

One Piece of the Whole

Teacher Evaluation as Part of a Comprehensive System for Teaching and Learning



BY LINDA DARLING-HAMMOND

The United States is at a critical moment in teacher evaluation. The evaluation process is undergoing extensive changes, some of them quite radical, in nearly every state and district across the country. As we embark on these reforms, it is crucial for schools, teachers, and, especially, students

Linda Darling-Hammond is the Charles E. Ducommun Professor of Education at Stanford University, where she is the faculty director of the Stanford Center for Opportunity Policy in Education and the founding director of the School Redesign Network. She is a former president of the American Educational Research Association and a member of the National Academy of Education. This article is adapted, with permission of Teachers College Press, from Linda Darling-Hammond, Getting Teacher Evaluation Right: What Really Matters for Effectiveness and Improvement. Copyright 2013 by Teachers College, Columbia University. All rights reserved. The book was inspired by 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.

that new policies improve the quality of teaching while avoiding pitfalls that could damage education. It is imperative that we not substitute new problems for familiar ones, but that we instead use this moment of transformation to get teacher evaluation right.

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. 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.

These problems are long-standing. They were obvious when my colleagues and I first studied U.S. teacher evaluation systems in the

early 1980s.¹ As part of a Rand Corporation study, Arthur Wise, Milbrey McLaughlin, Harriet Bernstein, and I searched the country for effective evaluation systems and found ourselves rummaging for the proverbial needle in a haystack. We discovered only a very few that offered opportunities for teachers to set goals and receive regular, useful feedback, along with systems that could support both learning and timely, effective personnel decisions.

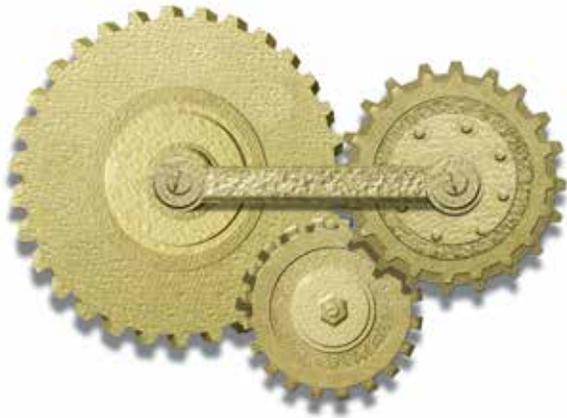
There were some bright spots, like the then-brand-new Toledo Peer Assessment and Review (PAR) model—a labor-management breakthrough that introduced intensive mentoring and peer evaluation for both novice teachers and struggling veterans, and that ensured serious decisions for tenure and continuation.* Also noteworthy was the Greenwich, Connecticut, model of teacher goal-setting and continuous feedback—which involved teachers in collecting evidence about their practice and student learning long before this was fashionable elsewhere. Although the use of some of these successful models has spread, the broad landscape for teacher evaluation has changed little, and impatience with the results of weak systems has grown.

As my colleagues and I found in our research nearly 30 years ago, and as I experienced as a high school teacher some years ago

promotion, advanced certification, and dismissal. As a consequence, most states in the country are in the process of dramatically overhauling their evaluation systems for both teachers and administrators.

Although there is widespread consensus that teacher evaluation in the United States needs serious attention, simply changing on-the-job evaluation will not, by itself, transform the quality of teaching. For all of the attention focused on identifying and removing poor teachers, we will not improve the quality of the profession if we do not also cultivate an excellent supply of good teachers who are well prepared and committed to career-long learning. And teachers' ongoing learning, in turn, depends on the construction of a strong professional development system and useful career development approaches that can help spread expertise. Finally, improving the skills of individual teachers will not be enough: we need to create and sustain productive, collegial working conditions that allow teachers to work collectively in an environment that supports learning for them and their students.

In short, what this country really needs is a conception of teacher evaluation as part of a *teaching and learning system* that supports continuous improvement, both for individual teachers



This country needs a conception of teacher evaluation as part of a *teaching and learning system* that supports continuous improvement.

myself, most teachers want more from an evaluation system. They crave useful feedback and the challenge and counsel that would enable them to improve. Far from ducking the issue of evaluation, they want more robust systems that are useful, fair, and pointed at productive development.

Today, teacher evaluation is receiving unprecedented attention, in large part because new teacher evaluation systems are a requirement for states and districts that want to receive funding under the federal Race to the Top initiative or flexibility waivers under No Child Left Behind. As teaching has become a major focus of policy attention, teacher evaluation is currently the primary tool being promoted to improve it. Federal requirements include the use of multiple categories of teacher ratings, rather than just “satisfactory” or “unsatisfactory,” based on multiple observations, feedback, and the use of student test scores to assess effectiveness. They also encourage the use of these evaluations to inform decisions about tenure and continuation, compensation,

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.

