evaluation</i> .....	15
4. <i>Create Structures to Support High-Quality, Fair, and Effective Evaluation</i> .....	28
5. <i>Create Aligned Professional Learning Opportunities</i> .....	33
Conclusion .....	39
Endnotes .....	40
Sidebars	
<i>New Mexico’s Standards-Based Teacher Evaluation System</i> .....	8
<i>Teacher Performance Assessment Elementary Literacy Portfolio</i> .....	11
<i>Standards-Based Evaluation in San Mateo, California</i> .....	17
<i>Use of Student Learning Evidence in Amphitheater, Arizona</i> .....	27
<i>Massachusetts’ Multiple Measures System of Evaluation</i> .....	29
<i>Criteria for Using Measures of Student Learning in Teacher Evaluations</i> .....	30
<i>Peer Assistance and Review</i> .....	34
<i>Criteria for an Effective Teacher Evaluation System</i> .....	38

## Executive Summary

Virtually everyone agrees that teacher evaluation in the United States needs an overhaul. Existing systems rarely help teachers improve or clearly distinguish those who are succeeding from those who are struggling. The tools that are used do not always represent the important features of good teaching. Criteria and methods for evaluating teachers vary substantially across districts and at key career milestones—when teachers complete pre-service teacher education, become initially licensed, are considered for tenure, and receive a professional license.

A comprehensive system should address these purposes in a coherent way and provide support for supervision and professional learning, identify teachers who need additional assistance and—in some cases—a change of career, and recognize expert teachers who can contribute to the learning of their peers.

This report outlines an integrated approach that connects these goals to a teaching-career continuum and a professional development system that supports effectiveness for all teachers at every stage of their careers.

### The Distinction between Teacher Quality and Teaching Quality

In the context of the current interest in measuring teacher effectiveness, it is important to distinguish between *teacher* quality and *teaching* quality. *Teacher quality* might be thought of as the bundle of personal traits, skills, and understandings an individual brings to teaching, including dispositions to behave in certain ways. *Teaching quality* refers to strong instruction that enables a wide range of students to learn. Teaching quality is in part a function of teacher quality—teachers' knowledge, skills, and dispositions—but it is also strongly influenced by the context of instruction: the curriculum and assessment system; the “fit” between teachers' qualifications and what they are asked to teach; and teaching conditions, such as time, class size, facilities, and materials. If teaching is to be effective, policymakers must address the teaching and learning environment as well as the capacity of individual teachers.

### Elements of a Teaching and Learning System

A high-quality *system* should create a coherent, well-grounded approach to developing teaching, crafted collectively by state and district leaders with teachers and their representatives. In addition to clear standards for student learning, accompanied by high-quality curriculum materials and assessments, this system should include five key elements:

1. **Common statewide standards** for teaching that are related to meaningful student learning and are shared across the profession: Standards for student learning can focus teachers' work and learning. They also permit aligned standards for teaching, like those created by the National Board for Professional Teaching Standards (NBPTS), for accomplished veteran teachers, and the Interstate New Teacher Assessment and Support Consortium (INTASC), for beginning teachers. These should guide preparation, licensing, on-the-job evaluation, and ongoing professional learning.

2. **Performance assessments, based on statewide standards, guiding state functions** such as teacher preparation, licensure, and advanced certification: Well-designed performance-based assessments, like those created by NBPTS and states like Connecticut and California, have been found both to *predict* and *develop* greater effectiveness. Such assessments—which look directly at teachers’ abilities to plan, teach, and assess learning—should be used to make key decisions regarding entry into the profession, readiness to be professionally licensed, and recognition of expertise. In a tiered system, these kinds of assessments can create a career continuum that fosters career development from early preparation to instructional leadership.
3. **Local evaluation systems aligned to the same standards**, which assess on-the-job teaching based on multiple measures of teaching practice and student learning: Such local evaluation systems should have three components, considered in relation to one another, in an integrated fashion:
  - **Standards-based evaluations of practice** based on observations, curriculum plans, assignments, and assessments revealing teachers’ classroom practice.
  - **Evidence of teachers’ contributions to the work of their colleagues and the school as a whole.** Since student learning gains are a function of teachers’ collective efforts, these valuable contributions should be part of the evaluation process.
  - **Evidence of teacher’s contributions to student learning** based on multiple sources of information reflecting classroom work and other assessments that are appropriate and valid for the curriculum and for the students being taught.
4. **Support structures** to ensure trained evaluators, mentoring for teachers who need additional assistance, and fair decisions about personnel actions. These structures should include, at minimum:
  - trained, skilled evaluators with deep knowledge of teaching and learning
  - supports, including mentoring, for teachers needing assistance
  - governance structures that enable sound personnel decisions
  - resources to sustain and monitor the system
5. **Aligned professional learning opportunities** that support the improvement of teachers and teaching quality: These should link both formal professional development and job-embedded learning opportunities to the evaluation system. Evaluations should trigger continuous goal-setting for areas teachers want to work on, high-quality professional development supports and coaching, and opportunities to share expertise.

To transform systems, incentives should be structured to promote collaboration and knowledge sharing, rather than competition, across organizations. Knowledge-sharing is needed to develop not only learning organizations but a learning-oriented system of education in which ongoing evaluation and inquiry into practice are stimulated within and across classrooms, across schools partnered within regions, and within the system as a whole.

## Measuring Student Learning

There is agreement that new teacher evaluation systems should look at teaching in light of student learning. One currently popular approach is to incorporate teacher ratings from value-added models (VAM) that use statistical methods to examine changes in student test scores over time. Unfortunately, researchers have found that:

- 1. Value-Added Models of Teacher Effectiveness Are Highly Unstable:** Teachers' ratings differ substantially from *class to class* and from *year to year*, as well as from one *test* to the next.
- 2. Teachers' Value-Added Ratings Are Significantly Affected by Differences in the Students Assigned to Them:** Even when models try to control for prior achievement and student demographic variables, teachers are advantaged or disadvantaged based on the students they teach. In particular, teachers with large numbers of new English learners and students with special needs have been found to show lower gains than the same teachers when they are teaching other students. Students who teach low-income students are disadvantaged by the summer learning loss their children experience between spring-to-spring tests.
- 3. Value-Added Ratings Cannot Disentangle the Many Influences on Student Progress:** —Many other home, school, and student factors influence student learning gains, and these matter more than the individual teacher in explaining changes in scores.

The limitations of value-added analysis do not mean that districts cannot include evidence of student learning in the evaluation process. Some districts use a variety of other measures of student learning in evaluations of teaching, such as evidence drawn from classroom assessments and documentation like the Developmental Reading Assessment; pre- and post-test measures of student learning in specific courses or curriculum areas (developed by individual teachers, departments, school faculty, or district faculty or staff); evidence of student accomplishments in relation to teaching activities, such as student science investigations, research papers, or art projects. Some districts use evidence from teachers' careful documentation of the learning of a set of diverse students over time, like that included in NBPTS portfolios.

The key is that the assessments be appropriate for the curriculum and the students being taught. This may mean the use of tailored assessments for certain students—such as English language proficiency tests for gauging the progress of new English learners—and the consideration of learning evidence in light of the teaching context.

## Criteria for an Effective Teacher Evaluation System

In conclusion, research on successful approaches to teacher evaluation suggests that:

- 1. Teacher evaluation should be based on professional teaching standards** and should be sophisticated enough to assess teaching quality across the continuum of development from novice to expert teacher.

2. **Evaluations should include multi-faceted evidence of teacher practice, student learning, and professional contributions** that are considered in an integrated fashion, in relation to one another and to the teaching context. Any assessments used to make judgments about students' progress should be appropriate for the specific curriculum and students the teacher teaches.
3. **Evaluators should be knowledgeable about instruction and well trained in the evaluation system**, including the process of how to give productive feedback and how to support ongoing learning for teachers. As often as possible, and always at critical decision-making junctures (e.g., tenure or renewal), the evaluation team should include experts in the specific teaching field.
4. **Evaluation should be accompanied by useful feedback, and connected to professional development opportunities** that are relevant to teachers' goals and needs, including both formal learning opportunities and peer collaboration, observation, and coaching.
5. **The evaluation system should value and encourage teacher collaboration**, both in the standards and criteria that are used to assess teachers' work, and in the way results are used to shape professional learning opportunities.
6. **Expert teachers should be part of the assistance and review process** for new teachers and for teachers needing extra assistance. They can provide the additional subject-specific expertise and person-power needed to ensure that intensive and effective assistance is offered and that decisions about tenure and continuation are well grounded.
7. **Panels of teachers and administrators should oversee the evaluation process** to ensure that it is thorough and of high quality, as well as fair and reliable. Such panels have been shown to facilitate more timely and well-grounded personnel decisions that avoid grievances and litigation. Teachers and school leaders should be involved in developing, implementing, and monitoring the system to ensure that it reflects good teaching well, that it operates effectively, that it is tied to useful learning opportunities for teachers, and that it produces valid results.

Initiatives to measure and improve teaching effectiveness will have the greatest payoff if they stimulate practices known to support student learning and are embedded in systems that also *develop* greater teaching competence. In this way, policies that create increasingly valid measures of teaching effectiveness—and that create innovative systems for recognizing, developing and utilizing expert teachers—can ultimately help to create a more effective teaching profession.

# Creating a Comprehensive System for Evaluating and Supporting Effective Teaching

## Introduction

*I have had administrators who never came into my classroom for formal observations or asked me for anything more than the initial planning/goal sheet. I have had administrators observe a formal lesson and put the feedback sheet in my box without ever having spoken to me about the lesson, and I have had years where I am just asked to sign the end-of-the-year evaluation sheet [without being observed].*

—Jane Fung, National Board Certified teacher and Milken Award Winner <sup>1</sup>

**V**irtually everyone agrees that teacher evaluation in the United States needs an overhaul. Existing systems rarely help teachers improve or clearly distinguish those who are succeeding from those who are struggling. The tools that are used do not always represent the important features of good teaching. It is nearly impossible for principals, especially in large schools, to have sufficient time or content expertise to evaluate all of the teachers they supervise, much less to address the needs of some teachers for intense instructional support. And many principals have not had access to the professional development and support they need to become expert instructional leaders and evaluators of teaching. Thus, evaluation in its current form often contributes little either to teacher learning or to accurate, timely information for personnel decisions.

Furthermore, criteria and methods for evaluating teachers vary substantially across schools and districts, and these are typically disconnected from the ways teachers are evaluated at key career milestones—when they complete pre-service teacher education, when they become initially licensed, and when they are tenured and receive a longer term professional license. Over the course of their careers, most teachers experience a cacophony of standards and directives—both in terms of *what* they are expected to teach and *how* they are expected to do so. In short, many states have no coherent system for evaluating and improving teaching, which makes it difficult to come up with effective solutions to the problems of teaching practice we face.

Today, much attention is focused on identifying and removing poor teachers. But what we really need is a conception of teacher evaluation as part of a *teaching and learning*

system that supports continuous improvement, both for individual teachers and for the profession as a whole. Such a system should enhance teacher learning and skill, while at the same time ensuring that teachers who are retained and tenured can effectively support student learning throughout their careers.

A highly skilled teaching force results from developing well-prepared teachers from recruitment through preparation through ongoing professional development. Support for teacher learning and evaluation need to be part of an integrated whole that promotes effectiveness during every stage of a teacher's career. Such a system must ensure that teacher evaluation is connected to—not isolated from—preparation and induction programs, daily professional practice, and a productive instructional context.

At the center of such a system are professional teaching standards that are linked to student learning standards, curriculum, and assessment, creating a seamless relationship between what teachers do in the classroom and how they are prepared and assessed. A productive evaluation system should consider teachers' practice in the context of curriculum goals and students' needs, as well as multi-faceted evidence of teachers' contributions to student learning and to the school as a whole.

Finally, a comprehensive system should address a variety of purposes: licensing, hiring, and granting tenure; support for supervision and professional learning; identification of teachers who need additional assistance and—in some cases—a change of career; and recognition of expert teachers who can contribute to the learning of their peers, both informally and as mentors, coaches, and teacher leaders. Some policymakers are also interested in tying compensation to judgments about teacher effectiveness, either by differentiating wages or by linking such judgments to additional responsibilities that carry additional stipends or salary.

This report outlines an integrated approach that connects these goals to both a teaching-career continuum and a professional development system that supports effectiveness for all teachers at every stage of their careers.

## Understanding *Teacher Quality* and *Teaching Quality*

In the context of the current interest in measuring teacher effectiveness, it is important to distinguish between *teacher quality* and *teaching quality*. *Teacher quality* might be thought of as the bundle of personal traits, skills, and understandings an individual brings to teaching, including dispositions to behave in certain ways. Research on teacher effectiveness, based on teacher ratings and student achievement gains, has found the following qualities to be important:

- strong content knowledge related to what is to be taught;
- knowledge of how to teach others in that area (content pedagogy) and skill in implementing productive instructional and assessment practices;

- understanding of learners and their development, including how to support students who have learning differences or difficulties, and how to support the learning of language and content for those who are not already proficient in the language of instruction;
- general abilities to organize and explain ideas, as well as to observe and think diagnostically; and
- adaptive expertise that allows teachers to make judgments about what is likely to work in a given context in response to students' needs.<sup>2</sup>

Most educators, parents, and policymakers would also include important dispositions in this list, such as the willingness to:

- support learning for all students,
- teach in a fair and unbiased manner,
- adapt instruction to help students succeed,
- strive to continue to learn and improve, and
- collaborate with other professionals and parents in the service of individual students and the school as a whole.

These qualities, supported by research on teaching, are embodied in the standards adopted by the National Board for Professional Teaching Standards and, at the beginning teacher level, by the states involved in the Interstate New Teacher Assessment and Support Consortium (INTASC), operating under the aegis of the Council of Chief State School Officers (CCSSO). This consortium of states has taken a leading role in developing beginning teacher standards and assessments that define a common knowledge base for teaching, based on research on development, learning, curriculum, and teaching.

