

GRADING**THE**GRADERS

A REPORT ON TEACHER EVALUATION REFORM
IN PUBLIC EDUCATION

BY THOMAS TOCH

MAY 2016

Center on the Future of
AMERICAN
EDUCATION
at Georgetown University

About the Author

Thomas Toch is a research fellow at Georgetown University's McCourt School of Public Policy and the founding director of the Center on the Future of American Education. He can be reached at Thomas.toch@georgetown.edu and is on Twitter at [@thomas_toch](https://twitter.com/thomas_toch).

About FutureEd

The forthcoming Center on the Future of American Education is an independent, solution-oriented think tank at Georgetown University's McCourt School of Public Policy, committed to bringing fresh energy to the causes of excellence, equity, and efficiency in K-12 and higher education.

TABLE OF CONTENTS

1	Grading the Graders
3	The New Evaluation Landscape
7	Impact
11	Challenges
20	Duncan’s Gambit
26	Recommendations
30	Endnotes
35	Bibliography
39	Acknowledgments

Grading the Graders

Type “teacher evaluation” into the search engine of the District of Columbia school system’s website and pages of information cascade down the screen: the District’s rating standards, performance categories, technical reports on the role of student test scores in grading teachers, and much more. The same is true of the Tennessee State Department of Education’s site, where the Tennessee Educator Acceleration Model—known as TEAM and used to rate the majority of teachers in Tennessee’s public schools—is parsed in great detail.

The comprehensiveness of the information hardly seems surprising, given the centrality of teachers to the education enterprise and the fact that taxpayers spend upwards of half a trillion dollars a year on public school teacher salaries and benefits.

But until recently, there was scant attention paid to teacher evaluation in American public education. The standard evaluation model for the nation’s 3.1 million teachers was a cursory visit once a year by a principal wielding a checklist, looking for clean classrooms and quiet students—superficial exercises that didn’t even focus directly on the quality of teacher instruction, much less student learning.¹

Because public school teachers have traditionally been hired, paid, and promoted strictly on the basis of their college credentials and their years in the classroom, there were few incentives for school systems to thoughtfully compare teacher performances. And most school systems didn’t. They gave nearly every teacher satisfactory ratings and rarely fired anyone for under performance, while the absence of meaningful measures of teacher quality made rewarding talent and other steps to strengthen the profession nearly impossible to implement.²

Beginning in 2009, policymakers’ stance toward teacher performance changed dramatically. Spurred by a powerful confluence of factors—alarming reports of school districts carrying incompetent teachers on their payrolls,³ studies undermining public education’s culture of credentialism,⁴ the emergence of new ways of measuring teachers’ impact on student achievement,⁵ the ascendancy of vocal reform advocates,⁶ and, above all, forceful federal incentives⁷—nearly every state strengthened teacher evaluation through new mandatory state models, mandates for new local systems, or a menu of state and local options.⁸

Under what has been one of the most rapid and wide-ranging policy responses in the history of public education, forty-six states have demanded more comprehensive teacher scrutiny, and nearly two dozen have directed school districts to weigh teaching performance in addition to experience when granting tenure and the substantial job protections it provides.⁹ Long sought by reformers, these and other changes with potentially far-reaching consequences for teacher quality and student learning would have been unimaginable on such a scale in the past, given public education’s bureaucratic ethos and tradition of industrial-style teacher unionism. As late as 2009, no states

required teacher performance to be part of tenure decisions, the National Council on Teacher Quality reports.¹⁰

But now, the most powerful catalyst of the reform movement—the Obama administration’s financial and regulatory incentives for state and local leaders to take more seriously the task of indentifying

The absence of meaningful measures of teacher quality made rewarding talent and other steps to strengthen the profession nearly impossible to implement.

who in the teaching profession was doing a good job, and who wasn’t—has been eliminated. An improbable but influential alliance of teacher unions wanting to end the new scrutiny of their members and Tea Party activists and congressional Republicans targeting the Obama incentives as part of a larger anti-Washington campaign led lawmakers to end the incentives under the new federal Every Student Succeeds Act and effectively

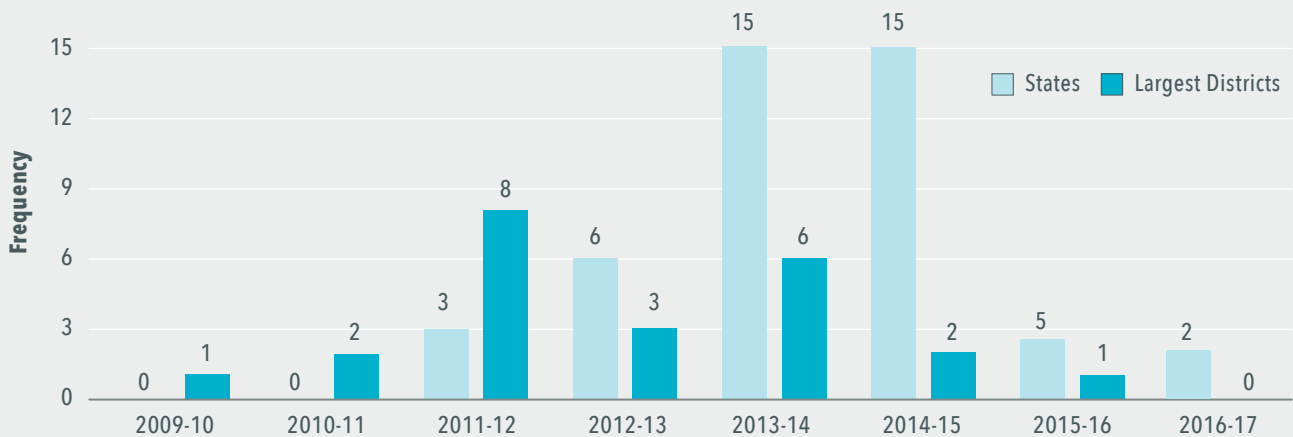
banning the US Secretary of Education from promoting teacher-performance measurement in the future. Once relevant federal regulations expire at the end of August, state and local policymakers may return to the superficial teacher practices of the past.

The dismantling of the Obama reforms has been accompanied by a narrative on both the left and right that the campaign for more meaningful ways to measure teacher performance has been ineffective, more hurtful than helpful to the teaching profession and of scant consequence to students.

But a large and growing body of state and local implementation studies, academic research, teacher surveys, and interviews with dozens of policymakers, experts, and educators all reveal a much more promising picture: The reforms have strengthened many school districts’ focus on instructional quality, created a foundation for making teaching a more attractive profession, and improved the prospects for student achievement.

Many of the new evaluation systems are in early stages and are far from perfect, the research for

The Introduction of New Teacher Evaluation Systems (46 States and 23 of Nation’s 25 Largest School Districts)



SOURCE: Steinberg, Matthew P. and Morgaen L. Donaldson. “The New Educational Accountability: Understanding the Landscape of Teacher Evaluation in the Post-NCLB Era.” Education Finance and Policy. (Forthcoming).

this report makes clear. Technical problems plague many of the new models. Many school leaders are unprepared for the new roles and responsibilities required of them under the reforms, and many school districts have struggled with the price tag of more comprehensive measurement systems. An underlying tension between the use of evaluations to make employment decisions and their potential to help teachers strengthen their performance has left teacher morale badly damaged in some places.

There is also no doubt that the decision of former Education Secretary Arne Duncan to have states stress student test scores in new teacher evaluation systems—while also having states introduce new, more demanding tests linked to the Common Core State Standards—made an already challenging task vastly more difficult and handed opponents an easy way to attack the Obama reforms.

Yet the deployment of new evaluation systems has led to the establishment of clearer teaching standards in many states and school systems. It has forced school leaders to prioritize classrooms over cafeterias and custodians (exposing how poorly prepared many principals are to be instructional leaders) and sparked conversations about good teaching that often simply didn't happen in the past in many schools.

By linking employment to student achievement for the first time in public education's history, enabling smarter staffing decisions, and providing a foundation for new roles and responsibilities for teaching's most talented practitioners, the transformation in teacher evaluation has put teaching, long an occupation of last resort, on the path to becoming a far more vibrant, performance-driven profession.

Importantly, what started as an accountability-driven attempt to remove bad teachers from the profession is, in some states and school systems, increasingly prioritizing ways to help teachers improve their practice through the sorts of systematic feedback they say they want but rarely get in public education—a shift reflecting

an emerging consensus among educators, policy actors, and reformers themselves that the nation can't simply fire its way to a strong teaching force.

The movement's effect on student achievement won't be clear for several years, but early evidence from the places that have had comprehensive evaluation reforms in place the longest is encouraging.

This report examines the changes that the evaluation reform movement has brought to the nation's schools, the consequences of those changes, and emerging solutions to the many and perhaps predictable challenges that have arisen in the course of pursuing fast-paced change at the core of the educational enterprise.

The New Evaluation Landscape

The flimsy checklists of the past have given way in many states and school systems to much more substantial evaluation strategies.

Concerned about relying on superficial standards and principals' personal preferences, the new evaluation systems typically have clearer, more detailed teaching standards and measurement guidelines. The District of Columbia Public Schools' 2009 IMPACT teacher-rating system included a Teaching and Learning Framework that established explicit instructional expectations for the first time in the District's history.

Most states and districts are deploying variations of standards developed in the 1990s by teaching expert Charlotte Danielson, who breaks the craft of teaching into four broad "categories" (planning and preparation, classroom environment, instruction, and professional responsibilities), 22 "themes" (ranging from demonstrating subject knowledge to designing ways to motivate students to learn), and 77 "key skills" (such as when and how to use different groupings of students and the most effective ways to give students feedback on their work).¹¹

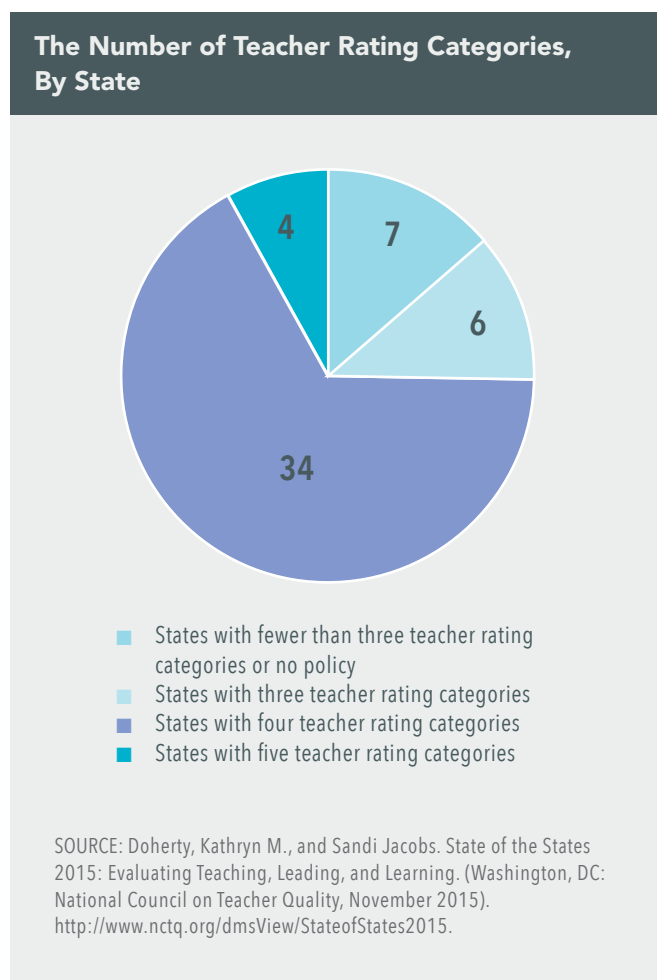
Danielson replaces the traditional binary teacher ratings of “satisfactory” or “unsatisfactory” with four classifications—“unsatisfactory,” “basic,” “proficient,” and “distinguished”—in every area of her evaluation design. Because the Obama administration required states and school districts to do the same under its Race to the Top and NCLB waiver programs, the multiple-category model is nearly universal among the new evaluation systems, with some jurisdictions even adding a fifth performance level. In Washington, DC, teachers are rated either “ineffective,” “developing,” “minimally effective,” “effective” or “highly effective.”