Of all the lessons for teacher evaluation in the current era, perhaps this one is the most important: that we not adopt an individualistic, competitive approach to ranking and sorting teachers that undermines the growth of learning communities. Research shows that student gains are most pronounced where teachers have greater longevity and work as a team.² (See the sidebar on page 6 for an example of how this collective approach can work.) At the end of the day, collaborative learning among teachers will do more to support student achievement than dozens of the most elaborate ranking schemes ever could.

How Should We View the Improvement of Teaching?

Some proponents of teacher evaluation reforms have conjectured that if districts would eliminate the bottom 5 to 10 percent of teachers each year, as measured by value-added student test scores, U.S. student achievement would increase by a substantial

*To learn more about peer assistance and review, see “Taking the Lead: With Peer Assistance and Review, the Teaching Profession Can Be in Teachers’ Hands,” by Jennifer Goldstein, in the Fall 2008 issue of *American Educator*, available at www.aft.org/pdfs/americaneducator/fall2008/goldstein.pdf.

amount—enough to catch up to high-achieving countries like Finland.³ However, there is no real-world evidence to support this idea and quite a bit to dispute it.

In fact, high-achieving Finland* does not do what these advocates propose. Rather than focusing on firing teachers, it has one of the strongest initial teacher education systems in the world, and leaders credit that system with having produced nationwide improvements in student learning.⁴ There is relatively little emphasis in Finland on formal on-the-job evaluation, and much more emphasis on collaboration among professionals to promote student learning. In truth, we cannot fire our way to Finland. If we want to reach the high and equitable outcomes it has achieved in recent years, we will have to *teach* our way to stronger student learning by supporting teachers' collective learning.

Despite the current focus on in-service evaluation, a highly skilled teaching force results from developing well-prepared teachers from recruitment through preparation via ongoing professional development. Support for teacher learning and evaluation needs 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 stan-

*For more on education in Finland, see "The Professional Educator: Lessons from Finland," by Pasi Sahlberg, in the Summer 2011 issue of *American Educator*, available at www.aft.org/pdfs/americaneducator/summer2011/Sahlberg.pdf.

dards that are linked to student learning standards, curriculum, and assessment, thereby 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 multifaceted evidence of teachers' contributions to student learning and to the school as a whole. And it should create the structures that make good evaluation possible: time and training for evaluators, the support of master or mentor teachers to provide needed expertise and assistance, and high-quality, accessible learning opportunities supporting effectiveness for all teachers at every stage of their careers.

If learning to teach is to be a cumulative, coherent experience, a common framework should guide a comprehensive system that addresses a variety of purposes:

- Initial and continuing teacher licensing;
- Hiring and early induction;
- Granting tenure;
- Support for supervision and professional learning;
- Identification of teachers who need additional assistance; and
- Recognition of expert teachers who can contribute to the learning of their peers, both informally as colleagues and formally as mentors, coaches, and teacher leaders.

The system must also allow for the fair and timely removal of teachers who do not improve with feedback and assistance. It may also be asked to support decisions about compensation, as policymakers are increasingly interested in tying compensation

When Evaluation Supports a Collective Perspective

Lynne Formigli, a National Board Certified Teacher in science and a leader in her local union, describes how participating in the alternative evaluation program in the Santa Clara Unified School District helped her reach her goal of improving student writing and learn much more in the process:

I teamed up with a seventh-grade writing teacher and an eighth-grade writing teacher. Our focus was on how we teach writing at different grade levels. We spent time observing each other teaching the writing process. Afterward, we met and compared our observations. We came away with specific ways to improve our students' writing, as well as ideas for integrating writing throughout all grade levels and subjects. Observing other teachers helped me understand how critically important modeling is, allowing me to overcome my fear of giving students the answers when I give them examples.

We found that in our search to help students be more effective communicators, we had all developed similar tools to scaffold their writing. During our discussions, we were excited to consider

the impact on our students if we standardized the tools we use, so students would recognize them from class to class, grade level to grade level. As we continue to work toward that goal as a school, we have the added benefit of increased communication and collaboration among teachers. The end result is of great benefit to the students we teach every day.

Formigli's principal also learned from the experience. After Formigli and her two colleagues presented a summary of their work and a reflection on the process, he wrote in his formal evaluation narrative:

At the middle school level, it is beneficial when students can see a common strand run through their instructional day. When something learned in science is tied to something learned in English, both make more sense. When instruction is coordinated from subject to subject and then from one grade level to the next, we not only have good education, we have magic. And that is what Lynne [and her colleagues] created. ... Participating in

the reflective discussion related to the alternative evaluation project was an evaluation-supervision highlight for me. We spoke about the writing process, genre, cross-grade and cross-subject education, staff development opportunities, standards, the need to share learning experiences, validation, and a host of other things.