As these standards have been built into licensing and preparation requirements in more than 40 states over the last decade, they have provided a means to develop a stronger foundation for effective teaching. Evidence from several studies suggests that teacher certification has become a stronger predictor of teacher effectiveness in states that have raised their standards.<sup>3</sup>

*Teaching quality* refers to strong instruction that enables a wide range of students to learn. Such instruction meets the demands of the discipline, the goals of instruction, and the needs of students in a particular context. Teaching quality is in part a function of teacher quality—teachers' knowledge, skills, and dispositions—but it is also strongly influenced by the context of instruction. Key to considerations of context are the curriculum and assessment systems that support teachers' work, the “fit” between teachers' qualifications and what they are asked to teach, and teaching conditions. An excellent

teacher may not be able to offer high-quality instruction in a context where she is asked to teach a flawed curriculum unsupported by appropriate materials or assessments. Similarly, a well-prepared teacher may perform poorly when asked to teach outside the field of his or her preparation or under poor teaching conditions—for example, without adequate teaching materials, in substandard space, with too little time, or to classes that are far too large.

Strong teacher quality may heighten the probability of effective teaching, but does not guarantee it. Initiatives to develop teaching quality and effectiveness must consider not only how to identify, reward, and use teachers' skills and abilities, but also how to develop teaching contexts that enable good practice. Thus, if teaching is to be effective, the policies that construct the learning environment and the teaching context must be addressed along with the qualities of individual teachers.

## A Systemic Approach to Teacher Evaluation and the Support of Effective Teaching

There is a growing realization that we need a more systemic approach to building teacher effectiveness. For example, in *Gearing Up: Creating a Systemic Approach to Teacher Effectiveness*, a recent task force of the National Association of State Boards of Education emphasized the importance of creating a more aligned system, beginning with recruitment and preparation and continuing through evaluation and career development.<sup>4</sup>

A high-quality teacher evaluation *system* should create a coherent, well-grounded approach to developing teaching, crafted collectively by state and district leaders with teachers and their representatives. In addition to clear standards for student learning, accompanied by high-quality curriculum materials and assessments, this system should include five key elements:

- 1) **Common statewide standards** for teaching that are related to meaningful student learning and are shared across the profession
- 2) **Performance assessments, based on these standards, guiding state functions** such as teacher preparation, licensure, and advanced certification
- 3) **Local evaluation systems aligned to the same standards**, for evaluating on-the-job teaching based on multiple measures of teaching practice and student learning
- 4) **Support structures** to ensure trained evaluators, mentoring for teachers who need additional assistance, and fair decisions about personnel actions

- 5) **Aligned professional learning opportunities** that support the improvement of teachers and teaching quality

Based on research and evidence of best practices, this paper outlines how each of these five elements should operate.

## 1. Start with Standards

“If you don’t know where you are going, any road will take you there,” observed the Cheshire Cat in *Alice in Wonderland*. So it is in education. Without a clear understanding of what students should learn and how teaching can support them, it is easy to wander aimlessly. Developing a shared vision of educational goals and supportive instruction is essential to a building a system that can support effective teaching.

The Common Core State Standards (CCSS), which most states have recently adopted, are one effort to achieve a more common vision of educational purpose. The CCSS seek to provide “fewer, clearer, and higher” expectations for learning across the grade levels in English language arts and mathematics. These standards are intended to provide guidance for understanding how students learn in a progressive fashion along skill strands, as well as what should be taught to enable them to be both college- and career-ready by the end of high school.

States that have not adopted the Common Core standards can accomplish these goals in state-level standards. The key is to establish a clear conception of the learning objectives and kinds of instruction that will support disciplinary (and interdisciplinary) understandings within and across content areas. Such conceptions need to be supported by thoughtful curriculum frameworks and materials, as well as assessments that meaningfully evaluate what students know and can do across the full scope of the standards.

The recently adopted CCSS and similar state standards in the United States are very similar to the core curriculum expectations articulated in countries like Finland, Japan, Singapore, and South Korea. These countries have produced a set of standards, curriculum supports, and associated assessments that provide an essential context for teacher development and evaluation. A major part of teachers’ ongoing professional learning takes place as they develop, in collaboration with their colleagues, the specific lessons and assessment tools they will use in the classroom. These collaborative learning opportunities become very analytic and intensive with the use of strategies like Lesson Study, action research about practice, or Learning Circles.

Agreement about learning goals for students allows teaching standards to be aligned. The National Board for Professional Teaching Standards has created benchmarks for how accomplished veteran teachers can enact the kind of learning envisioned by student learning standards. Similarly, the Interstate New Teacher Assessment and Support Consortium (INTASC) has revised its model licensing standards for beginning

teachers, which have been adopted by over 40 states, to reflect the kind of teacher knowledge, skills, and understandings needed to implement the CCSS and other standards for student learning.

These efforts to align learning standards with teaching standards are expressed in performance terms: what teachers should know and be able to do to support student learning, rather than merely how many hours they should sit in classes or workshops to gain credits. Such efforts have begun to focus teacher preparation and development on effective practice.

However, there has been remarkably little effort to connect these standards to districts' on-the-job evaluations of teachers. Local teacher evaluations vary widely across districts, and are often based on checklists of teacher behaviors that are not associated with effectiveness. Thus, teachers encounter a wide variety of disjointed signals over the course of a career, and opportunities to develop high-quality practice are missed at every turn. A comprehensive approach to teacher evaluation would create more useful assessments for state licensure and advanced certification, and would use these as a framework for more meaningful local evaluations that occur on the job.

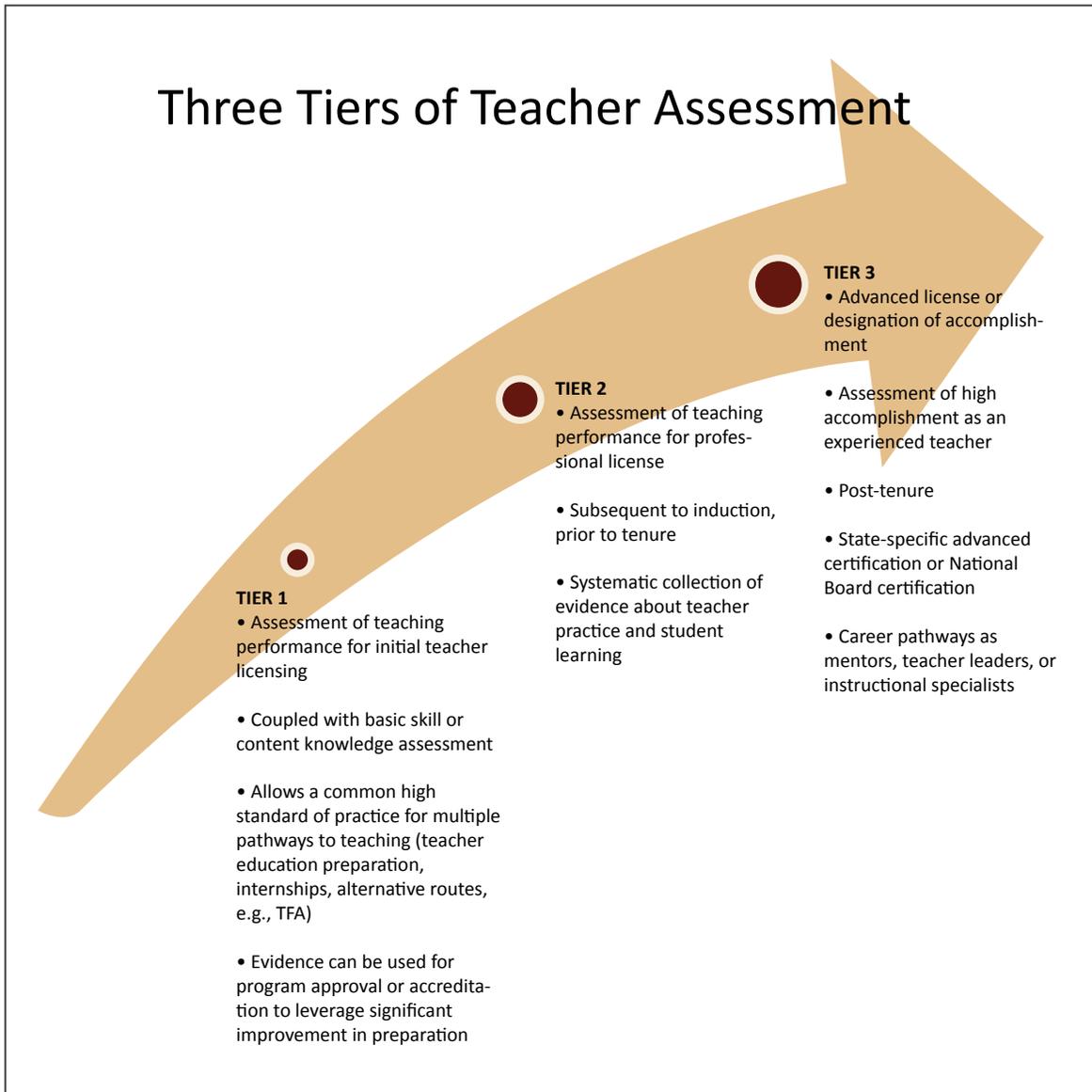
## 2. Create Performance-Based Assessments

Well-designed performance-based assessments have been found to measure aspects of teaching that are related to effectiveness, as measured by student achievement gains. These include standardized *teacher performance assessments* like those used for National Board Certification and for beginning teacher licensure in states like Connecticut and California, as well as *standards-based teacher evaluation systems* used in some local districts (discussed in section 3). The value of such assessments is that they can both *document* and help teachers develop greater effectiveness, as participation in these assessments supports learning both for teachers, who are being evaluated, and for educators who are trained to serve as evaluators.

Ideally, states would create a tiered licensure system that licenses new teachers and recognizes accomplished teachers based on their demonstrated performance. The system would frame a career continuum for professional learning and advancement to which local evaluations are aligned. State performance assessments of teaching would be used at Tier 1 for the initial teaching license that grants permission to practice (usually called the preliminary or probationary license), and again at Tier 2, as an indication that the induction period has been successfully completed and the teacher has reached a professional level of competence associated with the granting of the professional license. This generally happens before local districts take up the issue of granting tenure, and can inform the tenure decision.

Some states have a third tier of recognition that certifies accomplished practice and allows teachers to take on more extended professional roles, such as mentor or master

# Three Tiers of Teacher Assessment



teacher, and may carry additional compensation. This designation is sometimes tied to National Board Certification, and other times to state-specific assessments. In between these junctures, local evaluations of job performance would be based on the same standards, so that teacher development is coherent and consistent across the career.

New Mexico is an example of a state that has created such a system of both state performance assessments for licensure purposes and locally aligned evaluations for personnel purposes, with the help of teacher associations, teacher educators, and researchers. (See sidebar, pages 8–9.)

**License Teachers Based on Performance.** To leverage stronger preparation and teacher quality, states should make initial licensing decisions based on greater evidence of teacher

## New Mexico's Standards-Based Teacher Evaluation System

In 2003, New Mexico enacted a Three-Tiered Licensure system, linked to compensation and based on an independent review process for teachers. Teachers must demonstrate increased levels of competence in order to progress from Provisional Teacher (Level I, which comprises the initial three years) to Professional Teacher (Level II) to Master Teacher (Level III). Those who achieve each level receive greater compensation and undertake greater responsibilities.

In the authorizing legislation, HB212, legislators recognized the need for New Mexico to develop a high-quality teaching force, stating:

Unless the state and school districts find ways to mentor beginning teachers, intervene with teachers while they still show promise, improve the job satisfaction of quality teachers and elevate the teaching profession by shifting to a professional educator licensing and salary system, public schools will be unable to recruit and retain the highest quality teachers in the teaching profession in New Mexico.

To advance from one licensure level to the next, teachers complete a Professional Development Dossier that provides evidence of performance along three dimensions: instruction, student learning, and professional learning. The dossier, which is modeled after the National Board Certification portfolio and includes teaching artifacts and samples of student work, is submitted electronically to a state evaluation board. Classroom evidence of student learning is presented by the teacher in relation to specific curriculum goals and instructional processes.

The evaluation of the dossier is conducted by two certified reviewers from outside the teacher's district, at least one of whom is certified in the submitting teacher's field. The reviewers are master teachers from across the state who are trained to meet a scoring reliability standard. Reviewer scores are monitored for consistency by a consulting partner organization, Resources for Learning, which is currently engaged in a joint effort with the University of New Mexico College for Education, New Mexico school districts, the Institute for Professional Development, and business and community partners to build an infrastructure for professional development to improve student achievement. Thus, the evaluation system and the professional development system are designed to be mutually reinforcing.

The performance indicators at each licensure level include nine competencies that span three critical areas: instruction, student learning, and professional learning:

1. The teacher accurately demonstrates knowledge of the content area and approved curriculum.
2. The teacher appropriately utilizes a variety of teaching methods and resources for each area taught.
3. The teacher communicates with and obtains feedback from students in a manner that enhances student learning and understanding.
4. The teacher comprehends the principles of student growth, development, and learning, and applies them appropriately.

*continued on next page*

5. The teacher effectively utilizes student assessment techniques and procedures.
6. The teacher manages the educational setting in a manner that promotes positive student behavior and a safe and healthy environment.
7. The teacher recognizes student diversity and creates an atmosphere conducive to the promotion of positive student involvement and self-concept.
8. The teacher demonstrates a willingness to examine and implement change as appropriate.
9. The teacher works productively with colleagues, parents, and community members.

The performance expectations (indicators) describe the observable teacher behaviors required for a teacher to “Meet Expectations” for the competency area, and these indicators become more challenging at higher levels. The decision to advance a teacher from one licensure level to another is based upon both the dossier assessment and the school district’s recommendation for advancement. The district recommendation relies in large measure upon the running records of performance that are kept as part of a teacher’s annual Professional Development Plan.

### Local Teacher Evaluations

The local teacher evaluation process is aligned with the state licensure system. New Mexico defines the following purposes for teacher evaluation: to assist in identifying and building upon teachers’ strengths; to serve as the basis for the improvement of instruction; to enhance the implementation of programs and curriculum; to address accountability and teacher quality; and to support fair, valid, and legal decisions for rehire, promotion, or termination.

Both districts and teachers must create Professional Development Plans. Districts must develop a written teacher performance evaluation plan that meets state requirements, including evaluation instruments that measure performance against the standards, a system for data collection that includes classroom observations, a process for providing feedback, and training for teachers and administrators. Classroom observations may be supplemented by videotapes, written documentation of activities, portfolios, reflective journals, and instructional artifacts.