Many states and districts have replaced the single-evaluator, single-metric model of the past—principals observing teachers’ classrooms once a year—with multiple measures, multiple evaluations, and multiple evaluators. By 2013, over three-

dozen states had introduced requirements or recommendations that teachers be evaluated on multiple measures.¹²

The first step in more meaningful evaluation has been to increase the number of classroom visits teachers receive to produce a clearer picture of teacher performance. In Tennessee, for example, it was required that teachers be observed only twice a decade. Now, teachers with over three years of experience have evaluators in their classrooms an average of four times a year, while teachers with less experience typically receive six visits annually. This resulted in nearly 300,000 structured classroom observations of the state’s 66,000 teachers in 2011-12, the first year of the state’s reforms, compared to some 20,000 more informal visits the year before.¹³

The transformation in Tennessee is typical. Researchers Matthew Steinberg of the University of Pennsylvania and Morgaen Donaldson of the University of Connecticut report that states with new evaluation systems require an average of four classroom visits a year.¹⁴



Multiple Raters

Principals still shoulder the bulk of classroom observations in many places. But amidst concerns about principals’ ability to keep up with these additional demands and about the fairness and effectiveness of their evaluations, a growing number of school districts from New Haven to Santa Fe are adopting multi-rater evaluation systems using a combination of administrators, master educators, and even teachers’ peers.¹⁵ “[D]istricts aim not only to relieve principals but, more important, to lend new perspectives, deeper expertise and greater objectivity to the evaluation process,” writes Taylor White in a recent Carnegie Foundation study of the emerging multi-rater systems.¹⁶

New Haven public schools provide extra eyes on low-performers and teachers eligible for performance-pay bonuses by employing a cadre of “third-party validators,” most of whom have worked in various capacities in other nearby districts, but none of whom are teachers or administrators in

the New Haven district.¹⁷ In another model, the District of Columbia system recruits nationally for “master educators,” subject-matter and grade-level specialists who rate teachers throughout the school system in their respective fields.

Similarly, four states—North Carolina, South Carolina, New Jersey, and Maryland—require multiple raters for new teachers and low-performers.¹⁸

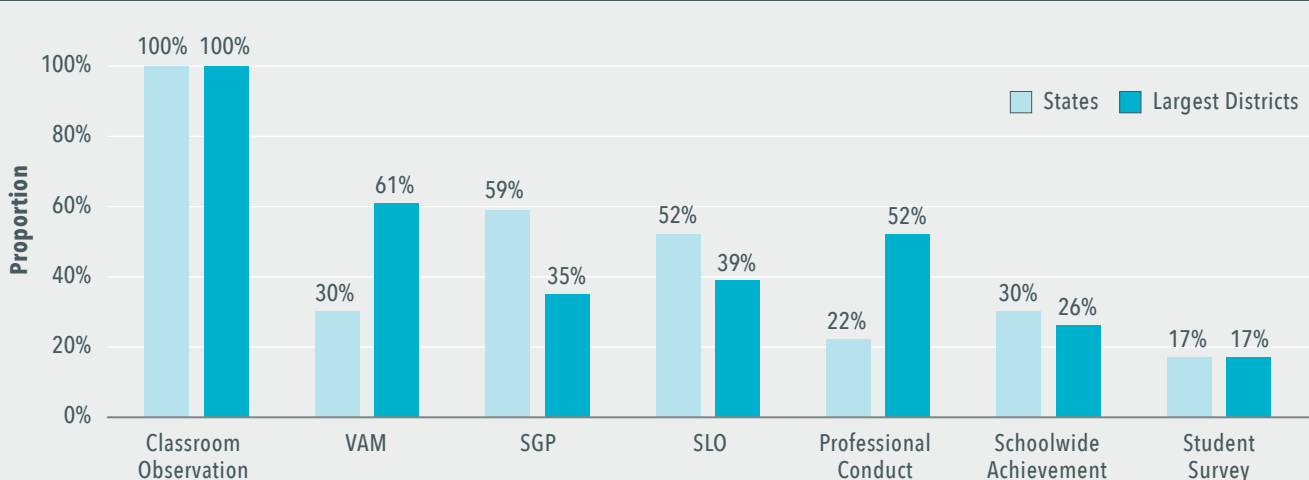
In a sharp break from the tradition of separate labor and management functions introduced (along with collective bargaining) in the 1960s, some school districts are having teachers evaluate colleagues. Several districts in Maricopa County, Arizona, in greater Phoenix, use over three-dozen such educators to evaluate peers who teach the same subjects and grades. Some districts, including Montgomery County, Maryland, and Toledo and Cincinnati in Ohio, have teacher-to-teacher “peer review” programs through agreements with their local teacher unions that predate today’s evaluation reforms.¹⁹

Multiple raters produce more dependable ratings and richer insights for teachers, research suggests. A three-year Measures of Effective Teaching (MET) study conducted by national researchers under the auspices of the Bill & Melinda Gates Foundation concluded that “adding a second observer increases reliability significantly more than having the same observer score an additional lesson.”²⁰ A second study, by researchers John Tyler of Brown University and Eric Taylor of Stanford, found that Cincinnati’s comprehensive multi-rater observation system accurately predicts the achievement of teachers’ future students—and helps raise student achievement in the school district.²¹

Student Achievement

In what quickly became a source of tremendous controversy, early reformers, including US Secretary of Education Arne Duncan, demanded that student achievement be factored into teacher evaluations for the first time, on the grounds that student achievement is the most direct way to measure teacher performance and that it is what matters most in schools.

The Components of New Teacher Evaluation Systems, in 46 States and 23 of the Nation’s 25 Largest School Districts, 2014



[**VAM:** value-added measures **SGP:** student growth percentiles **SLO:** student learning objectives]

SOURCE: Steinberg, Matthew P. and Morgaen L. Donaldson. “The New Educational Accountability: Understanding the Landscape of Teacher Evaluation in the Post-NCLB Era.” Education Finance and Policy. (Forthcoming).

Yet they couldn't simply match one teacher's student scores against another's and hope to produce comparisons with any credibility. Many factors other than teaching influence student achievement, including prior teachers, family income, and levels of parental education.

The early reformers sought to address these realities through complex "value-added" calculations designed to level the playing field by removing factors outside teachers' control from the evaluation equation.

Adapting methodologies pioneered by University of Tennessee agricultural statistician William Sanders in the late 1980s and early 1990s, researchers working for the Tennessee Department of Education and the District of Columbia Public Schools started using predictions of students' standardized test scores based on their previous years' scores in order to rate teachers. If a teacher's students outperformed expectations, the teacher received an above-average value-added rating; underperforming students earned teachers lower scores.²² Now, most states or school districts engaged in teacher-evaluation reform use the strategy, with value-added scores (or a similar measure, student growth percentiles) counting for between 20 and 50 percent of teachers' ratings.²³

Yet only about 30 percent of the nation's teachers teach subjects or at grade levels with sufficient statewide standardized testing to generate value-added scores. Some jurisdictions, including Tennessee and Florida, have compensated by generating school-level value-added calculations and applying the results to the teachers of non-tested grades and subjects. Others have introduced new standardized end-of-course tests to provide the raw material of value-added scores.

But mostly, policymakers have plugged the gaps in standardized testing more informally, by measuring students' progress toward a wide range of learning goals set by teachers and their principals, known as student learning objectives.

These various new, largely untested, and imperfect approaches to gauging teachers' contributions to student achievement have plagued the reform movement technically and politically.

Student Surveys

A third, increasingly common component of the new generation of teacher-evaluation systems has been student surveys of teacher performance. With the help of companies like Tripod, Panorama Education, and My Student Survey, school systems are increasingly gauging teachers' effectiveness through their students' responses to such questions as "In this class, we learn a lot almost every day" (elementary schools), "My teacher wants me to explain my answers" (upper elementary schools), and "My teacher takes the time to summarize what we learn each day" (middle and secondary schools).²⁴

Tennessee school districts employing nearly a quarter of the state's teachers were using student surveys by 2013-14. The responses determined several percentage points of teachers' ratings,²⁵ a typical attribution, Steinberg and Donaldson report.²⁶ But in some places they're a larger factor: In Chicago, survey results count for 10 percent of teachers' evaluations, in Pittsburgh, 15 percent.²⁷

The Gates-funded MET study found Tripod's survey results to gauge teacher performance as effectively as classroom observations and value-added metrics. The research organization Mathematica Policy Research reached the same conclusion in a 2014 study of the student surveys, value-added scores, and observations in the Pittsburgh schools' new evaluation system.²⁸

Not surprisingly, the research also reveals that comprehensive teacher-evaluation models are stronger than the sum of their parts. "Multiple measures produce more consistent ratings than student achievement measures alone," the MET report noted.²⁹ And teachers say they get more out of comprehensive evaluations. In a 2013 survey of 20,000 teachers by Scholastic and the Gates Foundation, 36 percent of participants rated on

three or more metrics reported that their evaluations were “extremely” or “very” helpful, compared to 25 percent of those rated on fewer than three metrics.³⁰

Impact

The new, more comprehensive teacher measurement systems have put in motion half a dozen important improvements in public education.

Prioritizing Classrooms

The reform movement has made instructional quality a much higher priority in schools than it has been, prompting sustained discussions among teachers and administrators about effective teaching and how it is manifested in classrooms.

Under the new, more formal measurement systems, “teachers and principals have been forced to create a common language of instruction, something that didn’t exist in many schools,” says

No longer are teachers left to decipher what’s expected of them and what effective work looks like.