It is possible for evaluation to be structured in ways that support this collective perspective. However, it is equally possible for individually focused and competitively oriented evaluation and compensation practices to undermine collegial work, harming the chances for professional sharing and learning. If teachers are ranked and if rewards are competitively allocated, evaluation is likely to undermine efforts toward collective improvements, to the ultimate detriment of teacher and student learning.

Source: Accomplished California Teachers, *A Quality Teacher in Every Classroom: Creating a Teacher Evaluation System That Works for California* (Stanford, CA: National Board Resource Center, Stanford University, 2010). As featured in Linda Darling-Hammond, *Getting Teacher Evaluation Right* (2013).

to judgments about teacher effectiveness, either by differentiating wages or by linking such judgments to specific responsibilities and salary increments for more expert teachers. An approach that supports the development and sharing of greater expertise, rather than one that fosters competition and isolation, holds the most promise for improving teaching and learning overall.

Understanding Teacher Quality and Teaching Quality

In building a system, it is important not only to develop skills on the part of individual practitioners, but also to create the conditions under which practitioners can use their skills appropriately. The importance of this is easily seen if we think of medicine, where both the professional skills and professional contexts are relatively well developed through licensing of doctors and accreditation rules for hospitals, the places where many physicians practice.

It would do little good to prepare doctors through intensive residencies in their specialty area if pediatricians could be assigned to cardiac surgery or ophthalmologists were asked to treat spinal injuries. If out-of-field assignment were allowed (as it too often is in teaching), the quality of medical care would suffer even if individual doctors were highly skilled in their fields. Similarly, a cardiologist supported by the latest technology and medical resources

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 he or she is asked to teach a flawed curriculum or lacks appropriate materials. 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 with classes that are far too large. Conversely, a less skilled teacher may be buoyed up by excellent materials, strong peer support for lesson planning, and additional specialists who work with students needing extra help.

The extent to which teachers experience dissimilar teaching conditions—and students experience very different learning conditions—has been made clear in the school finance lawsuits brought in many states, which describe in vivid terms the differences between rich and poor schools. In *Williams v. California*, for example, teachers, parents, and students from low-income communities described overcrowded schools that had to run multiple shifts each day and multiple shifts during the school year, alternating on-months and off-months for different cohorts of students



Initiatives to develop teaching quality must consider not only how to identify, reward, and use teachers' skills and abilities, but also how to develop contexts that enable good practice.

is clearly more effective than one who has no access to heart monitors, surgical equipment, defibrillators, or medication. The quality of care is determined equally by the skill of physicians and the resources that are available to them to do their jobs.

Similarly, if one wants to ensure high-quality instruction, it is important to attend to both *teacher* quality and *teaching* quality. Teacher quality might be thought of as the bundle of personal traits, skills, and understanding an individual brings to teaching, including dispositions to behave in certain ways, such as collaborating with colleagues and adapting instruction to help students succeed. Teaching quality, as distinct from teacher 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, including factors aside from what the teacher knows and can do.

cycling in and out of the building; classrooms with more than 40 students without enough desks, chairs, and textbooks for each student to have one; lack of curriculum materials, science equipment, computers, and libraries; and crumbling facilities featuring leaky ceilings and falling ceiling tiles, sometimes overrun with rodents, and lacking heat and air conditioning. Not surprisingly, these underresourced schools also had high levels of teacher turnover, making it difficult to create a coherent curriculum or develop common practices to support student learning.⁵

These kinds of conditions can undermine the effectiveness of any teacher. Even where teachers have equivalent skills, there is little doubt that the quality of instruction is greater in a school with high-quality and plentiful books, materials, and computers; a coherent, well-designed curriculum; well-lit, properly heated, and generously outfitted classrooms; small class sizes; and instructional specialists, than it is when students must learn in overcrowded, unsafe conditions with insufficient materials, poorly chosen curriculum, large classes, and no instructional supports.⁶

Strong teacher quality may heighten the probability of effective teaching, but it 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. 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 Evaluating and Supporting Teaching

We need a more systemic approach to building and sustaining teacher effectiveness. Despite the apparent single-minded emphasis on teacher evaluation from some policy quarters, the importance of a more comprehensive approach is gaining currency. For example, 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.⁷

A high-quality teacher evaluation *system* should create a coherent, well-grounded approach to developing teaching, created 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 elements:

1. Common statewide standards for teaching that are related to meaningful student learning and are shared across the profession;
2. Performance-based 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 properly trained evaluators, mentoring for teachers who need additional assistance, and fair decisions about personnel actions; and

A Glimpse into High-Quality Evaluation

Louisa, a fourth-year science teacher, sits down to discuss her teacher development portfolio with her evaluator. Her portfolio by now contains documentation and analysis of her work from the end of her preservice program through her first three years in the classroom. It also contains records and assessments of professional development projects she has done over the last three years. Louisa and her evaluators had selected these projects at different times in her first years of teaching to help her attend to the needs they identified together.