Ongoing, formative evaluation records are kept over the three-year Level I licensure period that precedes tenure and the Level II professional license. The documentation consists of a running record of authentic information about a teacher’s performance in the areas of instruction, student learning, and professional learning.

The teacher’s Professional Development Plan (PDP) is filed at the beginning of each school year. The teacher and principal establish measurable objectives for the nine teacher competencies. Together they develop a written plan that articulates goals (including the competencies and indicators to be addressed), an action plan, observable results, and a written reflection of the PDP (including an analysis of student achievement and learning growth). The principal observes the teacher during the year and the teacher collects evidence of accomplishment of the objectives. Before the end of the school year, the teacher and the principal meet to assess how well the PDP was carried out and the extent to which measurable objectives were achieved. This sets the stage for next year’s goals and for a process of continual improvement.

competence than merely completing a set of courses or surviving a certain length of time in the classroom. Since the 1980s, the desire for greater confidence in licensing decisions has led to the introduction of teacher licensing tests in nearly all states. However, these tests—generally multiple-choice tests of basic skills and subject matter—are not strongly predictive of teachers’ abilities to effectively teach children. Furthermore, in many cases, these tests evaluate teacher knowledge *before* they enter or complete teacher education, and hence are an inadequate tool for teacher education accountability.

Since the 1990s, several states have incorporated performance assessments in the licensing process. These measures of performance—which can provide data to inform the accreditation process—have been found to be strong levers for improving preparation and mentoring, as well as determining teachers’ competence. For example, the Performance Assessment for California Teachers (PACT) requires teachers to document their plans for a unit of instruction linked to the state standards, adapt them for special education students and English language learners, videotape and critique lessons, and collect and evaluate evidence of student learning. It is scored by school-based and university-based teacher educators, who are trained to produce reliable scores that are calibrated and audited. The Connecticut Beginning Educator Support and Training (BEST) assessment used a similar portfolio for granting the professional license for beginning teachers (year 2 or 3 in the profession).

Like the National Board assessments, beginning teachers’ ratings on the PACT and the BEST assessments have been found to predict their students’ achievement gains on state tests.<sup>5</sup> This form of *predictive validity* has not been established for traditional teacher tests, but is essential to making the claim that an assessment measures the right things on which to focus teachers’ attention and learning.

Currently, more than 25 states have joined together in a Teacher Performance Assessment Consortium<sup>6</sup> to create a common version of an initial licensing assessment, based on the work done in these states, which could be used nationwide to make preparation and licensing performance-based and grounded in teachers’ abilities to support student learning. This assessment, currently being piloted, is based on teaching standards that are linked to the CCSS, and will ultimately be embedded in states’ curriculum frameworks. The assessment ensures that teachers-in-training can plan, teach, and evaluate student learning effectively. (See sidebar on Teacher Performance Assessment, page 11.)

A more advanced version of the assessment could also be used at the point of the professional license (at the end of the probationary period), and to guide the mentoring process during the induction period. More than 40 states currently require some form of induction for beginning teachers, but these programs are rarely guided by a clear vision of what teachers should be able to do by the end of that period. Since the professional license is generally granted just before tenure decisions are made by local districts, this assessment could inform those decisions as well. States and districts that have adopted performance assessments to guide induction and decisions about licensing and tenure have supported

# TEACHER PERFORMANCE ASSESSMENT ELEMENTARY LITERACY PORTFOLIO

Assessment Components	Evidence Submitted
-----------------------	--------------------

## Task 1: Planning Instruction and Assessment

<ul style="list-style-type: none"> <li>▶</li> <li>▶</li> <li>▶</li> <li>▶</li> <li>▶</li> </ul>	<p>Provide relevant information about your instructional context.</p> <p>Select a learning segment of 3–5 sequential lessons that teach literacy skills and strategies and support students to comprehend and/or compose text</p> <p>Create an instruction and assessment plan for the learning segment that focuses on a key literacy concept and considers your students’ strengths and needs.</p> <p>Explain what you know about your students and the thinking behind your plans.</p> <p>Make daily notes about the effectiveness of your teaching for your students’ learning.</p>	<ul style="list-style-type: none"> <li>■ Information about the Learning Context</li> <li>■ Lesson Plans for Learning Segment               <ul style="list-style-type: none"> <li>■ Lesson plans</li> <li>■ Instructional Materials</li> <li>■ Assessment Tools/Procedures and Criteria</li> </ul> </li> <li>■ Planning Commentary</li> </ul>
---	---	---

## Task 2: Instructing and Engaging Students in Learning

<ul style="list-style-type: none"> <li>▶</li> <li>▶</li> </ul>	<p>Submit video clips from lessons where you engage your students to develop literacy strategies to comprehend and/or compose text.</p> <p>Analyze your teaching and your students’ learning in the video clip(s)</p>	<ul style="list-style-type: none"> <li>■ Video Clip(s)</li> <li>■ Instruction Commentary</li> </ul>
--	---	---

## Task 3: Assessing Student Learning

<ul style="list-style-type: none"> <li>▶</li> <li>▶</li> <li>▶</li> </ul>	<p>Analyze class performance from one assessment completed during the learning segment. Identify three student work samples that illustrate trends in student understanding within the class</p> <p>Select and analyze the learning of two focus students in more depth, and document your feedback on their work.</p> <p>Provide the assessment task and evaluation criteria.</p>	<ul style="list-style-type: none"> <li>■ Student Work Samples</li> <li>■ Evidence of Feedback</li> <li>■ Assessment Commentary</li> </ul>
---	--	---

## Task 4: Analyzing Teaching

<ul style="list-style-type: none"> <li>▶</li> </ul>	<p>Using notes you have recorded throughout the learning segment, respond to commentary prompts to explain what you have learned about your teaching practice and two or three things you would do differently if you could teach the learning segment over. Explain why the changes would improve your students’ learning.</p>	<ul style="list-style-type: none"> <li>■ Analyzing Teaching Commentary</li> </ul>
---	---	---

## Task 5: Academic Language in Literacy (evidence is gathered across tasks as noted)

<ul style="list-style-type: none"> <li>▶</li> <li>▶</li> </ul>	<p>Select one key language demand related to the literacy central focus. Explain how you will support students with varied language needs.</p> <p>Cite evidence of opportunities for students to understand and use the targeted academic language in: 1) the video clips from the Instruction task; OR 2) the student work samples from the Assessment task.</p> <p>Analyze the effectiveness of your language supports.</p>	<ul style="list-style-type: none"> <li>■ Planning Commentary</li> <li>■ Instruction Commentary</li> <li>■ Assessment Commentary</li> </ul>
--	---	--

much more purposeful and focused mentoring, with greater attention to a shared vision of good practice.

In states that have already used them, performance assessments of beginning teachers have been found not only to measure features of teaching associated with effectiveness, but actually to help develop effectiveness at the same time—both for the participants and for the programs that prepare them. Candidates report that they have learned more about teaching from completing the assessment. As one noted:

*For me, the most valuable thing was the sequencing of the lessons: teaching the lesson, and evaluating what the kids were getting, what they weren't getting, and having that be reflected in my next lesson ... the "teach-assess-teach-assess-teach-assess" process. You're constantly changing: You may have a plan or a framework, but you know that that has to be flexible, based on what the children learn that day.*

University and school faculty score these portfolios using standardized rubrics in moderated sessions following training, with an audit procedure to calibrate standards and ensure reliability. Faculties then use the PACT results to revise their curriculum. The scoring participants describe how this process creates a shared understanding of good teaching, focuses them on how to improve preparation, and creates a foundation for planning teacher induction and professional development.

*This [scoring] experience ... has forced me to revisit the question of what really matters in the assessment of teachers, which—in turn—means revisiting the question of what really matters in the preparation of teachers.*

—A teacher education faculty member

*[The scoring process] forces you to be clear about "good teaching"; what it looks like, sounds like. It enables you to look at your own practice critically, with new eyes.*

—A cooperating teacher

*As an induction program coordinator, I have a much clearer picture of what credential holders will bring to us and of what they'll be required to do. We can build on this.*

—An induction program coordinator

Teacher education programs receive detailed, aggregated data on all of their candidates by program area and dimensions of teaching, and use the data to improve their curriculum, instruction, and program designs. Using these aggregated data for accreditation will ultimately provide a solid basis for deciding which program models should be approved and expanded, and which should be closed if they cannot improve enough to enable most of their candidates to demonstrate that they can teach. With the addition of the incen-

tives for National Board Certification, these assessments would provide a continuum of measures that both identify and help stimulate increasing effectiveness across the career.

**Use National Board Certification to Recognize Accomplished Practice.** A standards-based approach to assessing teachers was initially developed through the work of the National Board for Professional Teaching Standards, which developed standards for accomplished teaching in more than 30 teaching areas defined by subject matter and developmental level of students. The Board then developed assessments that assemble evidence of teachers' practice and performance into a portfolio that includes video-tapes of teaching, accompanied by commentary, lesson plans, and evidence of student learning. These pieces of evidence are scored by trained raters who are expert in the same teaching field, using rubrics that define critical dimensions of teaching, such as planning based on knowledge of students' learning; instruction that uses effective strategies responsive to students' needs; and assessment and feedback that inform future plans and allow students to improve their work. Designed to identify experienced, accomplished teachers, National Board Certification has been implemented by a number of states as the basis for salary bonuses and other forms of teacher recognition, such as selection as a mentor or lead teacher.

A number of studies have found that the National Board Certification assessment process distinguishes teachers who are more effective in raising student achievement from others who do not achieve certification.<sup>7</sup>

Equally important, many studies have found that teachers' participation in the National Board process supports their professional learning and stimulates changes in their practice. Teachers note that the process of analyzing their own and their students' work in light of standards enables them to better assess student learning. It also helps them to evaluate the effects of their own actions and change them when necessary. Finally, teachers often develop new practices that are called for in the standards and assessments.<sup>8</sup> Teachers report significant improvements in their performance in each area assessed—planning, designing, and delivering instruction; managing the classroom; diagnosing and evaluating student learning; using subject matter knowledge; and participating in a learning community—and observational studies have documented that these changes do indeed occur.<sup>9</sup>

Furthermore, school-wide participation in National Board Certification can help teachers build their collective effectiveness. For example, the turnaround strategy at once-failing and now-much improved Mitchell Elementary School in Phoenix was to increase teacher expertise using the National Board certification process, supported by ESEA Title II funding. In this low-income Latino community, where most students are English language learners, more than 60% of the teachers—most of whom are from the community and reflect their student population—are either National Board Certified or in the process of earning certification. Mitchell teachers claim the National Board process transformed the school, as they have worked collectively to better understand

their teaching. Not only has the school's achievement dramatically improved, but teacher turnover is no longer a problem.<sup>10</sup> As the district's associate superintendent, Suzanne Zentner, noted, "We believe in the National Board Certification process as an approach to ... closing the achievement gap."<sup>11</sup> Other schools across the country have experienced similar successes using this strategy in recent years.<sup>12</sup>

**Create a Continuum of Teacher Effectiveness Throughout the Career.** If teachers are better supported and selected for tenure in the early years of the career, the prospects for developing a highly effective teacher corps will be much enhanced. Using tools like those described above, states could create a tiered system of recognition and compensation using performance assessments at each of three junctures: entry into the profession (initial licensure), movement to professional licensure or tenure status, and the recognition of accomplished teaching that signals teachers who might also be tapped for leadership roles in the profession.

Progress has been made in creating career development systems that can recognize excellent teaching, reward it, and tap the expertise of such teachers on behalf of broader school improvements. Performance-based evaluation and career development systems have been created at the local level in districts ranging from Cincinnati, Ohio, to Denver, Colorado, to Rochester, New York. These initiatives generally have several features in common:

- All have been co-developed and collaboratively implemented with teachers.
- Typically, evaluations occur at several junctures as teachers move from their *initial license*, through a period as a *novice or resident teacher* under the supervision of a mentor, to designation as a *professional teacher* after successfully passing an assessment of teaching skills.
- *Tenure*, which grants due process rights is a major step that either accompanies or follows the achievement of professional teaching status and is tied to a serious decision after rigorous evaluation of performance in the first several years of teaching, incorporating administrator and peer review by expert colleagues.
- *Lead teacher status*—which triggers additional compensation and access to differentiated roles—may be determined by advanced certification from the National Board for Professional Teaching Standards and/or other evidence of performance through standards-based evaluation systems. Such systems both encourage and measure effective teaching, and can be combined with other evidence of desirable teacher practices and student learning to identify accomplished teachers.

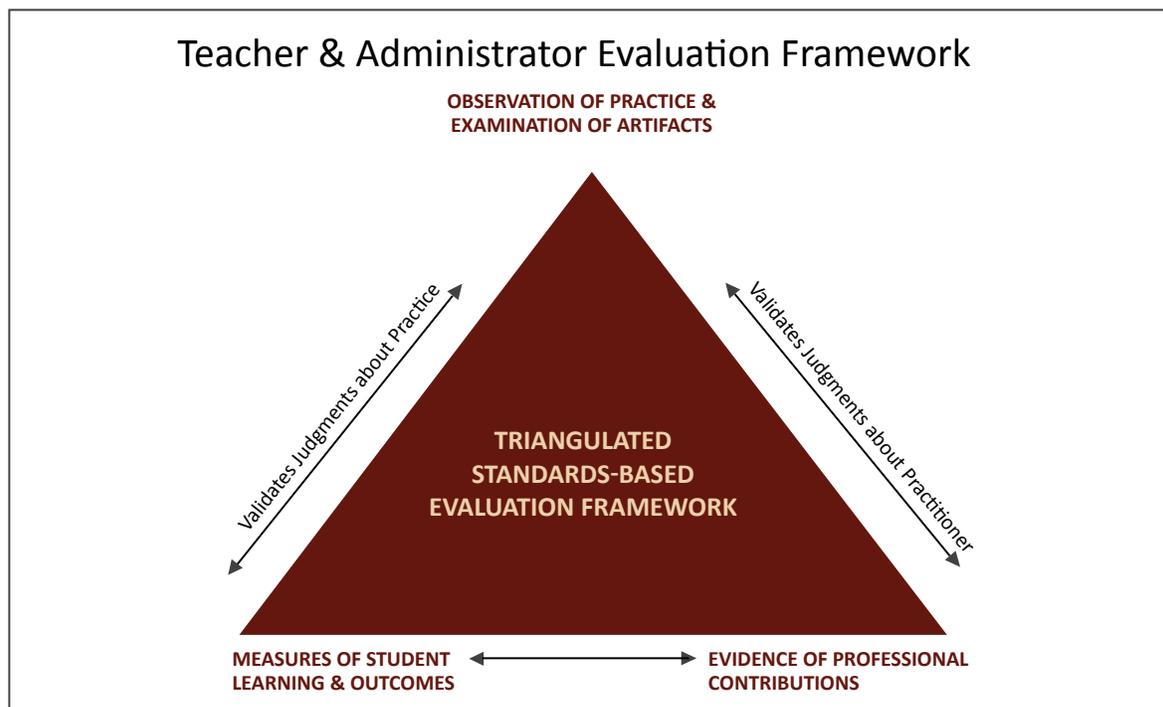
Where such systems have been put in place, evaluation systems have been carefully designed to provide a comprehensive picture of what teachers do, along with the results of those practices.