Charlotte Danielson, whose consulting company, The Danielson Group, employs 35 consultants to help states and school districts implement her Framework for Teaching.³¹

Danielson’s analysis is widely shared. “Teachers say, ‘I haven’t had this level of conversation about instruction with principals in 25 years in the classroom,’” says Tysza Gandha, a human capital expert formerly at the Southern Regional Education Board (SREB), and the author of two studies on the evolution of evaluation reforms in southern and mid-Atlantic states.³²

No longer are teachers left to decipher what’s expected of them and what effective work looks like. “There’s a clearer understanding of what

standards of performance are” under the new evaluation systems, says Heather Peske, associate commissioner for educator policy in Massachusetts, where a new evaluation model stresses teacher-improvement projects.³³

The press for stronger teacher evaluations has forced many school leaders to prioritize their classrooms over bus schedules and other daily demands, another promising step. “It has truly shifted the focus in schools from operations to instruction,” says Noah Bookman, who led the creation of the new Los Angeles evaluation system and is now director of accountability for a consortium of urban California school districts.³⁴

“Classroom doors are opening; it’s a hugely important change in the profession,” says educational consultant Joanne Weiss, who was the first director of the Department of Education’s Race to the Top program before serving as Secretary Duncan’s chief of staff.³⁵

Without evaluation reforms serving as a catalyst, it would have been difficult to break school leaders out of their traditional roles and routines, experts say. “There are a ridiculous number of legitimate things for principals to focus on, so getting them to focus on instruction is really hard,” says Heather Kirkpatrick, chief people officer at Aspire Public Schools, a charter management organization that is revamping its evaluation system with funding from the Gates Foundation.³⁶

Removing Low Performers

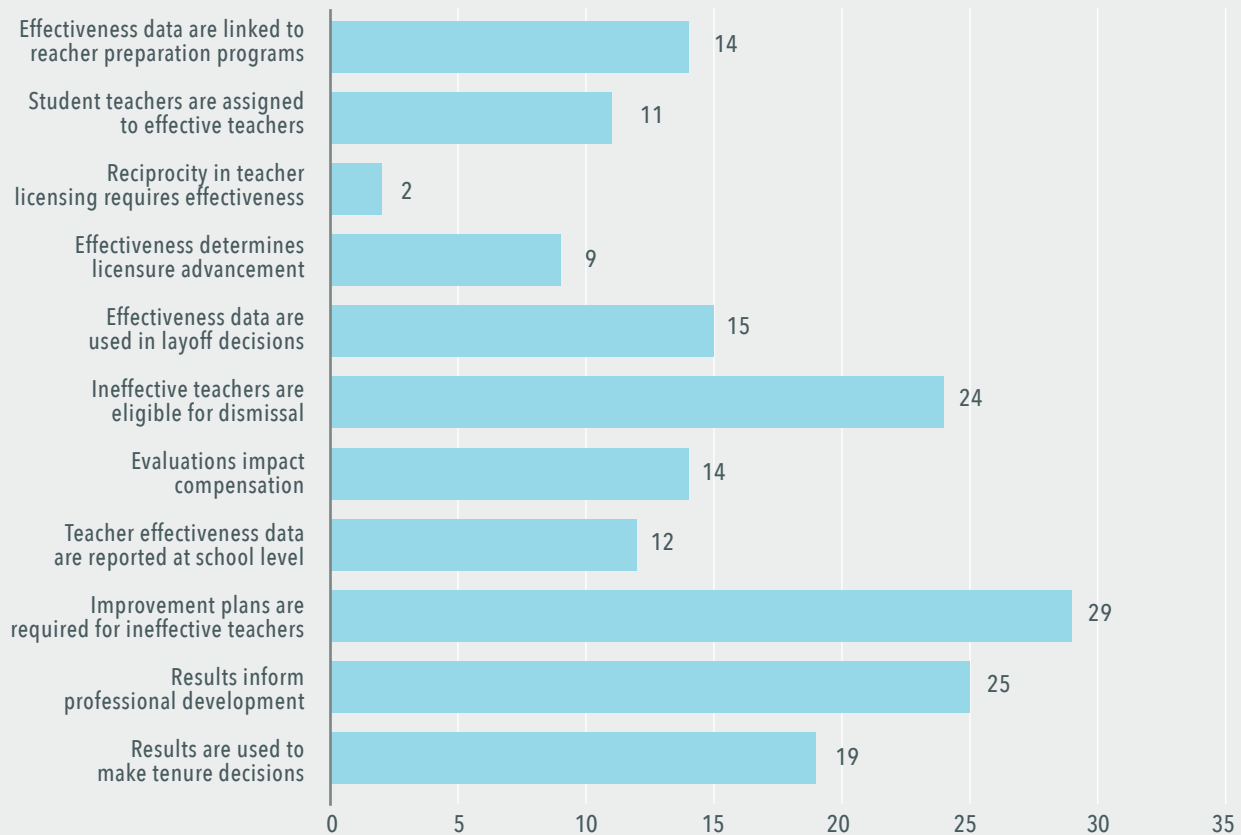
Early evaluation reformers prioritized removing bad apples from the nation’s classrooms, bolstered in their beliefs by new studies suggesting substantial differences in teachers’ impact on student test scores and by economic models showing that removing public education’s lowest-performing teachers would lift student achievement significantly. “The bottom end of the teaching force is harming students,” Stanford’s Eric Hanushek charged. “Allowing ineffective teachers to remain in the classroom is dragging down the nation.”³⁷

Recently, Matthew Kraft of Brown University and Allison Gilmour of Vanderbilt studied teacher ratings in roughly half of the nearly four dozen states with new evaluation systems and reported that a median of only 2.7 percent of teachers have received “unsatisfactory” ratings, even though principals they surveyed in one large urban school system suggested there were many more low-performing teachers than that in their schools.³⁸

Some commentators have pointed to the study and others like it as evidence that the evaluation reform movement hasn’t made much of a difference. But there are other ways to interpret the picture Kraft and Gilmour paint.

While there’s certainly plenty of work to do to improve the quality of the new evaluation systems—improvements that are likely to bring the differences in teacher performance into sharper focus—the proportion of unsatisfactory ratings the researchers found is about three times what it was before the introduction of the new grading systems. The New Teacher Project (now TNTP) studied Chicago’s teacher ratings between 2003 and 2006 and found that 87 percent of the city’s 600 public schools—including 69 the city had declared “educationally bankrupt”—didn’t issue a single “unsatisfactory” rating in those years.³⁹

Uses of New Teacher Evaluation Data



SOURCE: Doherty, Kathryn M., and Sandi Jacobs. *State of the States 2015: Evaluating Teaching, Leading, and Learning*. (Washington, DC: National Council on Teacher Quality, November 2015). <http://www.nctq.org/dmsView/StateofStates2015>.

Kraft and Gilmour concluded that the results represent “a meaningful increase” in the identification of underperformers since the pre-reform era. And in places with the best new evaluation systems, the numbers are substantially higher, with significant numbers of weak teachers being fired for the first time in public education’s recent history.

Take the District of Columbia, where teaching standards are clear and high and teachers are rated multiple times by both building administrators and outside observers, and where observation scores are combined with student-achievement results and other measures. Last year, only 80 percent of the District’s teachers were rated “effective” or “highly effective” under Washington’s comprehensive, seven-year-old rating system. Another 17 percent were placed on two levels of probation. Three percent were fired.⁴⁰

Other districts now releasing low-performing teachers regularly include Nashville, Memphis, Houston, and New Haven.⁴¹

There is also evidence that the threat of dismissal is encouraging weak teachers to leave the teaching profession voluntarily. Researchers Thomas Dee of Stanford and James Wyckoff of the University of Virginia recently reported that the departure of low-rated teachers under the District of Columbia’s

The sense of professional isolation, of being imprisoned in their classrooms, that public school teachers have long lamented is giving way in some schools to a focus on helping teachers improve their craft.

new evaluation system, combined with a push to recruit strong replacements, “substantially improves teaching quality and student achievement in [Washington’s] high-poverty schools.” The annual

attrition level of underperformers—those rated “ineffective” or “minimally effective”—was 46 percent, more than three times the departure rate among high-performers.⁴²

Of course, it only helps to remove bad teachers if good teachers replace them. Dee and Wyckoff found that Washington students with replacement teachers learned the equivalent of between a third and two-thirds of a year of additional study in math, and nearly as much in reading. But Washington is a magnet for talented millennials, giving its school district a recruiting advantage others may not have.

Beyond Bad Apples

If the early leaders of the reform movement were eager to get tough on bad teachers and reward good ones (the Obama administration’s Race to the Top competition required states to use new evaluation results in decisions on “compensation, promotion, retention, tenure, certification, and dismissal”), the evaluation reforms now working their way into public schools are changing the professional dynamic of public school teaching in another way. The emergence of shared instructional standards, the increased presence of principals in classrooms, and the evaluator-teacher conversations that are part of many of the new evaluation systems are creating more professional working environments for teachers.

The sense of professional isolation, of being imprisoned in their classrooms, that public school teachers have long lamented is giving way in some schools to a focus on helping teachers improve their craft. This is a shift teachers say they value greatly, and that may, if it spreads and deepens, be a significant consequence of today’s reforms.

Indeed, there’s a growing sense among reformers that the new generation of teacher evaluations should help teachers improve their craft instead of merely informing employment decisions, that strengthening the profession requires upgrading teachers’ performance instead of merely removing bad apples.

The Tennessee Department of Education has partnered with Brown University researchers John Tyler and John Papay to create an algorithm that matches teachers who do well on components of the state's new teacher-evaluation systems with colleagues who struggle. In an experiment, teachers in schools where the peer partnerships were introduced were more supportive of tougher teacher evaluations than their colleagues in schools without the program, and the partnership schools turned in higher student test scores in subsequent years. The program now reaches half of Tennessee's schools.⁴³

The District of Columbia has augmented its early commitment to greater teacher accountability with an ambitious new Teacher Data and Professional Development (TDPD) initiative that combines information on students, teachers, teaching standards, teaching strategies, and curricula within a single digital platform to create personalized, computer-generated professional-development plans for teachers based on their evaluation results. As part of the initiative, the school district's curriculum division commissioned a video company that had done work for the Discovery Channel and National Geographic to capture the city's best teachers demonstrating the district's nine teaching standards at every grade level in every subject—a project dubbed Reality PD.

New companies like BloomBoard and TeachBoost have begun drawing on the results of the new evaluation systems to provide teachers with personalized “playlists” of model lessons, readings, and other improvement materials based on their evaluation results. San Francisco-based Smarter Cookie enables teachers to upload videos of themselves teaching lessons and have them critiqued by trained coaches.

Even the Bill & Melinda Gates Foundation, an early advocate of tougher accountability in the teaching profession, is increasingly investing in the improvement of teacher practice, including working with 300 master teachers to build a digital archive of 15,000 exemplary lessons.⁴⁴

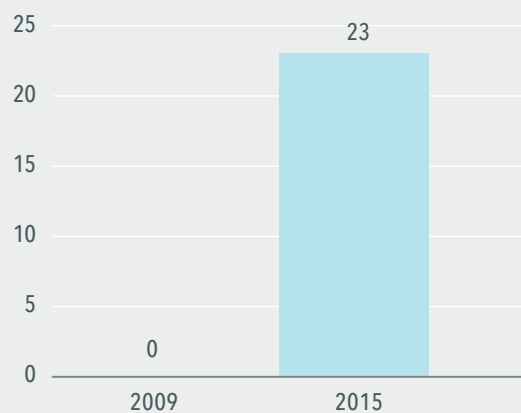
Although the new strategies haven't been extensively studied yet, they represent an encouraging new avenue to strengthen teachers' grasp of their subjects and of the best teaching strategies. In the past, public education has spent billions of dollars annually trying to improve teachers through what has been mostly a patchwork of widely disparaged workshops frequently having little to do with teachers' individual needs.