Susannah, who is Louisa's current evaluator, is a 15-year veteran science teacher at the same school. She is released from her classroom duties for three periods each day to work as a member of the district evaluation team. In that role, Susannah observes her colleagues, prepares written evaluations, meets with teachers to discuss or plan observations, and attends meetings where the district team reviews evaluations and individual professional development plans. The district evaluation team is composed of accomplished classroom teachers, administrators from each school site, and the district Peer Assistance and Review coordinator. The team's job is to review the evaluations of teachers to ensure that each of them is meeting performance expectations, progressing along the teacher development continuum, and receiving good counsel about ways to improve. When there are serious concerns about a teacher's performance, the team sends in another evaluator to validate the concern and help the team recommend a course of action that may range from targeted coaching to dismissal.

Louisa opens her observation notebook to the page that contains notes about the lesson that Susannah observed the previous day. Susannah has already given Louisa a copy of the observation notes she made and questions for Louisa to think about before they meet. Louisa has added some reflections about the lesson and questions she wants to explore with Susannah. Louisa has brought some writing her students did that morning in response to a question she posed when they came into class. Susannah asks Louisa for her own assessment of the lesson and, in particular, how she thinks the discussion went. Louisa is very proud that during the discussion, she had to interject to clarify questions only three times. She points to evidence in the discussion of the content mastery students showed. However, there is a discrepancy between what occurred during the discussion and evidence of content mastery in the students' writing that Louisa has brought along.

In her observation notes, Susannah cites many of the same kinds of evidence that Louisa has discussed. She points out that the students still struggle to explain their thinking clearly. She directs Louisa's attention to the students' use of questions to one another and their limited reference to the informational texts they had read. This is an "aha moment" for Louisa.

"Oh," she says, "this is what we've talked about when we have been trying to figure out why the kids do poorly on comprehension questions on informational texts!" She is referring to the meeting they had after they had looked at some of the school's standardized test data alongside other assessments. Louisa had complained several times about how few

questions her students asked about their reading and how literal their conversations about their reading often were. She suggested that students' lack of questions might well be related to their ability to pose questions about the text as they read.

Susannah reminds Louisa that inquiry in science means being able to ask "why?" at the appropriate times. Louisa knows this and recognizes that posing questions while reading is a way readers probe their own understanding. If students were not doing that during reading, then very likely they would not notice that their own written or verbal explanations did not offer the receiver opportunities for clear understanding.

"What should I do about this?" Louisa asks. Susannah suggests that Louisa and her colleagues, who have been doing some research on students' reading in science, invite one of the English teachers, who has taught reading to English language learners for several years, to come to their next research meeting to help them explore strategies to try with their own students.

Susannah's role will be to focus her observations on helping Louisa reflect on the success of the strategies she uses. As Susannah looks for evidence of teaching standards in Louisa's work this year, they agree that Louisa should focus on the effective teaching skills that she brings to solving this problem. They conclude by filing the observations, the records of their conversations, and agreements in the year 4 section of Louisa's portfolio. Thus begins a new chapter in Louisa's documentation of her professional journey.

5. Aligned professional learning opportunities that support the improvement of teachers and teaching quality.

Each of these five elements should operate within a system that supports effective teaching and learning.

It is easy for procedures to overwhelm purpose in almost any reform, and this is particularly true for teacher evaluation. As states and districts develop new approaches, it will be important for them to think strategically about how to accomplish their goals—putting in place the necessary systems and supports that allow educators to focus productively on improving teaching. As new practices are implemented, districts will also need to study and refine them, always mindful of keeping their eyes on the prize: more responsive and effective teaching in each classroom and across the school as a whole.

This focus on effective instruction has taken on a new sense of urgency as the pressures for improved student achievement have intensified. As a result, many initiatives to measure and improve teaching effectiveness through evaluation have emerged. 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 instruction focused on meaningful 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.

Several important conditions are necessary to create productive systems: (1) state licensing systems must be coordinated with local evaluation; (2) evidence about teachers' practice must be integrated with appropriate evidence about student learning; and (3) evaluations must be connected with both individual and collective professional learning. Where these elements are in place, the evaluation experience can support the development of sophisticated teaching. (See the sidebar on page 8, which

Louisa's case illustrates the learning that a coordinated evaluation and support system could produce. As a fourth-year teacher, Louisa has been developing her skills and documenting her practice around the same teaching standards from her preservice program throughout her first three years in the classroom. The portfolio

on the same teaching standards that are used to accredit Louisa's preparation program, so her training was organized to ensure that she would master the tested knowledge and skills. The assessment helped strengthen her preparation and her readiness to teach. The coherence of her experience was further enabled by the

entrants meet them—should complement the role of local districts, making it more possible for them to support the ongoing development of teachers who have met that initial bar.