### 3. Build a Standards-Based System of Local Evaluation

If local school districts' teacher evaluation systems were grounded in the same standards as state licensing and certification systems, they could jointly reinforce teacher learning and development. This strategy has guided the work in New Mexico described earlier. Such local evaluation systems should have three components:

- 1. Standards-based evaluations of practice.** These evaluations should be based not only on observations or videotapes of classroom practice, but also an understanding of teachers' curriculum plans, assignments, and assessments in relation to the content they teach and the students they serve.
- 2. Evidence of teachers' contributions to the work of their colleagues and the school as a whole.** Since education is a team sport and student learning gains are a function of teachers' collective efforts, these contributions are valuable and should be part of the evaluation process.
- 3. Evidence of teachers' contributions to student learning.** Such evidence should be evaluated using multiple sources of information from both classroom documentation and other assessments, where these are appropriate and valid for the curriculum and for the students being taught.

These data should be considered in relation to one another, in an integrated fashion, as suggested by the Teacher and Administrator Evaluation Framework developed by the Massachusetts Teachers Association (below).<sup>13</sup> The three factors should not be separately weighted, as they are interdependent and must be considered in relation to the teaching context.



## Standards-Based Evaluations of Practice

Standards-based evaluations of practice used by some districts have been found to be significantly related to student achievement gains and to help teachers improve their practice and effectiveness.<sup>14</sup> Like the teacher performance assessments described above, these systems for observing teachers' classroom practice are based on professional teaching standards grounded in research on teaching and learning. They use systematic observation protocols, administered by trained evaluators, to examine teaching along a number of dimensions, such as classroom organization and management, planning a well-organized curriculum, making subject matter accessible to students, assessing student learning, and differentiating instruction to meet student needs.

These protocols generally provide indicators of teaching effectiveness associated with a set of professional standards that are concrete enough to guide observations and feedback to teachers—the standards describe practices shown by research to be associated with student learning. These protocols differ from open-ended forms that allow evaluators to determine, idiosyncratically, what is important in the classroom, and from old-style behaviorist approaches that list discrete teaching behaviors that may or may not support learning (bulletin boards are neat, teacher keeps a brisk pace of instruction, objectives are on the board). (See the example in the sidebar, page 17.)

In a study of three districts using standards-based evaluation systems, researchers found significant positive correlations between teachers' ratings and their students' gains in standardized test scores.<sup>15</sup> In the schools and districts studied, formative and summative assessments of teachers were based on well-articulated standards of practice assessed through observations of teaching, pre- and post-observation interviews, and, sometimes, artifacts such as lesson plans, assignments, and samples of student work.

Critical to the success of such systems is that they include multiple classroom observations by expert evaluators across the year. These systems look at multiple sources of data that reflect a teacher's instructional practice, and they provide timely and meaningful feedback to the teacher. In larger schools, where there may be as many as 100 teachers, it is impossible for such intensive evaluations to be conducted for every teacher every year by a lone school principal. (In business settings, for example, the appropriate span of control is generally considered to be 1 supervisor to 7 employees.) Districts that have successfully resolved the tension between the need for high-quality evaluation and principal time have typically included assistant principals, department chairs, and, sometimes, other master teachers as evaluators, or have created a cycle of more intensive evaluations in periodic years (for example, once every two or three years) for teachers who have been well rated.

The set of studies on standards-based teacher evaluation suggest that the more teachers' classroom activities and behaviors reflect professional standards of practice, the more effective they are in supporting student learning.<sup>16</sup> These kinds of results led one researcher to conclude that tying teachers' advancement and compensation to their

## Standards-Based Evaluation in San Mateo, California

The California Standards for the Teaching Profession guide initial teacher licensing in California as well as evaluation systems in many local districts. The standards address the following areas:

1. Engaging and supporting all students in learning
2. Creating and maintaining effective environments for student learning
3. Understanding and organizing subject matter for student learning
4. Planning instruction and designing learning experiences for all students
5. Assessing student learning
6. Developing as a professional educator

Each standard contains five subcategories, rated using a rubric with levels from unsatisfactory to exemplary. In the San Mateo school district, evaluation is based on both supervisors' observations and collection of evidence about each of the standards. To supplement what is observed by the evaluator about a specific standard, teachers are invited to include evidence from their lesson plans, assignments, samples of student work, test scores and other evidence of student learning, student self-assessments, student or parent communications or evaluations, or videotapes of classroom practice. An example from the first standard follows:

### 1. Engaging and Supporting All Students in Learning

Element	Unsatisfactory	Satisfactory	Accomplished	Exemplary
<b>1.1 The teacher builds on the students' prior knowledge, life experiences, and interests to achieve learning goals.</b>	Makes limited connections between the learning goals and students' prior knowledge, life experiences, and interests. Does not encourage student questions or comments during a lesson.	Makes acceptable connections between the learning goals and students' prior knowledge, life experiences, and interests. Elicits some questions from students during a lesson to monitor student understanding.	Makes substantial connections between the learning goals and students' prior knowledge, life experiences, and interests. Elicits and uses questions and comments from students during a lesson to extend their understanding.	Employs strategies that allow all students to connect and apply their prior knowledge, life experiences, and interests to new learning and the achievement of learning goals. Builds on students' questions and comments during lessons to modify instruction.
<b>1.2 The teacher uses a variety of instructional strategies and resources that respond to students' diverse needs.</b>	Uses limited instructional strategies, but they lack variety, are poorly carried out, or are inappropriate to the students or to the instructional goals. Few adjustments are made to respond to students' needs.	Uses a selection of instructional strategies that are largely appropriate to the students and the instructional goals. They may lack variety or be less than responsive to students' needs.	Uses a variety of instructional strategies that are appropriate to the students and the instructional goals. The teacher carries out these strategies thoughtfully, making some adjustments that are responsive to students' needs.	Makes skillful use of a wide repertoire of instructional strategies to engage all students in learning, making adjustments while teaching to respond to students' needs.
<b>1.3 The teacher facilitates challenging learning experiences for all students in environments that promote autonomy, interaction, and choice.</b>	Directs most learning experiences, permitting limited autonomy, interaction, or choice.	Directs some learning experiences, and permits some autonomy, interaction, and choice.	Facilitates learning experiences to promote construction interactions, autonomy, and choice, and to encourage and support student involvement in learning.	Provides opportunities for students to collaborate in a variety of constructive interactions, autonomy, and choice, in the pursuit of significant learning.

Teachers set goals each year and identify evidence that can be used to evaluate progress toward the goals. Ratings determine the intensity of future follow-up. A teacher with more than two unsatisfactory ratings enters the Peer Assistance and Review program and receives intensive mentoring and ongoing evaluation from an expert teacher, before a decision is made by a panel comprised of both teachers and administrators regarding continuation or dismissal. All teachers receive ongoing feedback in light of the standards. The San Mateo Union High School District teacher evaluation handbook can be found at: <http://smuhsd.ca.schoolloop.com/file/1224132524944/1257605326962/7291446584551382814.pdf>

knowledge and skills—and then using evaluation systems that help develop those skills—may ultimately produce more positive change in practice than evaluating teachers based primarily on student test scores.<sup>17</sup> This prediction has proved accurate: Several recent major studies have demonstrated that student achievement does not increase when teachers are evaluated and rewarded based only on whether their students' scores increase.<sup>18</sup>

Teachers' performance on standards-based evaluations can be supplemented by other measures, also considered in the context of the content and students being taught, including:

- **contributions to school-wide goals**, including work with colleagues, parents, and students; contributions to curriculum development and alignment; school improvement initiatives; and engagement in broader professional learning; and
- **contributions to growth in student learning** (from classroom assessments and documentation, student work samples, and school, district, or state standardized tests, when valid and appropriate).

### Contributions to School-wide Goals

It is important to remember that education is a team sport, and that successful schools raise achievement because they assemble the right mix of skills and abilities and enable people to work collaboratively.

Over 90% of the nation's teachers report that their colleagues contribute to their teaching effectiveness.<sup>19</sup> In one recent study, economists were able to quantify the student learning gains generated by the collective expertise of teams of teachers. They found that most value-added gains are attributable to teachers who are more experienced and better qualified, and who stay together as teams within their schools. The researchers found that peer learning among small groups of teachers seems to be the most powerful predictor of improved student achievement over time.<sup>20</sup> Another recent study found that students achieve more in mathematics and reading when they attend schools characterized by higher levels of teacher collaboration for school improvement.<sup>21</sup>

Thus, it is important for any evaluation system to document and recognize not only aspects of teachers' knowledge and skills but also their contributions to the work of the school as a whole. These contributions can include specific kinds of *knowledge and skills*, engagement in *shared instructional practices* or specific student supports, and support for *collegial learning and school improvement*.

**Knowledge and skills:** Schools need a mix of knowledge, skills, and abilities among their faculties to inform curriculum decisions and to meet the needs of their students.

Aside from the knowledge of content and pedagogy that teachers generally acquire in their certification area, specialized knowledge about the teaching of English language learners or the teaching of special education students may be highly desirable in many school contexts. Knowledge of the home languages students speak is also essential for communicating with parents as well as students, and is especially useful when it allows teachers to engage children and families and help their colleagues do so. Proficiency in using specific educational techniques, such as Reading Recovery, may be important in certain contexts, as it supports both students and colleagues in expanding their success.

This approach stands in contrast to efforts to evaluate and compensate teachers based directly on students' test scores, which can create a number of unintended dysfunctional consequences, as we describe in the next section. Odden and colleagues note:

Knowledge- and skills-based compensation systems provide a mechanism to link pay to the knowledge and skills (and by extension, performance) desired of teachers. . . . The concept of knowledge- and skills-based pay in education was adapted from the private sector, where it was developed to encourage workers to acquire new, more complex, or employer-specific skills. Knowledge- and skills-based pay was also intended to reinforce an organizational culture that values employee growth and development and to create a clear career path linked to increasing professional competence.<sup>22</sup>

Odden and colleagues offer several examples of knowledge- and skills-based evaluation and compensation plans.<sup>23</sup> For example, Coventry, Rhode Island, provides stipends for National Board Certification and for teachers to develop skills in authentic pedagogy, self-reflection, differentiated instruction, and family and community involvement—all strategies that have been linked through research to student achievement. Douglas County, Colorado offers compensation for completing blocks of courses associated with district goals, such as student assessment or teaching diverse learners. Vaughan Learning Center, a charter school in Los Angeles, California, offers compensation for relevant degrees and certification, as well as for specific knowledge and skills relevant to the school's mission, such as literacy training, training for teaching English as a second language, training for the inclusion of special education students, and technology.

**Shared instructional practices:** As schools seek to offer a more coherent approach to instruction, they need to encourage teachers to share practices, especially those which have a positive impact on student achievement. For example, the use of formative assessment to provide feedback to students and opportunities for them to revise their work have been found in many dozens of studies to have large effect sizes on student learning gains.<sup>24</sup> Teachers who teach students specific meta-cognitive strategies for reading, writing, and mathematical problem solving have been found to produce increased student learning of complex skills.<sup>25</sup> And so on.

In some systems, teachers receive recognition for demonstrating that they have implemented particular new practices associated with school-wide or district-wide goals,

such as the use of common literacy practices across classrooms, the use of formative assessments in planning and modifying instruction, or the implementation of a new system of writing instruction. Where possible, these practices are documented, along with evidence of how the changes have affected student participation and learning. In addition to specific teaching practices, a teacher might document how she increased student attendance or homework completion through regular parent conferences and calls home. She might show evidence of changes in these student outcomes, as well as improved grades, graduation, and college-going. The rationale for using these measures of effective teaching practices is that they support teacher development and improve the conditions for student learning.

**Contributions to the work of colleagues and the school as a whole:** Finally, the ways in which teachers may support their colleagues and the work of the school as a whole should be recognized in an evaluation system, as is done in nations like Singapore and many leading districts in the U.S. These contributions could include developing and sharing curriculum; supporting colleagues through peer observations, mentoring, or coaching; taking leadership roles in school-improvement initiatives; engaging in outreach to parents; and a variety of other activities that enable the school team to become more effective. In Singapore, these collegial activities are weighted heavily in the evaluation, and are taken as a sign of leadership that can place a teacher on one of several paths on the career ladder.<sup>26</sup>

### **Evidence of Student Learning**

If student learning is the primary goal of teaching, it appears straightforward that it ought to be taken into account in determining a teacher's competence. One prominent proposal is to use *value-added student achievement test scores* from state or district standardized tests as a key measure of teachers' effectiveness.

**Issues with value-added achievement measures:** However, researchers have offered many cautions about the problems of basing individual teacher evaluations on annual student test scores. Some issues are technical and may ultimately be addressed by changes in testing systems: For example, "vertically scaled," curriculum-specific tests that measure student growth along the entire achievement continuum (rather than just measuring standards within a single grade level) are not available in most states and teaching areas.

Furthermore, the narrowness of current tests, which focus on basic skills and use primarily multiple-choice questions, raises concerns about teaching to tests at the expense of other kinds of learning, such as writing, inquiry, and complex problem-solving. And indeed, studies have found that teachers' measured effectiveness differs significantly depending on the tests that are used.<sup>27</sup> Teachers who are rated highly on VAM estimates of achievement on basic skills tests are often not rated highly when more conceptually oriented tests are used.