Smarter Employment Decisions

New information flowing from the improved evaluation designs is helping education leaders make smarter staffing decisions. Nearly two dozen states now require their school districts to weigh teacher performance when making tenure decisions, the National Council on Teacher Quality reports, a considerable shift from the pre-reform era when not a single state required its districts to consider any factors beyond years of experience.⁴⁵

Teacher layoffs have also traditionally been done strictly on the basis of seniority in public

States Requiring Teacher Performance as a Factor in Tenure Decisions



SOURCE: Doherty, Kathryn M., and Sandi Jacobs. *State of the States 2015: Evaluating Teaching, Leading, and Learning*. (Washington, DC: National Council on Teacher Quality, November 2015). <http://www.nctq.org/dmsView/StateofStates2015>.

education. Now, reductions-in-force in the District of Columbia and elsewhere prioritize teachers' evaluation results, helping to keep top performers in classrooms. Washington is also working to retain its best teachers by using evaluation results to reward them financially with bonuses and higher salaries.

Officials there and in Minneapolis are also using new evaluation data to identify the teacher training institutions delivering the strongest candidates—and targeting their teacher recruitment to those campuses.

A Foundation for New Teacher Roles and Responsibilities

The best of the new evaluation systems are creating a foundation for new, performance-based teacher responsibilities that reformers have long believed would make teaching a more attractive profession. With a dependable mechanism for identifying deserving teachers in place for the first time, New Haven, Connecticut, is among a growing number of districts where highly rated teachers have been tapped to serve as peer evaluators, mentors, or lead teachers—new roles that give teachers more compensation and higher status.⁴⁶

The District of Columbia is among the school districts that have launched formal teacher career ladders linked to new evaluation systems. The District's teachers are grouped into five categories based on experience and performance—"teacher," "established," "advanced," "distinguished," and "expert"—with the expectation that they will take on leadership responsibilities as they move up, yet another break from the traditional division of labor and management in public education. Tennessee draws from the ranks of its highest-rated teachers to staff a network of Core Coaches, who are leading the introduction of the Common Core State Standards to Tennessee's classrooms.⁴⁷

Helping the Bottom Line

Finally, early evidence suggests that evaluation reforms are improving public education's bottom line.

Beyond the studies linking new evaluation systems to higher student achievement in Tennessee and the District of Columbia, researchers John Tyler of Brown and Eric Taylor of Harvard found that Cincinnati's comprehensive evaluation system predicted the performance of teachers' future students. Mid-career teachers, their research revealed, increased student achievement more in the years after their first evaluation under the city's then-new evaluation system than they did in the years prior to their initial evaluations.⁴⁸

Challenges

The pace and scope of change in teacher evaluation have been dramatic in recent years. Add to this the complexity of the reforms involved, and it's no wonder a host of methodological and morale challenges has arisen that must be addressed if the reforms are to achieve their full potential to strengthen instruction, make teaching more attractive work, and raise student achievement.

The Obama administration's demands for a rapid redesign of evaluation systems under its Race to the Top funding incentives and its No Child Left Behind waivers, in particular, have left many state and local policymakers scrambling to create dependable evaluation systems that teachers will find credible. According to Education Counselor's Jess Wood, who has worked with many states on their new systems: "The federal government required lightning-fast responses from people who lacked experience and had few models to work from, resulting in a lot of design flaws."⁴⁹

Student Performance Measures

The most problematic—and controversial—aspect of the new educator evaluation systems has been the creation of student performance measures. While reformers' insistence on measurement is understandable—student achievement is, after all, what matters most in education—trying to do so systematically for the first time in public education's history has been challenging, to say the least.

Policymakers have relied on so-called value-added measures and student growth percentiles to gauge the performance of teachers in those grades and subjects for which students take standardized tests. Both strategies compare students' results with their own prior performance or the prior performance of students with similar academic and demographic profiles. The goal of value-added measures is to isolate teachers' contributions to their students' success.

There is a consensus among measurement experts that value-added calculations (and student growth percentiles) do effectively identify high- and low-performing teachers.⁵⁰ For example, five leading researchers hired to study value-added measurements by the Carnegie Foundation for the Advancement of Teaching concluded in a 2015 summary report that “value-added measures meaningfully distinguish between teachers whose future students will consistently perform well and teachers whose students will not.” Leading researchers also have found that students of teachers with higher value-added ratings enjoy greater success in their subsequent schooling and beyond.⁵¹

But weaknesses in the measures have led to mistaken ratings, undermining teachers' confidence in the measures. “You absolutely can use value-added measures in responsible ways,” says Scott Marion, executive director of the New Hampshire-based Center on Assessment. “But policymaking has run far ahead of practice.”⁵²

The basic problem is that value-added ratings are for a number of technical reasons inherently imprecise, or “noisy,” in researcher parlance. As a result, they misrepresent the “true” performance of many teachers.

There are often big year-to-year shifts in individual teachers' scores, resulting in high-performing teachers earning low ratings, and low-performers earning high marks. Researchers Peter Schochet and Hanley Chiang of Mathematica have found that when teachers' ratings are based on three years' worth of student scores, value-added ratings

should be expected to misidentify a quarter of teachers judged “highly effective” and an equal number judged “ineffective.”⁵³ Closer to a third would be misclassified if rated on the basis of a single year's scores.

While value-added scores reliably spot teachers at the top and bottom of the performance range, they aren't particularly helpful in identifying the differences among the nation's many mid-range teachers. And value-added ratings tend to favor teachers working with more-advantaged students.

Even when policymakers control for students' socioeconomic background, value-added calculations can't fully account for factors beyond a teacher's control, like the quality of her principal, the performance of other teachers, and school safety. Trying to level the playing field among teachers in different schools by taking students' backgrounds and prior performance into account in calculating value-added scores has proven tougher than expected.

Value-added systems are not designed to produce information geared toward improvement in instructional practice.

As a result, top teachers have an incentive to avoid working in struggling schools, and the reform goals of improving instruction in high-need schools and using student achievement to evaluate teachers are set in conflict with each other.

Policymakers also have struggled with how to apportion students for value-added ratings under the growing number of flexible teaching arrangements in public education, where team-teaching, learning that blends face-to-face instruction with technology, and other innovations, are expanding. This is a particular problem for special education teachers, who serve approximately 13 percent of the nation's students, but who frequently share instruction with regular classroom teachers.⁵⁴

Finally, student test scores are lagging indicators of teacher performance. Because students traditionally take standardized tests in the spring, teachers aren't able to act on the results until the following school year—that is, if they're able to act on them at all. Value-added systems are not designed to produce information geared toward improvement in instructional practice.

The Other 70 Percent

Complicating things further is the fact that only about 30 percent of the nation's teachers instruct in subjects or at grade levels for which students take standardized tests.⁵⁵ Producing dependable student achievement ratings for the other 70 percent has been very difficult.

Some school districts have augmented statewide standardized testing with new local tests. Miami-Dade, the nation's fourth largest district, wrote new tests for more than 1,000 courses for its new teacher evaluation system, at a cost of about \$3 million.⁵⁶

This new testing for teacher evaluations has helped fuel a vocal anti-testing movement in the country that has damaged both teacher evaluation reform and the introduction of the rigorous Common Core State Standards. In what has proven to be a problematic decision, former US Secretary of Education Arne Duncan pressed states to emphasize student test scores in new teacher evaluation systems at the same time that he introduced the Common Core standards and new testing regimes linked to the standards. In early 2016, new US Secretary of Education John King called on policymakers to halt the expansion of testing for teacher evaluation.

Some states and school districts have pursued a different strategy, calculating the school-wide growth in student achievement on existing standardized tests and assigning the results to teachers in non-tested grades and subjects. But this school-wide strategy has proven equally controversial, prompting embarrassing headlines about math results being used to rate gym

teachers, and calculus teachers being judged on reading scores. Sandi Jacobs of the National Council on Teacher Quality calls the approach “indefensible.”⁵⁷

However, the most common student achievement measure for teachers working outside of tested grades and subjects has been progress toward student learning objectives, or SLOs—grade-level and subject-specific outcomes that teachers select with their principals and typically measure through locally created, non-standardized assessments such as essays or projects. SLOs give teachers more ownership of the student achievement dimension of their evaluations, and surveyed teachers report liking them more than statewide standardized test scores because they're more relevant to their day-to-day teaching. But they, too, are flawed measures of teacher performance.⁵⁸

One problem is that many teachers and principals lack the “assessment literacy” required to create tests of sufficiently high quality to measure teacher performance confidently under SLO regimes. “Principals aren't psychometricians,” says Luke Kohlmoos, director of Tennessee's teacher evaluation system from 2012 to 2014, who calls SLOs “the least stable, least predictive [measure] of future [student] performance.”⁵⁹

SLOs are also time-consuming and therefore expensive to create. And because they're typically non-standardized assessments, they make dependable teacher-to-teacher comparisons next to impossible.⁶⁰ “They require a lot of effort for very, very little information,” Kohlmoos said. In the District of Columbia, SLOs count for just 15 percent of teachers' ratings because, in the words of one District official, “they're just not tight enough.”

But student achievement has a valuable role to play in teacher evaluations, experts say, despite the limitations of the various strategies for measuring teachers' contributions to student success. The “Duncan's Gambit” section below describes emerging strategies to address the measurements' weaknesses.

Inside Classrooms

While the new evaluation mandates have spurred fresh conversations about effective teaching in many schools and lent a renewed sense of purpose to observing teachers’ work, building the much more robust classroom observation systems necessary to make evaluations truly meaningful has been challenging. “[M]any states are facing technical, logistical, and human capital challenges in implementing and sustaining robust, reliable, and fair observation systems,” writes the

Southern Regional Education Board (SREB) in *Toward Trustworthy and Transformative Classroom Observations*, a 2015 study.⁶¹

Scrutinizing teachers at work with students is obviously critical to learning their strengths and weaknesses and to helping them improve their practice. But in much of public education, doing so represents a major cultural shift. Traditionally, principals functioned primarily as building managers rather than instructional leaders, and teachers were considered autonomous actors within their

State Authority Over Teacher Evaluations

	Single Statewide System	State Model for Districts with Possible Opt Out	State Criteria for District-Designed Evaluation Systems		Single Statewide System	State Model for Districts with Possible Opt Out	State Criteria for District-Designed Evaluation Systems
Alabama	●			Montana			●
Alaska			●	Nebraska			●
Arizona			●	Nevada			●
Arkansas			●	New Hampshire			●
California			●	New Jersey			●
Colorado			●	New Mexico			●
Connecticut		●		New York			●
Delaware	●			North Carolina		●	
DC			●	North Dakota			●
Florida			●	Ohio		●	
Georgia	●			Oklahoma	●		
Hawaii	●			Oregon			●
Idaho			●	Pennsylvania		●	
Illinois			●	Rhode Island		●	
Indiana			●	South Carolina	●		
Iowa			●	South Dakota			●
Kansas		●		Tennessee		●	
Kentucky		●		Texas		●	
Louisiana		●		Utah			●
Maine			●	Vermont			●
Maryland			●	Virginia			●
Massachusetts			●	Washington	●		
Michigan			●	West Virginia	●		
Minnesota			●	Wisconsin		●	
Mississippi	●			Wyoming			●
Missouri		●					
				TOTAL	9	12	30

SOURCE: Doherty, Kathryn M., and Sandi Jacobs. *State of the States 2015: Evaluating Teaching, Leading, and Learning*. (Washington, DC: National Council on Teacher Quality, November 2015). <http://www.nctq.org/dmsView/StateofStates2015>.

classrooms. Many principals didn't have a say in which teachers taught in their buildings and weren't held accountable for their teachers' performance, while teachers were compensated strictly on the basis of seniority and college credentials. There was little reason to take observations seriously, and few people did.