Louisa's case also illustrates how the evaluation process can connect evidence of practice to evidence of student learning in



Support for teacher learning and evaluation needs to be part of an integrated whole that promotes effectiveness during every stage of a teacher's career.

she has maintained began with the performance assessment she completed at the end of preservice preparation to illustrate her ability to plan, teach, and assess students around the state student learning standards—and to reflect on her practice and outcomes in light of the state's standards for teaching.

This seamless experience was facilitated by an overhaul of the state system to require a teacher performance assessment for licensing, raising the bar for entry with a valid and authentic measure of whether new entrants can practice responsibly. The assessment (in this case, the Performance Assessment for California Teachers) is based

extension of these standards into her induction program and later on-the-job evaluation.

Creating coherence from preparation to practice will greatly improve the capacity of the teaching force. States such as Massachusetts, Minnesota, Ohio, and Washington are among those that have taken steps forward to create such coherence, by adopting performance assessments for licensing beginning teachers that are linked to standards for initial induction and ongoing evaluation. The role of the state—to establish professional standards and ensure, through profession-wide assessments for licensing, that all new

ways that move teaching forward. By looking at standardized test data, Louisa's department highlighted some areas for further exploration that might better support achievement. By looking, then, at authentic student work in the context of her current teaching, Louisa was able, with help from her evaluator, to see more clearly how her students were thinking and understanding, and to fine-tune her plans to strengthen their learning.

Source: Accomplished California Teachers, *A Quality Teacher in Every Classroom: Creating a Teacher Evaluation System That Works for California* (Stanford, CA: National Board Resource Center, Stanford University, 2010). As featured in Linda Darling-Hammond, *Getting Teacher Evaluation Right* (2013).

describes how these system elements can work together.) I elaborate on each of these aspects below.

Entering the Profession: Coordinating State Licensing and Local Evaluation

One of the reasons for current concerns about the capability of some members of the teaching force is the public perception that teacher education and licensing systems do not routinely guarantee competence when teachers enter the profession. Furthermore, there is a large disjuncture in most states between the standards used to guide preparation and licensing and those that come into play when teachers are on the job.

Fixing these problems is critical to developing a strong teaching profession. A profession is defined by having all entrants master a common body of knowledge and skills, grounded in research, reflected in professional standards, and used to advance clients' welfare. Professions enforce these standards through

Because teacher licensing tests, which are currently focused largely on basic skills and subject-matter knowledge, have not provided a meaningful assessment of capacity to teach before entry, teaching has lacked this key element of a profession. The lack of a meaningful entry bar also means that the burden has fallen on school districts to figure out whether new teachers have mastered the basics for the classroom.* In teaching, it's time to create performance-based assessments for licensure and then to apply the same professional standards to local evaluation. This approach to assessment has been at the heart of recent recommendations from the two largest national teachers' unions. In *Transforming Teaching*, the National Education Association called for a career continuum based on national professional teaching standards that guide preparation and teacher performance assessments completed before licensure.⁸ In *Raising the Bar*, the American Federation of Teachers called for a "bar exam" for teaching that offers a nationally available performance assess-



For teaching to be comparable to other professions, we need clear professional standards both for state licensing and for on-the-job evaluation.

licensing examinations that measure the capacity to apply knowledge responsibly—such as the bar exam in law, licensing examinations in medicine, and the portfolios required for architectural registration.

Professional licensing and certification assessments are administered outside of the context of preparation or employment, so that they represent the knowledge and skills of the field as a whole, not just the views of a particular institution. They are scored by professionals who are trained to a common standard. The assessments also exert influence over preparation programs, because they help define the curriculum to be taught as they instantiate much of the knowledge and many of the skills candidates are supposed to learn. In the employment context, local institutions, such as hospitals, law firms, and architectural firms, make the judgments of competence, but they use the standards of the profession to establish whether professionals have engaged in appropriate practice or malpractice.

For teaching to be comparable to other professions, we need clear professional standards against which teachers are assessed both for state licensing and for on-the-job evaluation. These should be reflected in a continuum of performance assessments that validly and reliably measure actual teaching performance at key career junctures—initial licensing, the achievement of the professional license, and the designation of accomplished practice—as well as in on-the-job evaluation systems.

ment for licensure, along with evidence of competence in the subject area and strong clinical training.⁹

The InTASC (Interstate Teacher Assessment and Support Consortium) standards, adopted by more than 40 states, undergird new performance-based assessments for entry that have been developed by the profession—that is, by teachers and teacher educators across the country. These include, for example, the Performance Assessment for California Teachers (PACT), the Oregon Teacher Work Sampling (TWS) System, and the Teacher Performance Assessment (edTPA) recently piloted in more than 20 states.