Other concerns are more fundamental to the realities of teaching and schooling: There are difficulties in attributing student gains to specific teachers, and challenges in disentangling teacher effects from those of school and home conditions, as well as from other student factors. Among these influences on learning are multiple teachers, parents, tutors and out-of-school learning supports, and a variety of school factors.<sup>28</sup> As Henry Braun of the Educational Testing Service concluded in his summary of the research on value-added models:

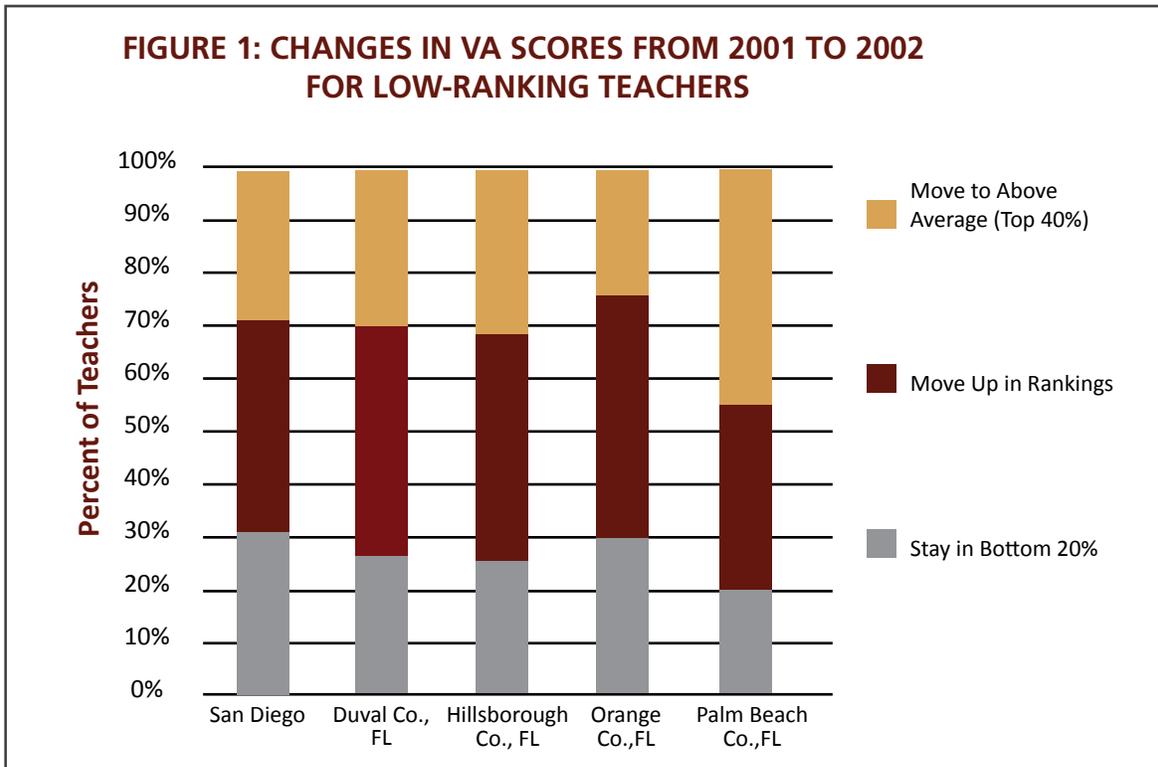
It is always possible to produce estimates of what the model designates as teacher effects. These estimates, however, capture the contributions of a number of factors, those due to teachers being only one of them. So treating estimated teacher effects as accurate indicators of teacher effectiveness is problematic.<sup>29</sup>

Indeed, the most optimistic estimates conclude that only about 7% to 10% of the overall variation in student achievement can be attributed to a student's individual teachers.<sup>30</sup> The largest influences, typically accounting for about 60% of the variance, are socio-economic factors associated with individual students and the collective composition of the classroom and school. The remaining variation is either a function of school factors other than the teacher—for example, the influences of specific curricula, the availability of useful learning materials, the time teachers have with students, class sizes, and the availability of instructional specialists—or is unexplained. In addition, on spring-to-spring measures of achievement like those offered by most state tests, the summer learning loss that substantially affects the scores of lower-income students also reduces their measured learning gains and is misattributed to their next year's teacher.<sup>31</sup>

Because of the many influences on student learning, studies have found that ratings of teacher effectiveness are highly unstable. One study of five large urban districts, for example, found that among top-ranked teachers (the highest 20%) in one year, only 25% to 35% were similarly ranked a year later, while a comparable proportion had moved to the bottom rankings. And among the lowest rated, only about 20% to 30% were similarly ranked a year later, while 25% to 45% moved to the upper tier of the rankings over the course of a year.<sup>32</sup> (See Figure 1, page 22.)

If value-added ratings were really measuring a teacher's basic competence or effectiveness, such wild swings would not occur. In fact, research has found that the test-score gains associated with teachers are affected by differences in the students who are assigned to them, even if those differences are supposed to be "controlled" in the statistical model. The same teacher appears more effective when he is teaching more advantaged students than when he is teaching students who experience educational challenges of various kinds, especially those who have special education needs or who are English language learners.<sup>33</sup>

Such distortions result in misestimates of teaching effectiveness and therefore create disincentives for teachers to teach those students who have the greatest needs. Such



disincentives could inadvertently reinforce historical practices of disproportionately assigning inexperienced teachers to the neediest students. They also exacerbate incentives for schools to push out high-need students.

To get a glimpse of how these factors operate, consider the results of a California study, which found that teachers’ value-added ratings were significantly related to their students’ race/ethnicity, income, language background, and parent education, despite the fact that these variables were “controlled” in the statistical models. In one example, the rating for an experienced English teacher jumped from the very lowest decile in one year to the very highest decile the next year. Between these two years, the proportion of English learners in her classroom dropped from nearly 60% to under 5%, and the proportions of Hispanic and low-income students also decreased while parent education levels increased. (See Figure 2, page 23.)

Similarly, in a study in Houston, Texas, where teachers are evaluated for dismissal and merit pay using a value-added system called EVAAS, teachers generally receive lower EVAAS ratings when they are teaching larger numbers of mainstreamed English learners. Some highly respected teachers with strong supervisory ratings have been dismissed after taking on such classes in the fourth grade, where students are first transitioned into mainstream classes. One of those dismissed had previously had exemplary ratings every year and was voted Teacher of the Year. As two teachers commented to researchers conducting the study:

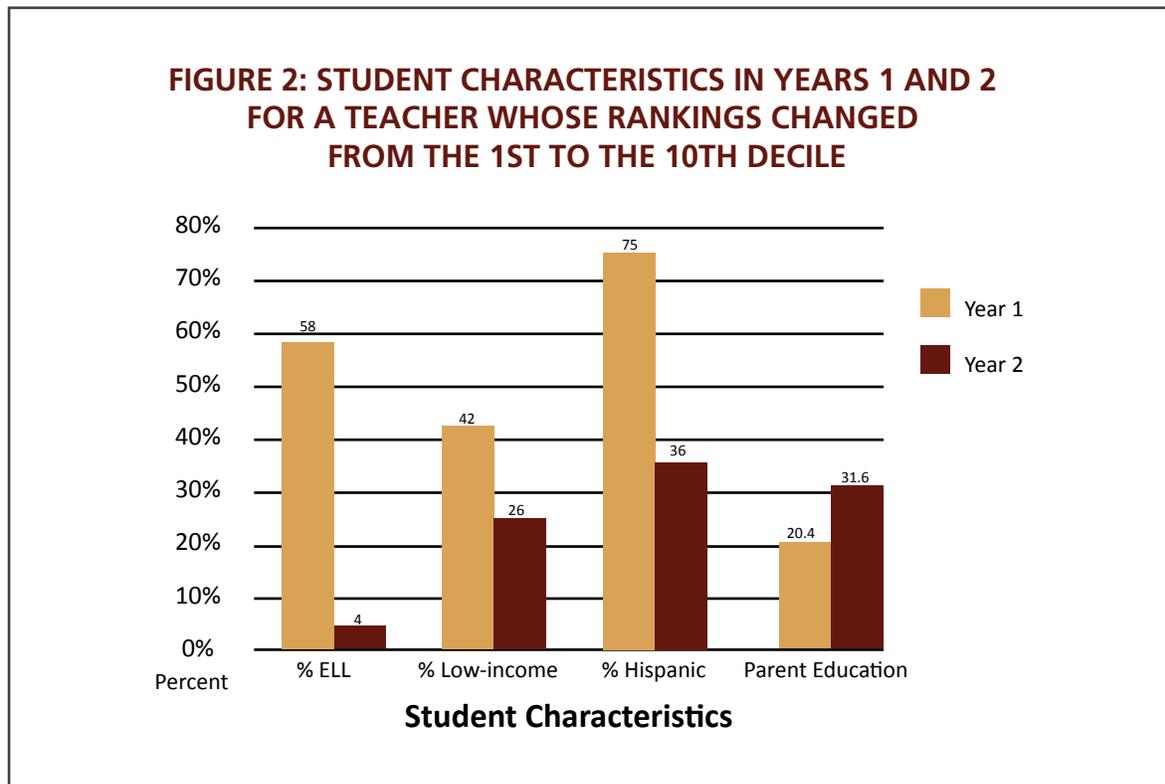
I went to a transition classroom, and now there's a red flag next to my name. I guess now I'm an ineffective teacher? I keep getting letters from the district, saying "You've been recognized as an outstanding teacher" ... this, this, and that. But now because I teach English-language learners who "transition in," my scores drop, and I get a flag next to my name.

I'm scared to teach in the fourth grade. I'm scared I might lose my job if I teach in an [ELL] transition grade level, because I'm scared my scores are going to drop, and I'm going to get fired because there's probably going to be no growth.

Similar problems occur for teachers working with large numbers of special education students and with teachers working in gifted and talented programs, where students have already topped out on the test.<sup>34</sup>

These and other concerns have caused leading research organizations to warn that value-added modeling (VAM) is not appropriate for making decisions about individual teachers. A major report by the RAND Corporation concluded:

The research base is currently insufficient to support the use of VAM for high-stakes decisions about individual teachers or schools.<sup>35</sup>



Similarly, Henry Braun stated in his review of research:

VAM results should not serve as the sole or principal basis for making consequential decisions about teachers. There are many pitfalls to making causal attributions of teacher effectiveness on the basis of the kinds of data available from typical school districts. We still lack sufficient understanding of how seriously the different technical problems threaten the validity of such interpretations.<sup>36</sup>

And the National Research Council's Board on Testing and Assessment concluded that:

VAM estimates of teacher effectiveness ... should not be used to make operational decisions because such estimates are far too unstable to be considered fair or reliable.<sup>37</sup>

Good systems must be designed so that teachers are not penalized for teaching the students who have the greatest educational needs. Rather, they should explicitly seek to provide incentives that recognize and reward teachers who work with challenging students.

Thus, to understand how teachers influence student learning, more information about teachers' practices and context are needed than value-added measures can provide. Student learning evidence needs to be multi-faceted and accompanied by an analysis of the teacher's students and teaching context. It must be integrated with evidence about teachers' practice, and its use should be focused on improving teaching.

**Using other evidence of student learning:** The shortcomings of value-added analysis do not mean that districts cannot recognize and reward teachers for producing strong student learning, or create incentives for them to help other teachers and serve the neediest students. It is possible to use other measures of student learning in evaluations of teaching, such as pre- and post-tests of learning conducted by districts or schools, or even learning evidence assembled by teachers themselves.

Such evidence can be drawn from classroom assessments and documentation, using instruments like the Developmental Reading Assessment, or pre- and post-test measures of student learning in specific courses or curriculum areas (developed by individual teachers, departments, school faculty, or district faculty or staff). Other possibilities include evidence of student accomplishments in relation to teaching activities, such as student science investigations, research papers, or art projects. Some districts use evidence from teachers' careful documentation of the learning of a set of diverse students over time, like that included in National Board Certification portfolios.

Analysis of standardized test results could be included, where appropriate, with attention to the relationship of the tests to the curriculum and their appropriateness for

the students being taught. Different assessments may be needed for different students—for example, English language proficiency tests for new English learners and IEP-related measures or alternative assessments for some special education students.

The evidence can be assembled by the teacher in a teaching portfolio, demonstrating and explaining the progress of students on a wide range of learning outcomes in ways that take students' starting points and characteristics into account. In some schools, teachers use their own fall and spring classroom assessments as a way of gauging student progress, including measures tailored to the learning goals of specific students (for example, special education students or English language learners). Some schools and districts have common assessments that are used in particular grades and courses. As part of a portfolio of evidence, such measures can document teaching effectiveness in achieving specific curriculum goals.

Measures of student learning in specific subject areas may include scored writing samples (including first to last drafts) or reading samples, mathematics assessments, assessments of science or history knowledge, or even musical performances. These measures typically provide better indications of student learning in a specific course or subject area because they are curriculum-specific. They are also more likely to capture the effects of a particular teacher's instruction and be available for most or all students. A teacher might even document the Westinghouse science competition awards she helped students win, or specific breakthroughs achieved by her students, with evidence of her role in supporting these accomplishments.

Some innovative teacher evaluation programs (in Rochester, New York and Denver, Colorado, for example, as well as selected districts in Arizona) have found valid ways to include various types of evidence of student learning. In Rochester's career ladder, for example, evidence of student learning, determined by the teacher, is assembled in the teacher's portfolio. In Denver's *Procomp* system,<sup>38</sup> teachers set two goals annually in collaboration with the principal, and document student progress toward these goals using district, school, or teacher-made assessments to show growth.

In award-winning Long Beach, California, a predominantly minority district widely recognized for its achievement gains, teachers are evaluated, through observations, on their performance in relation to the California Standards for the Teaching Profession. (See sidebar on San Mateo, California, page 17, for a list of these standards.) In addition, teachers and administrators together set goals for student progress and improvements in practice at the school level, as teams within departments or grade levels, and as individuals. Progress toward these goals is taken into account in both self-evaluations and supervisory evaluations. The evaluatee proposes how achievement of his or her objectives can be assessed, using evidence such as:

- Teacher observation and judgment
- Anecdotal and cumulative records

- Success and progress on a continuum of learning or a course of study
- Teacher, department, or school-made tests for pre-testing and post-testing
- Curriculum-related tests
- Use of audio-visual documentation if desired and available
- Student self-evaluations
- Evaluative discussion with students and parents
- Records of students' past learning performances
- Files of students' work collected to show growth
- Action research<sup>39</sup>

The district creates explicit and ongoing opportunities for schools, departments, and grade level teams to review student work and test score data of various kinds, to evaluate progress within and across classrooms, to discuss curriculum and teaching strategies, to problem solve around the needs of individuals and groups of students, and to plan for improvements.