As a result, when the new systems came cascading down on the nation's schools, many administrators weren't ready.

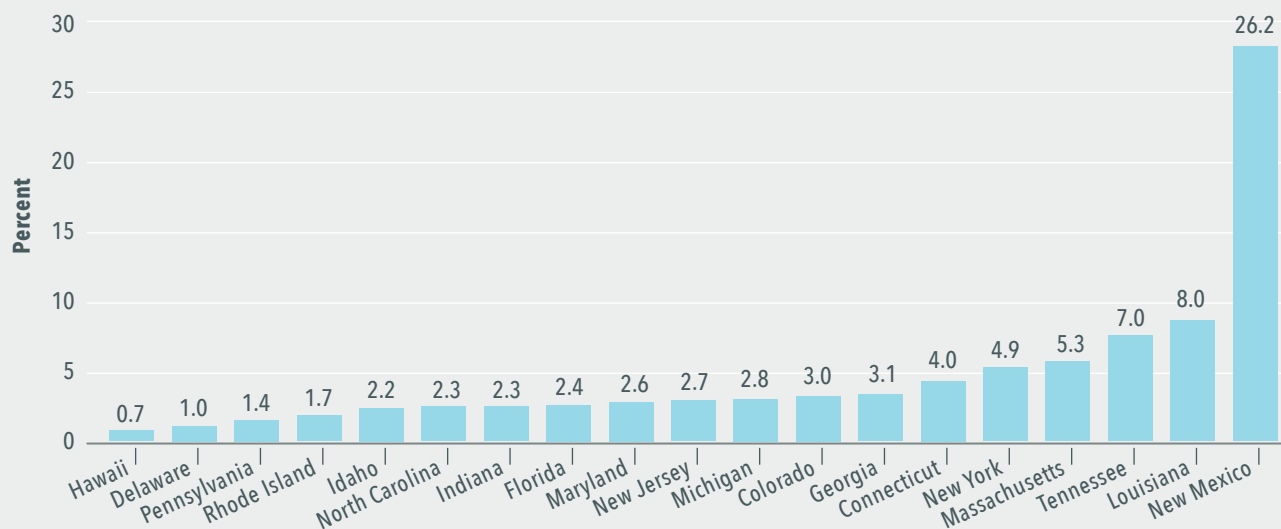
Unable to train sufficiently under tight federal timelines, many administrators didn't learn reliable methods of rating teachers.⁶² And even as training expanded, education leaders discovered that a single preparation session wasn't enough to ensure accuracy by individual raters or consistency across raters. "It takes constant vigilance to ensure that a three is a three is a three," says Heather Kirkpatrick, chief people officer at Aspire Public Schools, a charter schools network that certifies the ability of its teacher-evaluators—who are both school administrators and peer teachers—to rate teachers accurately before sending them into classrooms.⁶³

Evaluators have found rating rubrics overly detailed and observations too time-consuming, leaving many administrators overwhelmed.⁶⁴ Seventy-five percent of the nation's principals reported in 2012 that their jobs had become too complex.⁶⁵

And even when principals do sense they have a handle on the new observation obligations, they're often not on the same page as their teachers, both because many schools lack a strong tradition (or any tradition) of discussion about teaching practices, and because teachers and principals frequently received different training on the new evaluation standards, or "rubrics." Fully 97 percent of principals in a 2014 survey of Indiana educators by Indiana University's Center on Education and Lifelong Learning said they had a strong grasp of their evaluation rubrics, compared to only 49 percent of teachers.⁶⁶

High principal turnover in many urban school systems exacerbates the problem,⁶⁷ as does the finding by researchers that observation ratings, like value-added scores, tend to be lower for teachers serving low-income and minority students, and

Percentages of Teachers in 19 States Rated Below Proficient, 2014-15



SOURCE: Kraft, Matthew A., and Allison F. Gilmour. "Revisiting the Widget Effect: Teacher Evaluation Reforms and the Distribution of Teacher Effectiveness." Brown University Working Paper, February 2016.

higher for teachers of advanced classes—another incentive for top teachers to abandon struggling schools.⁶⁸

Researchers have also revealed that principals tend to give their teachers low ratings reluctantly and routinely rate them more generously than outside observers, a problem called “building bias.”⁶⁹ “Principals are the same people who gave high scores under the old evaluation systems,” says Tom Kane, a Harvard education professor and a leader of the comprehensive MET study of new teacher evaluation strategies funded by the Bill & Melinda Gates Foundation. “It’s human nature; it’s very hard to give critical feedback to people you work with.”⁷⁰

These factors have combined to produce the high percentages of “proficient” ratings that Kraft and Gilmour found in their study. And a number of states have reported teachers’ classroom observation ratings outpacing their value-added scores—an argument for multiple-measures evaluation systems.⁷¹

Flawed Feedback

Regardless of how many passing grades they award, evaluators have struggled to use observation results to help teachers improve their instruction—the aspect of evaluation that teachers say they most value.

Here, too, principals and other evaluators have had to play catch-up, given their limited involvement in instruction in the past and the fact that the evaluation reform movement unfolded rapidly and was initially focused on holding teachers accountable for their performance rather than on improving their practice. The feedback that teachers received typically summarized their rating results and wasn’t connected to improvement opportunities.⁷² Some school districts even lacked the infrastructure to store observation results, much less a capacity to act on the information to strengthen instruction.⁷³

The disconnect between what teachers want from teacher evaluations and what many are getting is

captured in the 2014 survey by Indiana University’s Center on Education and Lifelong Learning. Researchers found that 95 percent of principals believed the post-observation sessions they held with teachers were “constructive,” compared to 53 percent of teachers. And only 30 percent of teachers agreed that their school districts’ evaluation systems drive professional development, compared to 75 percent of principals.⁷⁴

Part of the problem is that the human capital and instructional functions of many school districts aren’t well integrated, making it tougher to incorporate new teacher evaluation data into instructional systems. And many school leaders receive “little or no support from their [central offices] in identifying teacher [improvement] needs or for ensuring that teachers’ professional development opportunities match their needs,” Mathematica reported in 2014.

Addressing these challenges requires a substantial expansion of the instructional infrastructure of many school districts. With the teacher evaluation reform movement serving as a catalyst, this important work is underway in many places. (Given its centrality to the education enterprise, such an infrastructure should have been in place long ago.) But expansion is expensive. Increased observations, training, and data systems, not to mention the draw on school leaders’ time, are stretching the resources of many school districts. Fortunately, solutions are emerging.⁷⁵

Teacher Morale

Teachers have told a wide range of researchers that they value the shared language and more frequent conversations about effective teaching that the new evaluation systems have engendered in many schools. They particularly value evaluators’ guidance on improving their performance—support that was rare in public education in the past and that teachers say has made their work more attractive. Stephanie Reinhorn of Harvard’s Project on the Next Generation of Teachers found in a study of new teacher evaluations in high-poverty urban schools that “teachers craved opportunities

to receive detailed and useful feedback on their teaching practice as well as complementary support for improvement.”⁷⁶

But several elements of the evaluation reform movement have alarmed and angered many teachers, turning them against reform. These include the fast pace of mandated changes; the early focus on removing bad teachers; the heavy reliance on principals who turned out to be

Teachers have told a wide range of researchers that they value the shared language and more frequent conversations about effective teaching that the new evaluation systems have engendered in many schools.

underprepared; the new, untested use of student achievement results; the simultaneous rollout of the Common Core curriculum and new tests; poor-quality feedback in many jurisdictions; and a lack of improvement resources.

Inflammatory press commentary—notably a 2008 *Time* cover photograph of reformer Michelle Rhee standing in a classroom with a broom in her hand and a hard look on her face, and a 2010 *Newsweek* cover story declaring, “The Key to Saving American Education: We Must Fire Bad Teachers”—compounded the problem, causing many teachers to interpret evaluation reform as a threat rather than an opportunity to improve.

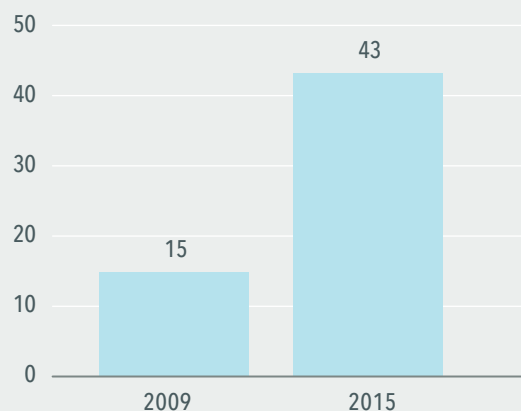
The 2014 survey by Indiana University captured the feelings of many teachers. While 84 percent of Indiana’s principals told the researchers they expected new state-mandated evaluation systems to strengthen teaching and learning, only 15 percent of the state’s teachers agreed.⁷⁷ Similarly, only 39 percent of teachers told the 2012 MetLife Survey of the American Teacher that they were very satisfied with their jobs, down from 62 percent in 2008, before the onset of the evaluation reforms.⁷⁸

Part of the problem was that in many places teachers played minor roles (or none at all) in the development of new evaluation systems. They found it difficult to trust a system on which their jobs depended but over which they had no control. In states where teachers had a seat at the table—Massachusetts, Kentucky, and in such school districts as New Haven and Cincinnati—teachers were more sympathetic to reform.

The new evaluation systems also revealed that many teachers didn’t trust their principals to rate them fairly. But more than any other objection, teachers hated being judged by their students’ test scores—even though the scores were predictive of teachers’ future performance. A national Gallup poll found that nearly nine out of ten teachers believed that tying test scores to teacher evaluations was unfair.⁷⁹

The new value-added systems were complicated—“the average person can’t understand it,” a recent Phi Beta Kappa graduate from a top university told me—and teachers felt them to be

States Requiring Inclusion of Student Achievement Data in Teacher Evaluations



SOURCE: Doherty, Kathryn M., and Sandi Jacobs. *State of the States 2015: Evaluating Teaching, Leading, and Learning*. (Washington, DC: National Council on Teacher Quality, November 2015). <http://www.nctq.org/dmsView/StateofStates2015>.

capricious, holding teachers responsible for student background and other factors outside of their control. (The Indiana University survey reported that 81 percent of school superintendents and 63 percent of principals “strongly agree” that teacher effectiveness affects student achievement, versus 25 percent of teachers.⁸⁰) Teachers teaching non-tested grades and subjects deeply resented being rated on the basis of other teachers’ test results, as happened in many places.