Furthermore, some states have envisioned a continuum in which beginning teachers are evaluated using performance assessments for initial and continuing licensure, and veteran teachers are considered for higher pay and leadership roles based in part on National Board Certification or similar assessments. Massachusetts, Ohio, and Washington are among the states that have created plans for a continuum of performance assessments to guide the teaching career. In 2003, New Mexico created a three-tiered licensure system at the state level, with locally aligned evaluations for on-the-job evaluation. Using a set of portfolios

*For more on entering the teaching profession, see "The Professional Educator: A New Path Forward," by Randi Weingarten, in the Spring 2010 issue of *American Educator*, available at www.aft.org/pdfs/americaneducator/spring2010/Weingarten.pdf.

modeled on that of the National Board, teachers must demonstrate increasing competence to progress from Provisional Teacher (the first three years) to Professional Teacher to Master Teacher. Each level is accompanied by increased compensation and responsibilities.¹⁰

Such an aligned system focuses teachers on what their students learn as a result of their teaching decisions, and on how to improve their effectiveness. Teachers feel they are learning as they both develop their own portfolios and score those of other teachers when they are part of the state scoring team. They also learn as they receive feedback on their work from colleagues, made more useful by the common language teachers are developing around their practice. And because yearly district evaluations are based on the same standards as the licensing assessments, teachers can continue to work on their practice coherently throughout their careers.

evaluators see firsthand what students know and can do before, during, and as a result of teaching. This evidence is directly associated with the curriculum and teaching goals, and it can include vivid examples of student thinking, reasoning, and performance on a wide range of knowledge and skills.

Although standardized test scores can give a general idea of the level of student achievement (typically limited to items that ask for recognition of information), the scores they report do not offer detailed insights into what students think or what they know how to do in practice. The scores that result from most current state tests are limited by the inability of the tests to assess achievement that requires communication, research, the production of new ideas, or the application of knowledge to new problems or situations. In addition, value-added measures based on these tests, which are not designed to measure achievement that is well above or below grade level, are both unstable and biased for teachers who serve certain groups of students. Finally, it is nearly impossible to attri-



Evaluations relying on a single test-based metric sitting in isolation alongside a rating based on classroom observations are not particularly helpful in either understanding or improving the quality of teaching, and may be harmful.

On-the-Job Evaluation: Integrating Evidence of Practice with Evidence of Student Learning

On-the-job evaluations should be based on the same teaching standards as performance assessments for entry. Furthermore, they should evaluate teacher effectiveness based on multiple measures of both practice and outcomes that are considered in an *integrated* fashion, including:

- Classroom observations and examination of other classroom evidence (e.g., lesson plans, student assignments, and work samples) using a standards-based instrument that examines planning, instruction, the learning environment, and student assessment;
- Evidence of student learning on a range of valid assessments that appropriately evaluate the curriculum and the students the teacher teaches, including students with special education needs and English language learners; and
- Teachers' contributions to colleagues and to the school. Connected, ongoing, high-quality professional learning opportunities should build strong professional learning communities and enable teachers to meet the standards.

Integrating authentic, rich evidence of student learning with the processes of evaluation—at the stage of goal-setting, throughout the course of the year, and at the end of teaching cycles (a year, a semester, or a unit of study)—can help teachers, mentors, and

bute student gains in test scores to a single teacher or to disentangle them from the many other influences on student learning, as well as the composition of the classroom.

Thus, evaluation systems that rely on a single test-based metric sitting in isolation alongside a rating based on classroom observations are not particularly helpful in either understanding or improving the quality of teaching, and may be harmful. Quite often, the two measures do not agree with one another, and the variations in the value-added metric are more related to changes in classroom composition—which students are assigned—than they are to any specific changes in teaching practice. A single test measure used for all teachers will, in some cases, also be invalid for particular students or a poor measure of the specific curriculum being taught.

To be useful, measures of teaching outcomes must be considered in a more nuanced analysis that is connected to the curriculum and students being taught, as well as to the practice of the teacher being evaluated. These measures may include test scores of various kinds, with greater weight placed on those that are the most direct measures of the content being studied and on those that are most appropriate for the students in the classroom. Measures should also include student work drawn from specific undertakings in the classroom that can be analyzed in terms of teachers' practices focused on particular learning goals. This kind of work can be used to closely evaluate the teaching-learning cycle and transform how

teachers think about and enact their practice. This approach is used in districts like Long Beach and San Mateo, California, and is encouraged in states like Massachusetts, Oregon, and Washington that call for multiple measures of student learning to be combined in a judgment system with evidence of teacher practice.

A recent study from the Consortium for Policy Research in Education describes the importance of connecting information on teacher practice to information on student learning. The study looked at whether instruction and student outcomes would be influenced by having teachers discuss evidence about their practice, derived from classroom observations, along with student

Although it may seem simpler in the short run to make teacher decisions based largely on a single set of student scores, this approach has thus far produced more heat than light in analyses of teaching, often creating greater confusion where more clarity is needed. Unskilled use of this kind of test score data can have damaging ramifications due to the miscalculation and potential loss of good teachers and the incentives for teachers to avoid the neediest students. Although attention to learning outcomes is important, the greatest benefits will be secured where multiple measures of learning are combined with evidence of practice to paint a meaningful picture of how teaching influences student progress.