Data analysis practices in Long Beach reflect what National Board Certified teacher Renee Moore suggests when she argues:

I would suggest that one important component of new student assessments is that the results be given not just to the individual teachers of those students, but that teachers be involved in the interpretation and discussion of test data together, in various configurations.... This type of data interpretation would be, in fact, a form of de facto peer evaluation.<sup>40</sup>

One outcome of these problem-solving processes in Long Beach is that the most expert teachers are encouraged to take on the highest-need students. Because gifted veterans can often move such students forward furthest, the students gain much more than they otherwise would. Meanwhile, other teachers who have easier classes can experience greater success, and everyone wins. Superintendent Chris Steinhauser believes that the use of state test scores in individual teacher evaluations would undermine the district's successful practices by penalizing teachers for taking on the toughest assignments and by undermining teacher collaboration. Instead, in Long Beach, the teacher evaluation system supports a culture of collective improvement.

Teachers' goal-setting to guide the collection of learning evidence plays a major role in Arizona's career ladder program, which has, for more than two decades, required the use of student learning data to ascertain teaching effectiveness in participating districts. (See sidebar, page 27.) In all of the systems created by local districts, evidence of student learning is combined with other evidence from standards-based teaching evaluations

## Use of Student Learning Evidence in Amphitheater, Arizona

More than 20 years ago, Arizona created incentives for districts to develop career ladders that base evaluation and advancement in part on evidence of student learning. One of these districts, Amphitheater, structures teachers' analysis of student learning evidence in several ways.

First, teachers set year-long goals for their students based on their initial achievement and state standards. Then, for one or more of these goals, they choose an assessment tool that they can use for a pre-assessment, mid-year assessment, and end-of-year assessment. In their evaluation, teachers provide copies of these assessments and a data sheet that lists outcomes for all of their students on each of these assessments. These data are the basis for further analysis. At the beginning, middle, and end of the year, teachers answer these kinds of questions:

1. Reflect on the student results from your assessment.
  - a. Identify and describe the areas of your students' strengths and weaknesses as they pertain to your goals (pre-assessment).
  - b. Identify why students regressed, stayed at the same level, and/or grew more than expected (mid- or post-assessment).
2. Looking at the results of students in three groups (high, medium, low), explain the instructional practices you will use to increase the achievement of each group.
3. How will students use assessment results to further their learning (e.g., student goal setting and reflection, etc.)?

Include three (3) student work samples evidencing the 21st Century Skills explained in the Student Achievement Plan. Submit samples of student work from high, medium, and low quality results with Student Work Sample Cover Sheets.

4. Describe specific actions taken for individual students who are not demonstrating growth, not meeting the growth targets or exceeding the identified growth targets.

***Interventions** are actions taken to individualize for students who are not making significant growth. **Extensions** are actions taken for students who have surpassed their growth targets before the end of the year.*

These data and reflections on student progress are part of the teacher's portfolio, along with a professional growth plan that outlines activities the teacher plans to pursue, their expected impact on student learning, and sources of evidence for evaluating that impact. In addition, teachers who want to advance on the career ladder must lead and participate in a collaborative action research group. The group explores a topic based on student needs. Teachers individually document the impact of instruction on student achievement.

For more information, see: [http://www.amphi.com/departments--programs/career-ladder/collaborative-action-research-\(car\)-2011-2012.aspx](http://www.amphi.com/departments--programs/career-ladder/collaborative-action-research-(car)-2011-2012.aspx)

conducted through classroom observation, and evidence of teachers' skills or practices as described earlier.

One study of the Arizona career-ladder programs found that, over time, participating teachers demonstrated an increased ability to create tools to assess student learning gains in their classrooms; to develop and evaluate pre- and post-tests; to define measurable outcomes in “hard-to-quantify” areas like art, music, and physical education; and to monitor student learning growth in relation to their action plans. They also showed a greater awareness of the importance of sound curriculum development; more alignment of curriculum with district objectives; and increased focus on higher quality content, skills, and instructional strategies.<sup>41</sup> Thus, the development and use of student learning evidence was associated with improvements in practice.

Some states are seeking to build similar systems. For example, in Massachusetts's new teacher evaluation system, collaboratively developed by the State Department of Education and the Massachusetts Teachers Association, along with other key experts and stakeholders, a multiple measures system incorporates all of these kinds of evidence. (See sidebar, page 29.)

#### **4. Create Structures to Support High-Quality, Fair, and Effective Evaluation**

One serious shortcoming of teacher evaluation reforms is that they have often focused on designing instruments for observing teachers, without developing the structural elements of a sound evaluation system. These elements should include, at minimum:

- trained, skilled evaluators
- supports for teachers needing assistance
- governance structures that enable sound personnel decisions
- resources to sustain and monitor system

Strong evaluation systems need principals and other evaluators with deep knowledge of teaching and learning, as well as an understanding of how to evaluate teaching, how to give useful feedback, and how to plan professional development that supports teacher learning. The lack of such knowledge and training has been a major problem for the validity, fairness, and utility of many teacher evaluation systems. The answers to these needs are severalfold:

- *Stronger principal preparation* coupled with the use of performance-based licensure for principals. Connecticut's approach is a good example: Principal preparation is focused on instructional leadership, teacher supervision, and professional development. Performance-based licens-

## Massachusetts' Multiple Measures System of Evaluation

Massachusetts has adopted a multiple measures system of evaluation for both teachers and administrators that considers practice, professional contributions, and student outcomes in an integrated process. The standards for teachers focus on:

1. **Curriculum, Planning, and Assessment:** Promotes the learning and growth of all students by providing high-quality and coherent instruction, designing and administering authentic and meaningful student assessments, analyzing student performance and growth data, using this data to improve instruction, providing students with constructive feedback on an ongoing basis, and continuously refining learning objectives.
2. **Teaching All Students:** Promotes the learning and growth of all students through instructional practices that establish high expectations, create a safe and effective classroom environment, and demonstrate cultural proficiency.
3. **Family and Community Engagement:** Promotes the learning and growth of all students through effective partnerships with families, caregivers, community members, and organizations.
4. **Professional Culture:** Promotes the learning and growth of all students through ethical, culturally proficient, skilled, and collaborative practice.

The collection of evidence used in the evaluation includes:

1. Multiple measures of student learning, including measures of student progress on classroom assessments aligned with the state standards; measures of student progress on learning goals set between the educator and evaluator for the school year; and district or state measures where appropriate and available
2. Judgments about practice based on observations and artifacts of teaching
3. Evidence of professional responsibilities and growth, such as self-assessments, peer collaboration, professional development linked to goals and or educator plans, and contributions to the school community and professional culture
4. Evidence of active outreach to and ongoing engagement with families
5. Student feedback with respect to teachers, and staff feedback with respect to administrators

There are no fixed percentages for any single source of data. Data are combined and ratings from unsatisfactory to exemplary are determined using multiple categories of evidence using either the state-developed rubric or one approved by the state.

<http://www.doe.mass.edu/lawsregs/603cmr35.html?section=07>.

## Criteria for Using Measures of Student Learning in Teacher Evaluations

**1. Assessment of teachers' contributions to student learning should rely on multiple measures of student learning, not a single value-added measure (VAM).** Researchers have found that teaching effectiveness ratings are affected by which measures of learning are used. VAM estimates using different tests thus produce different results. In addition, no single test measures all the important aspects of teaching and learning, and most state tests that are spring-to-spring measures add summer learning loss differentials to the other sources of error in looking at potential teacher effects. As a consequence, the system of evaluation should include multiple measures of student learning, which can be collected at the classroom, school, or district levels, and which evaluate a broad range of desired outcomes in different ways. Some of these might be tests. Others should be papers, projects, exhibitions, or portfolios of student work that demonstrate important skills requiring planning, research, revision, and demonstration of applied understanding. Some might be fall-to-spring measures (e.g. scored essays at the beginning and end of the year); others might show students' revisions of their products illustrating growth during the teaching, feedback, and revision process); still others might be periodic progress indicators using tools like the Developmental Reading Assessment. Equally important, teachers should be trained to collect, examine, interpret, and use evidence about student learning to reflect on and plan instruction, and to inform improvements.

**2. Measures of learning should reflect the curriculum a teacher is expected to teach and the full domain of skills and competencies students are expected to develop.** Thus, the collection of assessments should reflect the full content of the curriculum, including higher order thinking and performance skills, such as writing, investigation, research, and problem-solving. Measures should be continuously evaluated to determine the extent to which they address the range of knowledge and skills sought and their ability to capture student learning authentically.

**3. Valid measures must be used for all students, including those with special needs or limited English proficiency, as well as particularly high-achieving or low-achieving students.** This means that assessments should measure the full continuum of achievement where possible, and may need to be differentiated to be appropriate for some students. In some cases, certain students may need to be assessed separately using different measures.

**4. Test measures intended to indicate growth must capture learning and growth validly and reliably at the student's actual achievement level.** To evaluate how much a student has learned over a period of time, the assessment measures must reflect student performance at the level where a student actually achieves and measure changes accurately. Tests that measure only grade-level standards do not measure learning gains for students who are achieving below or above grade level. A test that has a low ceiling will not reveal gains made by students near the top of the distribution, and a test that has a high floor will not measure what might be substantial gains made by students who are achieving below that level. Thus, the teachers who teach these students will be disadvantaged in value-added comparisons even if statistical adjustments are made. Use of various measures should recognize the information offered by and limitations of each.

**5. The use of any value-added measure should take into account characteristics of the students and the context that affect student achievement gains.** Such factors include parent education, special needs of students (English learner and special education status, poverty, homelessness), student attendance, and classroom composition, in addition to the individual student's prior achievement. In particular, studies show that classroom composition greatly affects teachers' value-added scores. This information should be taken into account both in the models and in the overall analysis of information for the ultimate evaluation judgment. Other factors that may make a significant difference include class size, the quality and availability of curriculum materials, whether students also receive tutoring or related instruction from another teacher, etc. If these factors are not accounted for in the value-added model, they should be accounted for in the overall evaluation of a teacher.

*continued on next page*

**6. Value-added measures should be used only when there is a sufficient sample size and multiple years of data.** Studies find that many teachers have few students linked to them for whom data is available for both prior-year and current-year achievement. Other students who are mobile may have spent only a short time in a given teacher's classroom. Both of these are sources of considerable error. Year-to-year instability in teacher rankings is also very high. Many experts suggest that there should be at least 50 students (who have been with the teacher for a large majority of the year in each case) and at least 3 years of data to use in estimating a value-added score. Even with these considerations, it is important to recognize that multiple years of data may mask the year-to-year instability of scores, but do not eliminate the causes of such instability, which may often include the composition of classes that teachers teach.

**7. The evaluation system should consider evidence about student performance and teacher practice in an integrated fashion.** In order to identify how a teacher's practices are connected to and influence student learning, the system must look at both together, along with information about the students who are being served. Teacher evaluation must triangulate information about students, teaching, and outcomes, just as the evaluation of doctors considers patient outcomes in light of patient characteristics and the doctor's expertise in enacting professional standards of practice. An integrated evaluation of practice and outcomes is also necessary to reduce ambiguity in attributing gains in student learning to a particular teacher's contribution. Value-added systems face considerable challenges in distinguishing between the instruction a classroom teacher provides and the instruction provided by prior teachers, other concurrent teachers, resource specialists, or tutors. In evaluating teacher performance, linking what students have learned to what a teacher actually did in the classroom is critically important.

**8. Methods to verify the accuracy of the data should be used routinely.** Researchers have found many sources of error in VAM studies, including faulty student data and inaccurate links between teacher and student data. The roster and characteristics of students who are linked to teachers in a dataset must be verified before estimates are used in an evaluation. The most direct way to do this might be to provide all teachers with lists of the students to whom teachers are linked, including their demographic characteristics and course-taking records, so that teachers can check the lists for errors. Evaluating student records before analyses are done would also allow students to be flagged who have not been in the classroom of that teacher for a sufficient period of time, or who cannot be validly assessed by the specific tests being used.

**9. Value-added estimates should be weighted in the evaluation process commensurate with their limitations.** Given the large error range and instability associated with value-added estimates, and the fact that they represent many factors other than the individual teacher, it is clear that such estimates should be treated as error-prone, and should always be considered in conjunction with other evidence of student learning and teacher practice.

**10. The use of student learning evidence should be a source of continual study for educators, researchers, and systems.** Independent researchers should continually examine teacher evaluation systems and results to identify questions and problems and to suggest improvements. There should be a regular, thorough analysis of all evaluation data, including VAM estimates if they are used, to examine whether different measures (based on different tests or other learning evidence, different measures of teacher performance, or different models for analyzing and combining data) provide different estimates. Given all the questions about what VAM estimates measure, studies should be done annually to examine the congruency between VAM measures and other measures of student learning and teacher performance, without a presumption that VAM estimates are more accurate than other, more stable measures.

ing assessments require principals to demonstrate that they can evaluate videotapes of teaching accurately and then plan appropriate feedback and professional supports.

- *Specific, intensive training in evaluation and supervision.* Training for evaluators is conducted in states like Connecticut and in programs like the Teacher Advancement Program, which offers several days of training for principals in the use of the standards-based evaluation tools and strategies for providing useful feedback and follow-up.
- *Involvement of lead teachers with content expertise* in some aspects of supervision and evaluation. For example, some districts involve department chairpersons or other lead teachers matched to teachers' content areas.

Also important are the provision of supports for teachers who need additional assistance—something few principals have time to provide—and the creation of decision-making procedures and structures that allow personnel decisions to be made fairly and efficiently.

Rochester, New York has undertaken innovative approaches to provide greater content expertise in the evaluation process and to encourage teachers to work as - and be evaluated as - teams. There, teachers have the option to be observed and evaluated by a trained Lead Teacher/Peer Evaluator, if they so request, in addition to their administrative supervisor. Lead Teachers are selected by the Rochester Career In Teaching (CIT) Governing Panel, made up of six administrators appointed by the district and six teachers appointed by the teachers' union.