The anxiety teachers felt about the introduction of value-added ratings was compounded by the simultaneous rollout of demanding new national testing systems tied to the Common Core standards. The overlapping reforms meant that teachers would be evaluated on tougher tests pegged to a new, more demanding curriculum—leaving teachers in what many rightly argued was a deeply unfair situation. In the words of Heather Kirkpatrick of Aspire Public Schools: “We piloted our new evaluation system using the California state test [to calculate value-added scores], then all of a sudden we’re using new tests based on the Common Core. It has been very bad for morale.”⁸¹

The dual reforms also fueled the anti-Common Core and anti-testing movements that have gathered momentum recently. Such was the backlash that then-Secretary of Education Duncan was forced in late 2014 to declare a moratorium on the use of value-added scores in teacher evaluations.

Organizations representing other educators piled on. The National Association of Secondary School Principals, for example, in 2015 issued a statement declaring that value-added scores shouldn’t be used to make personnel decisions about teachers.⁸² Despite this, many of the so-called accountability hawks in the evaluation movement remained strongly committed to removing bad teachers and communicating the importance of classroom performance—so strongly committed, in fact, that they largely ignored the impact of value-added measures on teacher morale.

Teacher Unions

If the evaluation reforms have been unsettling for many teachers, they present both a tremendous challenge and an opportunity to teacher unions, the single most influential voice in public education.

Performance-based staffing and compensation systems represent a sharp break from traditional union-backed policies that differentiated between teachers largely on the basis of their credentials rather than by the quality of their work. At the same time, the new, more comprehensive evaluations signal positive change: a path to improving teacher performance, more professional working conditions, and increased professional opportunities.

Local unions in New Haven, Pittsburgh, and Hillsborough County, Florida, as well as state-level teacher associations in Massachusetts and Kentucky, have been partners in teacher-evaluation reform out of a belief that more meaningful evaluations could strengthen the profession and provide a foundation for improving teachers’ professional lives, transforming public schools into “learning communities” for teachers as well as students.

Steve Cantrell, who played a key role in the Bill & Melinda Gates Foundation’s three-year Measures of Effective Teaching project, says that “MET couldn’t have happened without tremendous union support” in the study’s seven school districts.⁸³

Yet while some unions have embraced reform, many more have not. The governing body of the National Education Association, the nation’s largest teacher union, declared in a 2014 resolution that “standardized tests...may not be used to support any employment action against a teacher.”⁸⁴ “The association...believes,” it wrote in another resolution, “that...compensation based on an evaluation of an education employee’s performance” is “inappropriate” and that “any additional compensation beyond a single salary schedule must not be based on education employee evaluation, student performance, or attendance.”⁸⁵

Randi Weingarten, the president of the American Federation of Teachers, which represents many teachers in urban school districts in the Northeast and Midwest, at first endorsed the calls for evaluation reform as a way of strengthening the teaching profession. “A strong teacher development and evaluation system is crucial to improving teaching,” she declared in a 2010 speech at the National Press Club in Washington, entitled “A New Path Forward: Four Approaches to Quality Teaching and Better Results.”⁸⁶ Such systems would include “student test scores based on valid and reliable assessments that show students’ real growth while in the teacher’s classroom,” she said, and “would inform tenure, employment decisions, and due process proceedings.”

But Weingarten and her organization reversed course in the face of a growing backlash against reform among the AFT’s regional leaders and mounting opposition within its rank and file to rating teachers with test scores.

Because student achievement was, for teachers, the most controversial component of teacher evaluation, Weingarten, a former lawyer and New York City labor leader, shrewdly sought to discredit teacher evaluation reform as a whole by treating student achievement as if it were the *only* component of new evaluation systems. “Test-based teacher evaluation has not worked,” she declared. In 2014, the AFT launched an anti-evaluation public relations campaign with the slogan “VAM is a sham.”⁸⁷

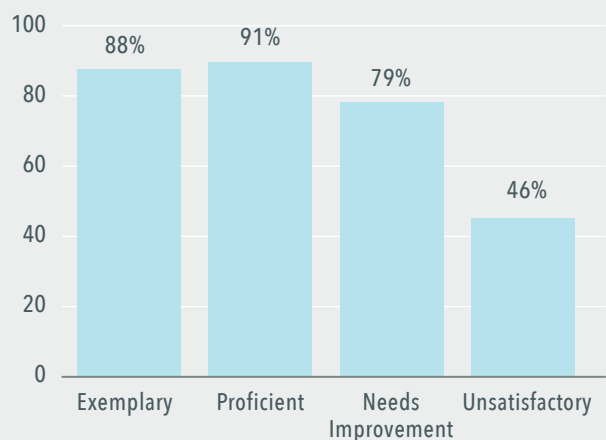
By the National Education Association’s own calculations, the union sent 255,000 emails to Capitol Hill, made 23,500 phone calls, and had 2,300 face-to-face meetings with lawmakers and their aides last year to ensure that the Obama teacher-evaluation-reform incentives and other accountability provisions didn’t make their way into the new federal Every Student Succeeds Act. It also spent \$500,000 on advocacy advertising in key Senate congressional districts promoting its anti-evaluation agenda—even though many of the new evaluation blueprints have paved the way for new teacher roles and responsibilities that the union’s own polling shows their members want. The smaller

AFT contributed some 125,000 phone calls and ran ads in *The New York Times* attacking Obama’s reforms, joining the NEA in an alliance with Tea Party advocates and congressional republicans to strip the Obama reforms from the new federal law.⁸⁸

Teacher unions have been equally aggressive in many states and school districts. In early 2015, for example, the New York State United Teachers, the AFT’s New York affiliate, picketed the state capitol in Albany as part of a (partially successful) lobbying campaign against Governor Andrew Cuomo’s effort to strengthen the state’s teacher evaluation system. In the wake of the battle, NYSUT President Karen Magee encouraged parents to “opt-out” of New York standardized tests “to try to subvert the [state’s teacher evaluation] rating system.”⁸⁹

And unions have taken their fight against evaluations to the judiciary in Florida and half a dozen other states, though they haven’t been successful in overturning any of the new systems yet.⁹⁰ In a recent example, a federal appeals court last rejected a union suit in Florida claiming that

Percentages of Massachusetts Public School Educators Teaching in 2014-15, by 2013-14 Evaluation Ratings



SOURCE: Brown, Catherine, Lisette Partelow, and Annette Konoske-Graf. Educator Evaluation: A Case Study of Massachusetts’ Approach. (Washington, DC: Center for American Progress, March 2016).

the use of student test scores violated teachers' 14th Amendment rights to due process and equal protection.⁹¹

Yet while the unions' resistance to the new evaluation systems suggests a simple calculation of risks outweighing benefits, the challenges presented by reform are in fact far more complex.

The unions' push to protect jobs is understandable—they have a long tradition of doing that. But establishing rigorous teaching standards, helping teachers reach those standards, and rewarding those who do—steps that the best new evaluation systems have introduced—are key to strengthening the teaching profession, and thus the long-term welfare of unions and their members.

Union leaders know this. “Professionalism comes from teachers owning the standards of the profession,” Marla Ucelli-Kashyap, assistant to the president for educational issues at the AFT, told me when we met in her office.⁹² And they know that their members want to replace the isolation of traditional public school classrooms with collegial working environments where the quality of their work matters.

“Teachers hate the blame and shame narrative [of the new evaluation systems], they hate value added,” says Rob Weil, also of the American Federation of Teachers. “But the polling numbers are completely different on the question of ‘do you want help improving your teaching’.”⁹³ Since the new evaluation systems supply the foundation for that help by indentifying teachers' strengths and weaknesses, there's seemingly a strong argument for teacher unions taking a more nuanced stance toward teacher evaluation reform.

Duncan's Gambit

Secretary of Education Duncan knew that most states and school systems lacked the ability to transform their teacher evaluation systems on the tight timelines required by his department's Race to the Top and NCLB waiver programs. The wisdom

of setting such aggressive expectations in the face of that reality was much debated among Duncan's senior staff.

But Duncan ultimately concluded that if state and local policymakers and practitioners “were not forced to grapple with the evaluation issue, they would continue to ignore it,” in the words of a participant in the department discussions. To Duncan, change in education required disrupting the status quo. If he could shake things up enough on the ground, he reasoned, people would “figure out how to put things back together again,” strengthening teacher evaluation in the process.

There's no doubt that the dramatic pace and scale of evaluation reform would have been unimaginable without Duncan's willingness to extend the federal government's reach into a core aspect of school operations—not to mention the powerful financial and regulatory incentives he created. But because of the speed, complexity, and reach of reform, the new teacher evaluation systems emerging in states and school districts are very much works in progress. Flawed, fractious, and incomplete, their return on investment is not yet fully visible.

Some states and districts are merely going through the motions of change, more compliant than committed to careful teacher evaluation and the opportunities it creates to improve staffing decisions and teacher performance. Many state departments of education and local school districts suffered budget cuts in the wake of the recent recession that have made it harder to respond to demands for reform.⁹⁴

However, the capacity of schools to conduct meaningful evaluation is increasing across a growing number of states and districts as solutions to the nascent evaluation systems' many implementation challenges emerge.

Simpler Rubrics

In an effort to help principals and other evaluators navigate teacher observations more efficiently and effectively, Charlotte Danielson and others are

creating simpler standards for measuring classroom performance, known as rubrics. Danielson says that the level of detail in her widely used Framework for Teaching “makes it cumbersome for everyday use,” and she is developing a new model organized around six “clusters” of effectiveness, down from the nearly two dozen criteria included in the original framework. The clusters are content knowledge, safe and effective learning environments, classroom management, student intellectual engagement, successful learning by all students, and teacher professionalism.⁹⁵

TNTP, the teacher reform and recruitment organization, has created an alternative to what it calls “overstuffed, clumsy [observation] rubrics” which focuses on four key questions: Are all students engaged in the work of the lesson from start to finish? Are all students working with essential content for their subject and grade? Are all students responsible for doing the thinking in this classroom? And do all students demonstrate that they are learning?⁹⁶

The District of Columbia, which in the 2009-10 school year introduced one of the nation’s first comprehensive evaluation systems, is paring down its observation criteria by two-thirds in 2016-17 in order to streamline evaluators’ work, says Michelle Hudacsko, the District’s teacher-evaluation director.⁹⁷ (Although the District is also adding content knowledge to its evaluation measures, to help teachers respond to the more demanding Common Core State Standards.)