Multiple measures of learning combined with evidence of practice paint a meaningful picture of how teaching influences student progress.

learning data. Compared with a control group of teachers who only discussed student data, the group that received feedback about their teaching in the same sessions where they discussed student learning data with colleagues exhibited more changes in their later instructional strategies of the kind emphasized in the feedback, and their students experienced significantly greater learning gains.¹¹

In this aspect of evaluation, especially, it is important to keep in mind that our goal is not to rank teachers on a single scale. It is to support high-quality instruction for all students—instruction that is well informed by a sophisticated understanding of what students are learning and how teaching can support their progress.

To accomplish this, we need more than valid instruments and tools to assess teaching. We also need structures that enable fair,

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 multifaceted 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 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.

—L.D.H.

effective evaluation by ensuring evaluator training; expert teachers who can provide intensive assistance to teachers in need; governance structures that oversee the process and enable timely, well-grounded personnel decisions; and resources that can support a manageable system. And finally, teachers should participate in developing the system and in the governance structure that supports the ongoing decision-making processes. These conditions address not only evaluation instruments or procedures, but also the policy systems in which they operate and the school-based conditions that are needed to stimulate continuous learning and improvement.

knowledge, build their skills, and improve instruction.¹³ For example, a comprehensive five-year study of 1,500 schools undergoing major reforms found that in schools where teachers formed active professional learning communities, achievement increased significantly in math, science, history, and reading, while student absenteeism and dropout rates were reduced. Further, particular aspects of teachers' professional community—a shared sense of intellectual purpose and a sense of collective responsibility for student learning—were associated with a narrowing of achievement gaps in math and science among low- and middle-income students.¹⁴



Teaching improves most in collegial settings where common goals are set, curriculum is jointly developed, and expertise is shared.

Learning Together: The Critical Importance of a Collective Perspective

I cannot stress enough that teaching improves most in collegial settings where common goals are set, curriculum is jointly developed, and expertise is shared. Although individual teacher evaluation can be a part of an educational improvement strategy, it cannot substitute for ongoing investments in the development and dissemination of profession-wide knowledge through pre-service preparation and work in professional learning communities.

Collegiality is encouraged when teachers' contributions to school improvement and collaboration with peers and parents are valued among the evaluation criteria, and when opportunities for analyzing teaching and learning are taken up by teaching teams and interwoven with opportunities for peer coaching and planning. Productive professional learning and effective coaching require communal engagement in sustained work on instruction over time. Successful practices also engage teams of teachers and administrators in the design and governance of the evaluation system, so that everyone develops shared standards of practice and a collective perspective on how to improve the work.

Research shows that when schools are strategic and persistent in creating productive working relationships within academic departments, across them, or among teachers schoolwide, the benefits can include greater consistency in instruction, more willingness to share practices and try new ways of teaching, and more success in solving problems of practice.¹² Perhaps the simplest way to break down professional isolation is for teachers to observe each other's teaching and to provide constructive feedback. Several large-scale studies have identified specific ways in which professional community-building can deepen teachers'

Strong professional learning communities require leadership that establishes a vision, creates opportunities and expectations for joint work, and finds the resources needed to support the work, including expertise and time to meet.¹⁵ Collaborative teacher teams can improve practice together by:¹⁶

- Examining data on student progress;
- Analyzing student work;
- Determining effective strategies to facilitate learning;
- Designing and critiquing curriculum units and lessons;
- Observing and coaching one another; and
- Developing and scoring common classroom-based assessments to measure progress.

Over time, this work can be more deeply supported if professional learning opportunities are conceptualized as part of a career continuum that encourages teachers to gain and share expertise. Productive career ladders (or lattices) can also create 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.

The lack of time for collaborative planning in most U.S. schools gives teachers few opportunities to develop sophisticated practice, although some restructured schools have redesigned the use of time and resources to support students and teacher learning with longer periods, shared planning time, and extensive ongoing professional development. It is possible to create the context for teachers to become more effective, but it may require thinking differently about some of the traditional "regularities of schooling."¹⁷

(Continued on page 44)

One Piece of the Whole

(Continued from page 13)

Comprehensive, coherent systems of teacher development and evaluation are needed to meet our goals of a high-quality education for all students. The key features of such systems (see the box on page 12) do exist in many schools and districts, although few places have stitched together all the components in a single tapestry. That is the critical work ahead. □