Teachers can also choose the Performance Appraisal Review for Teachers (PART). PART, negotiated in Rochester in 1987, represents a collegial approach to evaluation of teachers. Self-selected groups of teachers, in the same school or across schools, design an annual or multi-year project related to improved teaching and learning, set goals for themselves and are then evaluated, not just as individual practitioners but also as members of a teaching team. This promotes teacher collaboration and contextualizes teacher assessment within the reality of teachers' day-to-day actual work. PART has also promoted innovation, inter-disciplinary approaches, project-based learning and performance-based assessments, moving the collective practice of teaching forward.

Of course, an evaluation system based on standards of professional practice must also be able to remove individuals from the profession when they do not, after receiving assistance, meet professional standards. The most long-standing evaluation systems that have successfully supported evaluation and personnel actions for both beginning and veteran teachers are those that have used Peer Assistance and Review programs that rely on highly expert mentor teachers to conduct some aspects of the evaluation and

provide assistance to teachers who need it. The systems in Cincinnati, Columbus, and Toledo, Ohio; Rochester, New York; Poway and San Juan, California; Seattle, Washington have all been studied and found successful in identifying teachers for continuation and tenure as well as those needing intensive assistance and personnel action. These systems—collaborations between unions and school boards, which build in due process and assistance for teachers placed in intervention—have proven more effective than traditional evaluation systems at both improving and efficiently dismissing teachers while avoiding union grievances.

Key features of these systems thus include not only the instruments used for evaluation, but also the expertise of the consulting teachers or mentors—skilled teachers in the same subject areas and school levels, who have released time to serve as mentors to support their fellow teachers—and a system of due process and review that involves a panel of both teachers and administrators who make recommendations about personnel decisions based on evidence from the evaluations. This joint committee oversees the work of the mentor teachers who support both beginning teachers prior to tenure and veteran teachers who are struggling. Based on reports from both the mentor teachers and the principals, the committee decides which teachers will receive tenure, which will have another year to improve, and which will be dismissed. Similarly, the committee decides, before the school year is over, which teachers in the intervention program have improved sufficiently to be continued in the district.

In such systems, beginning teachers have been found to stay in teaching at higher rates because of the mentoring they receive, and those who leave are usually those the district has chosen not to continue rather than those who have quit. Among veteran teachers identified for assistance and review, many improve sufficiently to be removed from intervention status, and one third to one half leave by choice or by district request. Because teacher associations collaborate in creating and administering the programs, which are designed to ensure due process, there are no procedural grievances when a teacher is dismissed.

## **5. Create Aligned Professional Learning Opportunities**

Finally, it is important to link both formal professional development and job-embedded learning opportunities to the evaluation system. Evaluations should trigger continuous goal-setting for areas teachers want to work on, specific professional development supports and coaching, and opportunities to share expertise, as part of recognizing teachers' strengths and needs.

Furthermore, professional learning opportunities must be of high quality. Rather than the drive-by, spray-and-pray, flavor-of-the-month afterschool workshops that predominate in the United States, teachers should have access to the kind of sustained, focused learning that has been shown to improve practice. For example, a review of experimental studies found that, while professional development offerings of less than 14

## Peer Assistance and Review

Peer Assistance and Review (PAR) programs have demonstrated over 30 years that it is possible to evaluate teachers rigorously, support them intensely, and make personnel decisions effectively. The first Peer Assistance Review program began in Toledo, Ohio as a partnership between the school board and the teachers' union in the early 1980s. Union leader Dal Lawrence was convinced that "teaching would become a profession only when teachers, themselves, set standards for their work and decided who met those standards and deserved to teach."<sup>43</sup> To address these concerns, Lawrence proposed an intern program to the Toledo Public School District to better mentor and induct new teachers into the profession. The program also provides intensive support to veteran teachers who are struggling, and enables a timely and well-grounded decision to be made about tenure and continuation in the district.

Thirty years later, the Toledo PAR Program has deepened and become the blueprint for other PAR programs across the country. Such programs now exist in at least 41 districts in 13 states (California, Colorado, Florida, Illinois, Indiana, Maryland, Michigan, Minnesota, Missouri, New Mexico, New York, Ohio and Washington). A just-published study of PAR programs in California, where a state law encourages district adoption, documented the successes of the model in Poway and San Juan, demonstrating the capacity of the program to be successfully adapted to new contexts.<sup>44</sup>

### How PAR Operates

In Toledo, the PAR governing body consists of an Internal Review Board made up of nine members (five teachers and four administrators) who oversee the program. Although the number of board members differs in other districts, panels generally include nearly equal numbers of teachers and administrators, with a slight edge to teacher members. This governing body is responsible for overseeing the work of the mentor teachers, as well as evaluating accumulated evidence on a participating teacher and making final tenure and employment recommendations to the superintendent of schools.

**Selection of Consulting Teachers.** Consulting Teachers, who provide the support and evaluation, have at least five years of teaching experience, and undergo an intensive selection process that includes classroom observations, interviews, a review of teaching evaluations, and recommendations from peers and administrators. In the Toledo model, these mentors are employed full-time to support and evaluate approximately ten novice and/or struggling teachers over the course of an individualized intervention or mentorship period. They serve no more than three years before returning to the classroom, and are paid an annual stipend of \$5,000-\$7,000.

**Support and Evaluation.** Consulting Teachers design a support, intervention, and improvement plan based on the needs of each teacher. The support spans the five domains that comprise the evaluation criteria: planning and designing instruction, instruction, classroom management, assessment, and professional development. The specific tasks range from assistance with lesson planning and sharing of resources to observing classrooms and providing feedback on classroom management and instructional practices.

When working with experienced teachers, Consulting Teachers provide intensive guidance and direction, including helping them design and implement individualized improvement plans. Consulting Teachers make periodic reports to the Governance Board, marshaling accumulated

*continued on next page*

evidence about the teacher's professional practice. At the conclusion of the teacher's time in PAR (usually a year), the Governance Board examines the Consulting Teacher's reports, the administrator's evaluations, and other evidence and makes a recommendation about whether or not the teacher should be retained in the district.

**Program Effects.** Studies of PAR outcomes have found that beginning teacher retention rates have increased significantly and that those who leave are now primarily those the district does not renew, rather than candidates who become disenchanted with teaching. On the national level, the National Center for Educational Statistics reports that the new-teacher attrition rate for those who participate in PAR programs is 15%, compared to 26% for those who do not.<sup>45</sup> Among intervention candidates, many improve; those who do not improve leave teaching without extended legal proceedings, because due process is built into the design of the model. A Harvard Graduate School of Education study of Toledo reports that nearly 8% of program participants from 1981 through 2008 were either dismissed or resigned after the first year of participation in PAR.<sup>46</sup>

### PAR California-Style

In the early 1990s, teacher leaders from Poway studied Toledo's program and brought its key features to their district. As Poway has a long tradition of quality practice and careful teacher selection, the primary goal was not to get rid of bad teachers, but to further develop good ones. As Don Raczka, the retired president of the Poway Federation of Teachers, noted in 2000, "Today we have better trained teachers who are used to reflecting on their practice and talking about pedagogy. We want evaluation to make sense for them."<sup>47</sup> The program was also designed to work with novices and struggling teachers.

In 1999, the California State Legislature authorized a statewide grant program for PAR programs, as part of a hybrid model designed to work in tandem with an existing induction program called Beginning Teacher Support Assessment (BTSA). Leaders in San Juan created their model with these funds after studying the work in Poway and other successful districts.

As in Toledo (and other PAR districts), San Juan and Poway use the program to sort out ineffective teachers who do not improve. Historically, about two thirds of veterans identified for intervention improved substantially and successfully completed the program; about one third in each case resigned or were dismissed. Among beginning teachers, about 20% are not renewed as a result of the program. As SRI researchers noted, though, the effects of the program are much broader. For example, Poway's program for beginning teachers has served 1,875 individuals, more than 60% of the teachers currently in the district. Poway officials note that the success of the program in both building competence and weeding out poor performers at the beginning of their careers has helped to raise the overall quality of practice in the district. Both consulting teachers and those being mentored report that they became better teachers as a result of their careful analysis of and work on practice. Thus, there are relatively few teachers identified as struggling later in their careers.

The PAR Governing Boards make other contributions to teaching quality. The SRI study noted that the boards not only ensure that solid evidence is brought to bear on personnel decisions but also "turned out to be problem solving forums where district officials and union leaders collaboratively address routine operational and policy problems."<sup>48</sup> The decision-making processes help build a shared sense of responsibility and collaboration that extends into other areas of work, and helps union officials and district administrators maintain a focus on teaching and learning.

hours per year on a given topic had no effect on student learning, high-quality professional development programs of about 50 hours over a 6-to-12 month period increased student achievement by 21 percentile points on average.<sup>49</sup>

These high-quality opportunities are typically

- focused on the learning and teaching of specific curriculum content,
- organized around real problems of practice,
- connected to teachers' work with children,
- linked to analysis of teaching and student learning,
- intensive, sustained, and continuous over time,
- supported by coaching, modeling, observation, and feedback,
- connected to teachers' collaborative work in professional learning communities,
- integrated into school and classroom planning around curriculum, instruction, and assessment.

Such opportunities may include intensive institutes focused on particular strategies or on the teaching of specific curriculum, interspersed with opportunities for teachers to try things in the classroom, receive coaching, reflect together on their experiences, revise and revamp their approaches, and develop increasingly polished skills in an iterative cycle of practice, reflection, and fine-tuning. They may include collective opportunities to analyze observations or videos of teaching and/or samples of student work, study groups, action research projects, peer observations, and collaborative planning and evaluation in grade-level or departmental teams.<sup>50</sup>

Opportunities to study and learn specific strategies that are central to individual practice are important, and may be pursued through specific coursework or participation in subject-matter networks or conferences. It is also important to organize professional learning that creates collective capacity and curriculum coherence among the staff members in a school. The best systems create time for teachers to work and learn together during the school day, as is common in high-achieving nations in Europe and Asia, where teachers typically have 15-25 hours a week to plan and work together.<sup>51</sup> Over time, this collective investment reaps greater gains for student learning than the efforts of any one teacher alone.

Professional learning opportunities should also be conceptualized as part of a career continuum that encourages teachers to gain and share expertise and creates avenues for such sharing to occur, as teachers take on roles as mentor and master teachers, as curriculum and assessment specialists, and as leaders of school-improvement activities.

In addition, schools and districts must develop conditions that provide teachers and principals with sufficient organizational and instructional support to carry out a system of teacher evaluation that enables continuous learning. For example, teachers

and principals need the time and guidance to develop a shared understanding of effective teaching, to examine artifacts of practice for evidence of learning, to explore one another's assumptions about how learning occurs and what counts as evidence of learning, to promote reflection, and to learn how to provide effective feedback. Without these sorts of school-based conditions, the ability for a teacher-evaluation program to stimulate ongoing professional learning is severely handicapped.

The challenge of getting to scale with good educational practice<sup>52</sup> is in developing widespread educational leadership and expert teaching on the one hand, and encouraging the design of effective organizations on a system-wide scale on the other. The work of improving practice must therefore be conceptualized as collective rather than individual.

To transform systems, incentives must be structured to promote collaboration and knowledge sharing, rather than competition, across organizations. Knowledge-sharing is needed to develop not only learning organizations but a learning oriented system of education. This has been the primary strategy for improvement in Finland, for example, where ongoing evaluation and inquiry into practice are stimulated within and across classrooms, across schools partnered within regions, and within the system as a whole.<sup>53</sup>

Also key to developing such a system is the creation of networks that allow teachers, leaders, schools, and districts to learn from one another. Andy Hargreaves describes an initiative in England in which 300 schools that were declining in performance were networked with one another, provided with technical assistance and support from mentor schools, and given a small discretionary budget to support their efforts. Schools were also given a practitioner-generated list of strategies that had produced improvements in other schools. More than two-thirds of these “exceptionally energized” schools experienced gains over the next two years at rates double the national average, “without,” the researchers noted, “the characteristic mandates and prescriptions that had characterized English reforms before this point.”<sup>54</sup> An initiative in Ontario, Canada used similar school-to-school networking strategies and leveraged them further by identifying positive exemplars that schools could visit to see successful reforms in action. As these examples indicate, when schools learn to create better conditions for teaching and learning, individual and collective teaching practice can improve.

## Criteria for an Effective Teacher Evaluation System

- 1. Teacher evaluation should be based on professional teaching standards** and should be sophisticated enough to assess teaching quality across the continuum of development from novice to expert teacher.
- 2. Evaluations should include multi-faceted evidence of teacher practice, student learning, and professional contributions** that are considered in an integrated fashion, in relation to one another and to the teaching context. Any assessments used to make judgments about students' progress should be appropriate for the specific curriculum and students the teacher teaches.
- 3. Evaluators should be knowledgeable about instruction and well trained in the evaluation system**, including the process of how to give productive feedback and how to support ongoing learning for teachers. As often as possible, and always at critical decision-making junctures (e.g., tenure or renewal), the evaluation team should include experts in the specific teaching field.
- 4. Evaluation should be accompanied by useful feedback, and connected to professional development opportunities** that are relevant to teachers' goals and needs, including both formal learning opportunities and peer collaboration, observation, and coaching.
- 5. The evaluation system should value and encourage teacher collaboration**, both in the standards and criteria that are used to assess teachers' work, and in the way results are used to shape professional learning opportunities.
- 6. Expert teachers should be part of the assistance and review process** for new teachers and for teachers needing extra assistance. They can provide the additional subject-specific expertise and person-power needed to ensure that intensive and effective assistance is offered and that decisions about tenure and continuation are well grounded.
- 7. Panels of teachers and administrators should oversee the evaluation process** to ensure that it is thorough and of high quality, as well as fair and reliable. Such panels have been shown to facilitate more timely and well-grounded personnel decisions that avoid grievances and litigation. Teachers and school leaders should be involved in developing, implementing, and monitoring the system to ensure that it reflects good teaching well, that it operates effectively, that it is tied to useful learning opportunities for teachers, and that it produces valid results.