To help teachers and administrators see eye to eye on the issue of instructional standards, Louisiana and Tennessee have built video libraries that include examples of effective teaching at each level of performance under each standard.⁹⁸

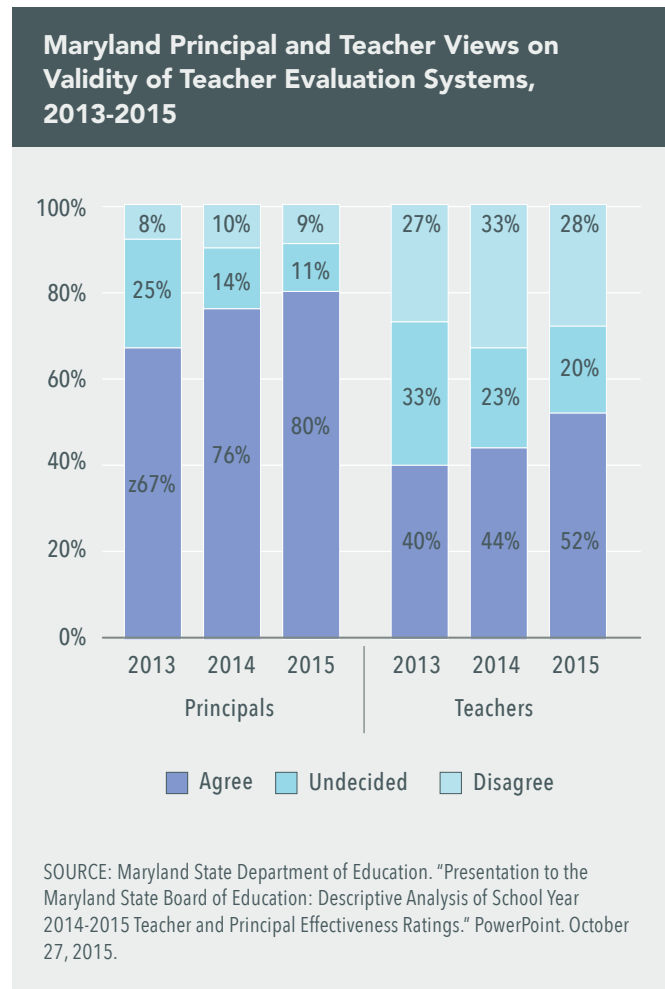
Tennessee has also developed a strategy to help principals and other raters produce consistent scores. The state’s department of education employs eight regional coaches to work with evaluators in schools with big discrepancies between teachers’ observation ratings and value-added scores—a signal that observers’ ratings are

out of kilter. To date, the program has improved the quality of observations in over 100 schools.⁹⁹

The District of Columbia, meanwhile, has a simple solution to the challenge of generating value-added scores for special-education teachers, many of whom work with students only part-time: The District excludes them from its value-added system.

Cutting Costs

Comprehensive teacher evaluation systems are more expensive than the superficial exercises they have replaced. The District of Columbia, for example, spent about \$1 million creating its ambitious IMPACT system, and in its first several years under the new system the school district spent roughly \$1,000 per employee to evaluate



4,000 classroom teachers and 2,300 aides, custodians, and other non-instructional school-based employees, or \$6.7 million of its \$758 million operating budget in 2011-12.¹⁰⁰

Massachusetts tapped into \$250 million in Race to the Top funding to provide its school districts with the guidance and infrastructure needed to build more robust evaluation systems—everything from model observation rubrics to student surveys and data systems to track the flow of evaluation results.

But the new systems represent expenditures in schools' core instructional work, precisely the sort of investment that policymakers should be making. In their 2011 study, John Tyler and Eric Taylor calculated that Cincinnati's comprehensive evaluation system pays substantial financial dividends in the form of increased student success.¹⁰¹

And strategies are emerging to reduce the cost of the new systems and increase their efficiency.

While policymakers initially required multiple classroom observations of all teachers, they have begun to differentiate these requirements based on teacher status. In Ohio, for example, districts adopting the state's model evaluation system can opt to evaluate top teachers bi-annually instead of annually, reducing the workload for observers and saving resources.¹⁰²

The Achievement First charter school network has taken a slightly different tack, reducing the number of formal evaluations for all teachers and replacing them with more informal “walkthroughs”—brief check-in visits focused on a single skill or behavior—out of a belief that they engender trust between teachers and observers.¹⁰³

The District of Columbia's public school system, which in recent years has lost thousands of students to charter schools, recently announced plans to eliminate its cadre of over three dozen master educators, the nationally recruited teaching experts paid by the District to rate teachers alongside school administrators. The District, which

had already reduced master-educator observations for teachers who routinely received proficient ratings, expects the move to lower the price of its IMPACT evaluation system “significantly,” says Hudacsko.¹⁰⁴ The step coincides with the winding down of a federal grant that funded IMPACT, and with the school district's decision to spend more resources on teacher professional development.

New Haven's decision to use external evaluators exclusively as a check on school principals' ratings of novices and low- and high-performers represents a middle ground in the effort to evaluate teachers more cost-effectively. It saves resources, but preserves the benefits of providing key groups of teachers with multiple perspectives on their work.

Improving Teacher Morale

States and school districts are also taking steps to make the new evaluation systems more amenable to teachers, with whom the success of evaluation reform ultimately rests.

Policymakers in growing numbers have reduced the use of student achievement results in calculating teachers' performance. Others have suspended their use until new, Common Core-based testing systems are fully introduced—something many teachers have demanded.

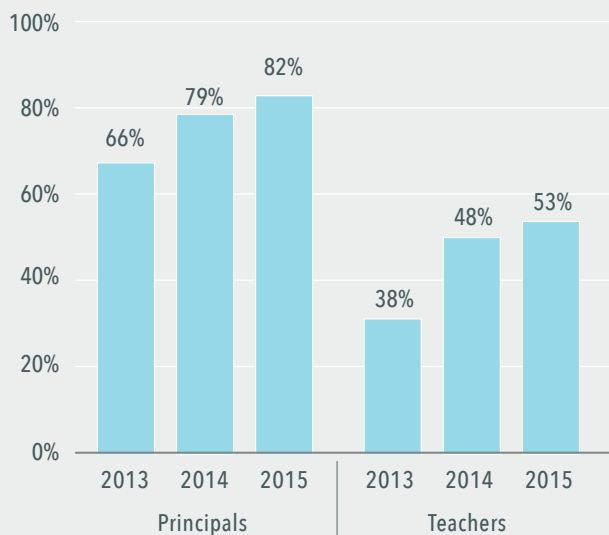
Even the District of Columbia, where student achievement was the foundation of Michelle Rhee's take-no-prisoners stance toward teachers when she launched IMPACT in 2009, reduced the weight of value-added scores in teacher ratings from 50 percent to 35 percent in the face of teacher pushback. And it eliminated whole-school value-added ratings altogether, to reduce the friction between teachers of tested and non-tested subjects and because whole-school ratings created disincentives for high-performing teachers to work in low-performing schools.

Others have sought to make student-achievement results more palatable to teachers by targeting the ways in which the information is used. One strategy, championed by Harvard's Tom Kane, involves

giving student achievement a substantial role in the evaluations of non-tenured teachers, since value-added measures predict future performance, while reducing the role of achievement results for tenured teachers, for whom improvement-oriented metrics are most important.¹⁰⁵ A second option involves using student-achievement metrics only as an initial screen to identify top and bottom performers, which the metrics do most reliably. Teachers at the extremes would then receive extensive in-classroom evaluation to determine the appropriateness of either remediation or rewards, but for the majority of teachers the metrics wouldn't be factored into the evaluations at all.¹⁰⁶ These ideas are explored in more detail in the Recommendations section below.

The evidence seems increasingly clear that providing teachers with multiple evaluations by multiple evaluators works better than relying exclusively on individual building administrators.

Percentage of Maryland Principals and Teachers Agreeing that State's New Evaluation System Improve Instruction, 2013-2015



SOURCE: Maryland State Department of Education. "Presentation to the Maryland State Board of Education: Descriptive Analysis of School Year 2014-2015 Teacher and Principal Effectiveness Ratings." PowerPoint. October 27, 2015.

Research reveals that multiple perspectives function as a check on building bias and yield more dependable ratings of teachers' work. Just as importantly, teachers say they value outside observers, especially subject-matter and grade-level experts, both because they don't always trust their principals to be fair and because they think they'll get more helpful feedback from instructional experts.

Other researchers have stressed the value of giving teachers a greater stake in the evolution of the new evaluation systems than was the case in many places early on. "You have to have teachers feel like they have a say, if you want buy-in," concludes University of Southern California researcher Julie Marsh from her study of the contentious implementation of the new evaluation system in Los Angeles.¹⁰⁷

There has been less controversy in states like Massachusetts, where teachers were involved in both drafting and implementation of the new evaluation system. The Revere Public Schools, for example, trained a cadre of teachers in the state's evaluation system and they serve as resources on the Massachusetts model for their colleagues. Says Heather Peske of the state's department of education: "We're having teachers be experts on implementation, rather than imposing reform on them."¹⁰⁸ In some school districts, including New Haven, Connecticut, teacher unions have co-authored new evaluation systems.¹⁰⁹

Tom Kane, meanwhile, has been exploring a different strategy to increase teacher buy-in. Over the past several years, he and Harvard colleagues have given two hundred teachers around the country video cameras to record their lessons. The teachers may select three or four lessons to share with their principals (for evaluation) and with outside master teachers (for informal feedback). Kane and colleagues found that teachers became more reflective and sympathetic concerning the process of evaluation under this "best foot forward" experiment. As an added benefit, principals gained the flexibility to score lessons remotely on their own schedules.¹¹⁰

Ultimately, teachers are more likely to back new evaluations if they perceive them as designed to help strengthen teaching, and a number of states and districts are making that aim an increasing focus of their evaluation systems.

To ensure the effectiveness of its evaluators, Tennessee requires them to accurately score videotaped lessons against teaching standards before they're allowed into classrooms. But the state also requires prospective evaluators to demonstrate their ability to lead effective post-observation conferences with teachers and give

Teachers say they value outside observers, especially subject-matter and grade-level experts, both because they don't always trust their principals to be fair and because they think they'll get more helpful feedback from instructional experts.

them well-targeted improvement plans.¹¹¹ The Uncommon Schools network of charter schools and the *New York City Department of Education* have developed video libraries to help principals and other observers deliver effective, rubric-based feedback to teachers.¹¹²

Some districts are building comprehensive supports for teachers on top of their evaluation systems. Beginning in 2016-17, the District of Columbia will assign the bulk of its teaching force to “content teams” under a new Learning Together to Advance Our Practice initiative. Led by instructional coaches, lead teachers, assistant principals and (at the high school level) department chairs, the teams will meet for 90 minutes a week to plan lessons, deepen their subject-matter knowledge, and review student work and school data. The team leaders will regularly observe teachers in their classrooms and provide them with more informal feedback than what is provided under the District's IMPACT evaluation system.¹¹³

Through such initiatives, districts hope to signal to teachers that they are valued professionals doing important work, while at the same time reducing the isolation of the classroom, enhancing collegiality, and promoting a general culture of improvement—things teachers say they prize.

While these initiatives boost teacher morale, they also serve a strictly practical function. Despite the reform movement's early singular focus on removing “bad apples” from the teaching pool, most school districts can't fire their way to a stronger teaching force for the simple reason that they cannot hire more qualified replacements.