Endnotes

1. Arthur E. Wise, Linda Darling-Hammond, Milbrey W. McLaughlin, and Harriet T. Bernstein, *Teacher Evaluation: A Study of Effective Practices* (Santa Monica, CA: Rand Corporation, 1984).
2. C. Kirabo Jackson and Elias Bruegmann, "Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers" (NBER Working Paper No. 15202, National Bureau of Economic Research, Cambridge, MA, 2009).
3. Eric Hanushek, "Lifting Student Achievement by Weeding Out Harmful Teachers," *Eduwonk*, October 31, 2011, www.eduwonk.com/2011/10/lifting-student-achievement-by-weeding-out-harmful-teachers.html.
4. For discussions of Finland's approach to teaching, see Linda Darling-Hammond, *The Flat World and Education: How America's Commitment to Equity Will Determine Our Future* (New York: Teachers College Press, 2010); and Pasi Sahlberg, *Finnish Lessons: What Can the World Learn from Educational Change in Finland?* (New York: Teachers College Press, 2012).
5. Jeannie Oakes, "Investigating the Claims in *Williams v. State of California*: An Unconstitutional Denial of Education's

Basic Tools?," *Teachers College Record* 106, no. 10 (2004): 1889–1906.

6. Oakes, "Investigating the Claims."

7. National Association of State Boards of Education (NASBE), *Gearing Up: Creating a Systemic Approach to Teacher Effectiveness* (Arlington, VA: NASBE, 2011).

8. National Education Association, *Transforming Teaching: Connecting Professional Responsibility with Student Learning* (Washington, DC: NEA, 2012).

9. American Federation of Teachers, *Raising the Bar: Aligning and Elevating Teacher Preparation and the Teaching Profession* (Washington, DC: AFT, 2012).

10. Darling-Hammond, *The Flat World and Education*.

11. Jonathan Supovitz, *The Linking Study—First Year Results: A Report of the First Year Effects of an Experimental Study of the Impact of Feedback to Teachers on Teaching and Learning* (Philadelphia: University of Pennsylvania, 2012).

12. Shirley M. Hord, *Professional Learning Communities: Communities of Continuous Inquiry and Improvement* (Austin, TX: Southwest Educational Development Laboratory, 1997); Bruce Joyce and Emily Calhoun, *Learning Experiences in School Renewal: An Exploration of Five Successful Programs* (Eugene, OR: Eric Clearinghouse on Educational Management, 1996); Karen S. Louis, Helen M. Marks, and Sharon Kruse, "Teachers' Professional Community in Restructuring Schools," *American Educational Research Journal* 33, no. 4 (1996): 757–798; Milbrey W. McLaughlin and Joan E. Talbert, *Professional Communities and the Work of High School Teaching* (Chicago: University of Chicago Press, 2001); and Fred M. Newmann and Gary G. Wehlage, *Successful School Restructuring: A Report to the Public and Educators* (Madison, WI: Center on Organization and Restructuring of Schools, 1997).

13. Anthony Bryk, Eric Camburn, and Karen Seashore Louis, "Professional Community in Chicago Elementary Schools: Facilitating Factors and Organizational Consequences," *Educational Administration Quarterly* 35, no. 5 (1999): 751–781; Andrew Calkins, William Guenther, Grace Belfiore, and Dave Lash, *The Turnaround Challenge: Why America's Best Opportunity to Dramatically Improve Student Achievement Lies in Our Worst-Performing Schools* (Boston: Mass Insight Education and Research Institute, 2007); Yvonne L. Goddard, Roger D. Goddard, and Megan Tschannen-Moran, "A Theoretical and Empirical Investiga-

tion of Teacher Collaboration for School Improvement and Student Achievement in Public Elementary Schools," *Teachers College Record* 109, no. 4 (2007): 877–896; Karen Seashore Louis and Helen M. Marks, "Does Professional Community Affect the Classroom? Teachers' Work and Student Experiences in Restructuring Schools," *American Journal of Education* 106, no. 4 (1998): 532–575; and Jonathan A. Supovitz and Jolley Bruce Christman, *Developing Communities of Instructional Practice: Lessons from Cincinnati and Philadelphia*, CPRE Policy Briefs RB-39 (Philadelphia: University of Pennsylvania, November 2003).

14. Newmann and Wehlage, *Successful School Restructuring*.

15. Dan Mindich and Ann Lieberman, *Building a Learning Community: A Tale of Two Schools* (Stanford, CA: Stanford Center for Opportunity Policy in Education, 2012).

16. Professional Development Partnership, "A Common Language for Professional Learning Communities," 2008, www.nj.gov/education/profdev/pd/teacher/common.pdf.

17. Seymour B. Sarason, *The Culture of the School and the Problem of Change* (Boston: Allyn and Bacon, 1982).

INTENTIONALLY
LEFT BLANK



Write to us!

We welcome comments on *American Educator* articles. Address letters to Editor, *American Educator*, 555 New Jersey Ave. NW, Washington, DC 20001, or send comments via email to ae@aft.org. Letters selected for publication may be edited for space and clarity. Please include your phone number or email address so we may contact you if necessary.