## Conclusion

In summary, teacher evaluation systems need to be considered not only in terms of evaluation instruments or procedures, but also in terms of the policy systems in which they operate and the school-based conditions that are needed to stimulate continuous learning and improvement. These conditions include:

- teacher participation in developing the system and supporting the ongoing decision-making processes;
- recognition and encouragement of collegial contributions to overall school success and clear criteria for accomplishment that all eligible teachers can achieve, rather than a quota system that pits teachers against each other;
- valid evidence of teacher effectiveness based on multiple measures, including
  - evaluation of practice based on multiple classroom observations and examination of other classroom evidence (e.g., lesson plans, student assignments and work samples) using a standards-based evaluation instrument that incorporates evidence of planning, instruction, the learning environment, and student assessment;
  - evidence of learning by the teacher's students on a range of valid assessments that appropriately evaluate the curriculum the teacher teaches;
  - evaluation of teachers' contributions to colleagues and to the school as a whole;
- consideration of the needs of the students the teacher serves and valid and appropriate assessment of teaching and learning for all students, including students with special learning needs and new English language learners;
- ongoing, high-quality professional learning opportunities that will enable teachers to meet the standards.

Many initiatives to measure and improve teaching effectiveness have emerged as pressures for improved student achievement have intensified. Such initiatives will have the greatest payoff if they stimulate practices known to support student learning and are embedded in systems that also *develop* greater teaching competence. Such systems will be based on professional teaching standards and the teaching of well-defined curriculum content. They will make intense use of coaching and offer extensive opportunities for teachers to help their colleagues and their schools improve. Policies that create increasingly valid measures of teaching effectiveness—and that create innovative systems for recognizing, developing and utilizing expert teachers—can ultimately help to create a more effective teaching profession.

## Endnotes

- <sup>1</sup> Accomplished California Teachers. (2010). *A quality teacher in every classroom: Creating a teacher evaluation system that works for California*. Stanford, CA: National Board Resource Center, Stanford University.
- <sup>2</sup> For a summary of studies, see Darling-Hammond, L., & Bransford, J. (2005). *Preparing teachers for a changing world: What teachers should learn and be able to do*. San Francisco: Jossey-Bass; Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Educational Policy Analysis Archives*, 8(1). Retrieved from <http://epaa.asu.edu/epaa/v8n1>; Wilson, S. M., Floden, R., & Ferrini-Mundy, J. (2001). *Teacher preparation research: Current knowledge, gaps, and recommendations*. Seattle: Center for the Study of Teaching and Policy, University of Washington.
- <sup>3</sup> Clotfelter, C., Ladd, H., and Vigdor, J. (2007). *How and why do teacher credentials matter for student achievement?* (NBER Working Paper 12828). Cambridge, MA: National Bureau of Economic Research; Goldhaber, D., & Brewer, D. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational evaluation and policy analysis*, 22(2), 129-145.
- <sup>4</sup> National Association of State Boards of Education (NASBE) Study Group on Teacher Preparation, Retention, Evaluation, and Compensation (2011). *Gearing Up: Creating a Systemic Approach to Teacher Effectiveness*. Arlington, VA; Author.
- <sup>5</sup> Newton, S. P. (2010). Predictive validity of the performance assessment for California teachers. Stanford, CA: Stanford Center for Opportunity Policy in Education. Retrieved from <http://scale.stanford.edu/>; Wilson, M., & Hallum, P. J. (2006). Using student achievement test scores as evidence of external validity for indicators of teacher quality: Connecticut's *Beginning Educator Support and Training* program. Berkeley, CA: University of California at Berkeley.
- <sup>6</sup> For more information, see <http://tpafieldtest.nesinc.com/>
- <sup>7</sup> See, for example, Bond, L., Smith, T., Baker, W., & Hattie, J. (2000). *The certification system of the National Board for Professional Teaching Standards: A construct and consequential validity study*. Greensboro, NC: Center for Educational Research and Evaluation; Cavaluzzo, L. (2004). *Is National Board Certification an effective signal of teacher quality?* (National Science Foundation No. REC-0107014). Alexandria, VA: The CNA Corporation; Goldhaber, D., & Anthony, E. (2005). *Can teacher quality be effectively assessed?* Seattle, WA: University of Washington and the Urban Institute; Smith, T., Gordon, B., Colby, S., & Wang, J. (2005). *An examination of the relationship of the depth of student learning and National Board certification status*. Boone, NC: Office for Research on Teaching, Appalachian State University; Vandevort, L. G., Amrein-Beardsley, A., & Berliner, D. C. (2004). National Board certified teachers and their students' achievement. *Education Policy Analysis Archives*, 12(46), 117.
- <sup>8</sup> Athanases, S. (1994). Teachers' reports of the effects of preparing portfolios of literacy instruction. *Elementary School Journal*, 94(4), 421-439.
- <sup>9</sup> Sato, M., Wei, R. C., & Darling-Hammond, L. (2008). Improving teachers' assessment practices through professional development: The case of National Board Certification. *American Educational Research Journal*, 45, 669-700; Tracz, S.M., Sienty, S. & Mata, S. (1994, February). *The self-reflection of teachers compiling portfolios for National Certification: Work in progress*. Paper presented at the Annual Meeting of the American Association of Colleges for Teacher Education, Chicago, IL; Tracz, S. M., Sienty, S., Todorov, K., Snyder, J., Takashima, B., Pensabene, R., Olsen, B., Pauls, L.,

& Sork, J. (1995, April). *Improvement in teaching skills: Perspectives from National Board for Professional Teaching Standards field test network candidates*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

- <sup>10</sup> This process is documented in the film, *The Mitchell 20*, Randy Murray (Executive Producer) and Andrew James Benson (Producer). United States: Randy Murray Productions. <http://www.mitchell20.com/>
- <sup>11</sup> Berry, B. (2009). *Keeping the promise: Recruiting, retaining, and growing effective teachers for high-needs schools*. Raleigh, NC: Center for Teaching Quality.
- <sup>12</sup> Dean, S. & Darling-Hammond, L. (forthcoming). *Using National Board Certification to Improve Schools*. Stanford, CA: National Board Resource Center and the Stanford Center for Opportunity Policy in Education.
- <sup>13</sup> Skinner, K. J. (2010). *Reinventing evaluation: Connecting professional practice with student learning*. Boston, MA: Massachusetts Teachers Association.
- <sup>14</sup> Milanowski, A., Kimball, S. M., & White, B. (2004). *The relationship between standards-based teacher evaluation scores and student achievement*. University of Wisconsin-Madison, Consortium for Policy Research in Education.
- <sup>15</sup> Milanowski, Kimball, & White, 2004.
- <sup>16</sup> Milanowski, Kimball, & White, 2004.
- <sup>17</sup> Hassell, B. C. (2002). *Better pay for better teaching: Making teacher compensation pay off in the age of accountability*. Progressive Policy Institute 21st Century Schools Project. Retrieved November 18, 2004, from <http://www.broadfoundation.org/investments/education-net.shtml>
- <sup>18</sup> Fryer, R. G. (2011). *Teacher incentives and student achievement: Evidence from New York City Public Schools* (NBER Working Paper no. 16850). Cambridge, MA: National Bureau of Economic Research; Martins, P. (2009). *Individual teacher incentives, student achievement and grade inflation* (Discussion Paper No. 4051). London, UK: Queen Mary, University of London, CEG-IST and IZA; Springer, M. G., Ballou, D., Hamilton, L., Le, V., Lockwood, J. R., McCaffrey, D. F., Pepper, M., and Stecher, B. M. (2010). *Teacher pay for performance: Experimental evidence from the Project on Incentives in Teaching*. Nashville, TN: National Center on Performance Incentives, Vanderbilt University.
- <sup>19</sup> MetLife Foundation. (2009). *The MetLife survey of the American teacher: Collaborating for student success*. New York: Author.
- <sup>20</sup> Jackson C. K., & Bruegmann, E. (2009, August). *Teaching students and teaching each other: The importance of peer learning for teachers*. Washington, DC: National Bureau of Economic Research.
- <sup>21</sup> Goddard, Y., & Goddard, R. D. (2007). A theoretical and empirical investigation of teacher collaboration for school improvement and student achievement in public elementary schools. *Teachers College Record*, 109(4), 877–896.
- <sup>22</sup> Odden, A., Kelley, C., Heneman, H., and Milanowski, A. (2001, November). *Enhancing teacher quality through knowledge- and skills-based pay* (CPRE Policy Briefs, R-34). Philadelphia: Consortium for Policy Research in Education, University of Pennsylvania.

- <sup>23</sup> Odden et al., 2001.
- <sup>24</sup> Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment and education: Principles, policy and practice*, 5(1), 7-75.
- <sup>25</sup> See Darling-Hammond & Bransford, 2005, for example.
- <sup>26</sup> Darling-Hammond, L., & Rothman, R. (2011). *Teacher and leader effectiveness in high-performing education systems*. Washington, DC: Alliance for Excellent Education and Stanford, CA: Stanford Center for Opportunity Policy in Education.
- <sup>27</sup> Lockwood, J. R., McCaffrey, D. F., Hamilton, L.S., Stetcher, B., Le, V. N., & Martinez, J. F. (2007). The sensitivity of value-added teacher effect estimates to different mathematics achievement measures. *Journal of Educational Measurement*, 44 (1), 47 – 67; Bill & Melinda Gates Foundation (2010). Learning about teaching: Initial findings from the Measures of Effective Teaching Project. Seattle: Author; Rothstein, J. (2011). Review of “Learning About Teaching: Initial Findings from the Measures of Effective Teaching Project.” Boulder, CO: National Education Policy Center; Corcoran, S. P., Jennings, J. L., and Beveridge, A. A. (2011). *Teacher effectiveness on high- and low-stakes tests*. Working paper, New York University, New York, NY.
- <sup>28</sup> For reviews, see Braun, H. (2005). *Using student progress to evaluate teachers: A primer on value-added models*. Princeton, NJ: ETS Policy Information Center; McCaffrey, D., Lockwood, J., Koretz, D., & Hamilton, L. (2005). *Evaluating value-added models for teacher accountability*. Santa Monica, CA: RAND Corporation.
- <sup>29</sup> Braun, 2005, p. 10.
- <sup>30</sup> Rivkin, S. G., Hanushek, E. A. & Kain, J. F. (2000). *Teachers, schools, and academic achievement* [revised] (Working Paper No. 6691). Cambridge, MA: National Bureau of Economic Research.
- <sup>31</sup> Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2007). Lasting consequences of the summer learning gap. *American Sociological Review*, 72, 167-180.
- <sup>32</sup> Sass, T. (2008). *The stability of value-added measures of teacher quality and implications for teacher compensation policy*. Washington DC: CALDER. See also Newton et al. for similar findings.
- <sup>33</sup> Newton, X., Darling-Hammond, L., Haertel, E., & Thomas, E. (2010). Value-added modeling of teacher effectiveness: An exploration of stability across models and contexts. *Educational Policy Analysis Archives*, 18 (23). Retrieved from <http://epaa.asu.edu/ojs/article/view/810>.
- <sup>34</sup> Amrein-Beardsley, A., & Collins, C. (in press). *The SAS® Education Value-Added Assessment System (EVAAS®): Its intended and unintended effects in a major urban school system*. Tempe, AZ: Arizona State University.
- <sup>35</sup> McCaffrey, D. F., Koretz, D., Lockwood, J. R., & Hamilton, L. S. (2005). *Evaluating value-added models for teacher accountability*. Santa Monica, CA: RAND Corporation.
- <sup>36</sup> Braun, 2005, p. 17.
- <sup>37</sup> National Research Council, Board on Testing and Assessment. (2009). Letter report to the U.S. Department of Education. Washington, D. C.: Author.

- <sup>38</sup> For more detail about the Denver Procomp system, see <http://denverprocomp.org>.
- <sup>39</sup> Long Beach Unified School District (2003). *Teacher evaluation handbook*. Long Beach, CA: Author.
- <sup>40</sup> Berry, B. & Daughtrey A., with Moore, R., Orphal, D., & Ratzel, M. (2012). *New Student Assessments and Advancing Teaching as a Results-Oriented Profession*. Washington, DC: Arabella Advisers.
- <sup>41</sup> Packard, R., & Dereshiwsky, M. (1991). *Final quantitative assessment of the Arizona career ladder pilot-test project*. Flagstaff, AZ: Northern Arizona University.
- <sup>42</sup> California State University Institute for Education Reform. (2000, March). *Peer assistance and review: Working models across the country*. Sacramento, CA: Author; Humphrey, D., Koppich, J., Bland, A., & Bosetti, K. R. (2011). *Peer review: Getting serious about teacher evaluation*. Menlo Park, CA: SRI International; National Commission on Teaching and America's Future (1996). *What matters most: Teaching for America's Future*. New York: Author.
- <sup>43</sup> Harvard Graduate School of Education. (n. d.). A user's guide to peer assistance and review. *The Project on the Next Generation of Teachers*. Retrieved October 1, 2011, from <http://www.gse.harvard.edu/~ngt/par/>.
- <sup>44</sup> Humphrey, Koppich, Bland, & Bosetti, 2011.
- <sup>45</sup> Marshall, R. (2008). *The case for collaborative school reform: The Toledo experience*. Washington, DC: Economic Policy Institute, p. 27.
- <sup>46</sup> Harvard Graduate School of Education. (n. d.). A user's guide to peer assistance and review.
- <sup>47</sup> California State University Institute for Education Reform, 2000.
- <sup>48</sup> Humphrey et al., 2011.
- <sup>49</sup> Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007–No. 033). Retrieved October 15, 2008, from [http://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/REL\\_2007033.pdf](http://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/REL_2007033.pdf)
- <sup>50</sup> Darling-Hammond, L., & Richardson, N. (2009, February). Teacher learning: What matters? *Educational Leadership*, 5(66), 46-53.
- <sup>51</sup> Darling-Hammond, L. (2010). *The flat world and education: How America's commitment to equity will determine our future*. New York, NY: Teachers College Press.
- <sup>52</sup> Elmore, R. (1996). Getting to scale with good educational practice. *Harvard Educational Review*, 66(1), 1-26.
- <sup>53</sup> Darling-Hammond, 2010.
- <sup>54</sup> Hargreaves, A. (2008). The coming of post-standardization: Three weddings and a funeral. In C. Sugrue (Ed.), *The future of educational change: International perspectives* (pp. 15-33). New York, NY: Routledge.



**Stanford Center for Opportunity Policy in Education**  
**Barnum Center, 505 Lasuen Mall**  
**Stanford, California 94305**  
**Phone: 650.725.8600**  
**[scope@stanford.edu](mailto:scope@stanford.edu)**

**<http://edpolicy.stanford.edu>**