Even some of the most prominent proponents of student achievement as a key measure of teacher performance now question the wisdom of the “bad apple” strategy. “If school systems can figure out a way to reduce [teacher] anxiety and support [improvements in] instruction, we're on a promising path,” says Steve Cantrell of the Bill & Melinda Gates Foundation, commenting on the arc of the reform movement. “If not, we're in trouble.”¹¹⁴

Tom Kane attributes some of that trouble to a one-size-fits-all approach to evaluation. “We failed to make a distinction between systems used to evaluate probationary teachers, where a high-stakes decision is required, where prediction is the point, and feedback to help existing teachers improve,” says Kane. “By treating untenured and tenured teachers alike, the accountability focus has made it hard to talk about evaluation as a feedback mechanism for post-tenured teachers. The failure to make that distinction has created the most anxiety, the most acrimony.”¹¹⁵

Teachers consistently say they want to work in environments where they feel valued and where their work is taken seriously, where they have opportunities to work with others to hone their craft. In a recent example, the Paris-based Organisation for Economic Co-operation and Development surveyed over 100,000 teachers from nearly three dozen developed nations and concluded that “the more frequently teachers collaborated with colleagues the higher their job satisfaction.”¹¹⁶

While some teachers are unlikely to support evaluation reform under any circumstances, there are signs that the recent efforts by policymakers to respond to teachers' concerns are paying dividends. Growing evidence suggests that the "reforms of the reforms"—especially the push for evaluation to play a greater role in helping teachers improve their practice—are bringing some teachers around to the new systems.

Consider Tennessee. In 2012, following the first year of the state's new evaluation system, 38 percent of Tennessee's 36,700 public school teachers told researchers they believed the new system improved teaching in the state, and 67 percent reported that they liked working in their schools. By 2015, the proportion of teachers believing that the new TEAM system was improving Tennessee teaching had increased to 68 percent, and 80 percent of the state's teachers liked their jobs.¹¹⁷

The increasing use of evaluation results to help teachers improve rather than merely reward or remove them, as well as the deepening involvement of teachers in some states and districts in the planning and implementation of reforms, may make it easier for teacher unions to embrace the reforms as a way to create more professional working environments for their members and strengthen the teaching profession.

Albert Shanker, the founding father of American teacher unionism, made that case three decades ago, as president of the American Federation of Teachers. "We don't have the right to be called professionals—and we will never convince the public that we are," he told a union convention in Niagara Falls in 1985, "unless we are prepared honestly to decide what constitutes competence in our profession and what constitutes incompetence and apply those definitions to ourselves and our colleagues."¹¹⁸

No evaluation system is perfect, and as this report makes clear, the work to create a new performance paradigm in public school teaching is far from complete, despite the magnitude of change in the past few years. A complex policy change at

the heart of the education enterprise is a long-term proposition. It is not going to be perfected overnight.

But we cannot build public school teaching into the profession that policymakers, taxpayers, and teachers themselves want—and that Al Shanker envisioned—using superficial evaluation systems as a foundation. It is simply impossible to strengthen instruction, make teaching more attractive work, and raise student achievement without understanding individual teachers' strengths and weaknesses and without making performance matter. You can't help people improve if you don't know what needs improving—even if the measurement of teacher performance is ultimately an inexact science.

The introduction of teacher evaluation reform to public education has been fast and furious, thanks to federal incentives. In many states and school districts the infrastructure of change is still catching up to reformers' aspirations. But signs of progress are increasingly visible. The hard-learned lessons of the past several years suggest that building on that progress, staying the course on reform despite the dissolution of the Obama incentives, is in the best interests of both students and teachers.

Recommendations

Here’s what state and local policymakers can do to improve the quality of the new evaluation systems, enhance their utility in schools, and, importantly, increase their legitimacy in teachers’ eyes.

The cost and complexity of the new evaluation systems argue for standardized statewide models like those in place in South Carolina and Delaware. It is vastly more efficient to rely on several dozen statewide systems than to expect the nation’s 13,300 school districts to respond to this challenge independently. As commentator Matt Miller wrote in *The Atlantic* at the end of the George W. Bush administration, relying on local school boards to meet today’s educational challenges would be “as if after Pearl Harbor, FDR had suggested we prepare for war through the uncoordinated efforts of thousands of small factories.”¹¹⁹

Many of the nation’s small school systems—some 6,900 enroll under 1,000 students—simply lack the resources and expertise to build dependable evaluation infrastructures, the Southern Regional Education Board (SREB) concluded.¹²⁰ Other studies suggest that when given autonomy under state systems, many districts wind up creating low-quality evaluation systems.¹²¹ And teachers trust statewide systems more than local models, SREB’s research has found. “The sweet spot,” former SREB researcher Tysza Gandha says, “would be having states provide [school districts with] several evaluation options.”¹²²

Comprehensive Models

It has also become increasingly clear that **comprehensive evaluation models** (those that combine classroom observations, student-achievement measures, student surveys, and perhaps other metrics) **yield evaluations that are more dependable, more likely to be trusted by teachers, and more likely to produce information that can help teachers improve their practice.**

Strengthening Classroom Observations

The first step in relegating superficial checklist observations to the past is **ensuring that teachers receive multiple observations during a school year by multiple evaluators**, including at least one evaluator from outside the teacher’s school. “Multiple evaluators is the message from the MET study that has been most ignored,” says Tom Kane of the Harvard Graduate School of Education and the principal author of the MET study. Yet multiple evaluations yield more reliable results and garner more respect from teachers.¹²³

Having teachers serve as evaluators of peers in other schools is also a sound strategy. “It gives teachers ownership of the process, a sense of professionalism, and encourages conversations around good teaching,” says Sandi Jacobs of the National Council on Teacher Quality. Teachers already play these roles in Cincinnati, New Haven, and other cities.¹²⁴

To encourage peer review, **ensure that state policies and local collective-bargaining contracts permit teachers to contribute to teacher evaluations.**

Require that all evaluators be certified before allowing them into classrooms. This would increase the reliability and utility of observations, strengthen administrators’ abilities as instructional leaders, and improve teacher morale. “Teachers say in surveys that they trust certified principals more,” says Luke Kohlmoos, the former director of Tennessee’s teacher evaluations.¹²⁵

Have observations done by content and grade-level experts as a way of connecting observations and post-observation feedback more closely to the “what” of instruction rather than just the “how.”

For new teachers and low-performers, focus the first evaluations of the year on feedback and don’t count the results toward the teachers’ final ratings—an approach that will lower anxiety levels and help to establish evaluation as an improvement tool.

To improve the quality of feedback teachers receive after their observations, **establish feedback protocols and audit a percentage of post-observation conferences.**

To lower costs, **require fewer observations of teachers who have been highly rated over the two previous years.**

To help counter inconsistency in observation scores and build teacher confidence in the system, **eliminate from final performance ratings any observation scores that fall substantially below the average.**

Joint teacher-administrator training on observation rubrics (and the components of evaluation systems more generally) helps ensure transparency and shared understanding.

Improve and Repurpose Student-Achievement Measures

Although student-achievement measures have supplied valuable new information for schools, policymakers, and researchers, they are far from perfect and policymakers should do everything possible to mitigate their weaknesses.

Use at least two years of student achievement data in teacher evaluation ratings. Research reveals that value-added ratings more accurately reflect teachers' true performance when they are based on at least two years of student test scores.

Weigh student achievement as no more than one-third of a teacher's rating. The MET study found that evenly weighting achievement, observations, and student surveys produced the most reliable overall results.¹²⁶

A lower student-achievement weighting can help reconcile two conflicting policy priorities: 1) ensuring that there are top teachers in challenging schools and 2) incorporating student-achievement results into teacher evaluations. Research shows that working in schools with disadvantaged students makes it tougher for teachers to earn higher value-added ratings. Reducing (but not eliminating) the role of student

achievement in teacher ratings would help relieve this tension and help make evaluation reform more palatable to teachers, whose aversion to value-added measures is clear in surveys.

Researchers at the Brookings Institution have proposed a related strategy of awarding **teachers extra credit on both observation ratings and student-achievement results** based on their students' demographics.¹²⁷

Harvard's Tom Kane, perhaps the nation's most prominent advocate of using student-achievement results in teacher evaluations, proposes another way to reduce the influence of such results in the face of strong teacher opposition. **Give value-added scores a substantial role in the assessment of probationary teachers,** Kane argues, upwards of 50 percent of teachers' total scores, while assigning a much lower weight to value-added ratings for tenured teachers. Value-added scores predict future performance, he argues, and that's exactly what we want to know about teachers early on in their careers, before they are granted tenure. "We can't ignore the fact that we have to make high-stakes decisions for probationary teachers," Kane says.¹²⁸

Another way to limit the impact of student-achievement results would be to use them only as an initial screen in the larger evaluation system. Because value-added

Elements of Effective Feedback Conferences

- Start with an affirmation of what's working;
- Allow teachers to share their perspectives first;
- Focus on improvement;
- Have a supportive demeanor;
- Build an improvement plan that's both parties agree on;
- Make the plan concrete, actionable.

SOURCE: Jeannie Myung and Krissia Martinez, "Strategies for Enhancing the Impact of Post-Observation Feedback for Teachers," Carnegie Foundation for the Advancement of Teaching, 2013.

scores are most useful for identifying the best and worst teachers in a school, Douglas Harris, the director of Tulane University’s Education Research Alliance for New Orleans, has proposed eliminating achievement results in teachers’ ratings generally and using them instead to target very low-scoring teachers for more intensive scrutiny, as well as to audit very high classroom-observation ratings and identify top-performing teachers for new roles and responsibilities. “Classroom observations should be the core of the [evaluation] process, it’s what’s most trusted,” says Harris, who proposes the VAM-as-screen strategy in his 2011 book, *Value-Added Measures in Education*.¹²⁹

Similarly, **policymakers should eliminate the use of school-wide student achievement results to measure individual teacher performance.** While it’s fine, in theory, to argue that every teacher bears responsibility for the performance of every student in her school, the hard-learned lessons of the past few years suggest that this strategy engenders too much ill-will between teachers to be worthwhile; teachers in non-tested grades and subjects resent being judged on the basis of their colleagues’ effectiveness.

To improve the quality of student learning objectives (SLOs) used to measure student progress for teachers in grades and subjects lacking standardized tests, Tom Kane proposes that **groups of teachers work to create common assessments in schools and school districts by grade and subject, and that teachers score the tests of students outside their classrooms**—steps that would engender faculty-wide conversations about standards and instruction. Kane’s is a promising if labor-intensive solution to the unreliability of most SLOs used in teacher evaluations today. An alternative would be to simply eliminate their role in evaluations.

Build Stronger Bridges to Professional Development

Policymakers have a chance to achieve a greater return on their investments in new evaluation systems by **systematically tying evaluation systems to professional-development efforts.**

The District of Columbia has sought to do that by creating new online resources linked to the school district’s teaching standards that teachers can use to address weaknesses revealed in their evaluations. And increasingly, it is having principals and instructional coaches incorporate those resources into more structured professional development activities that are complemented with instructional coaches working with teachers on targeted topics in six-week cycles. “You can’t merely present people with evaluations and assume they’ll improve, or even give them feedback and assume they’ll do professional development on their own,” says Scott Thompson, deputy chief for innovation and design in the District’s office of instructional practice.¹³⁰

Incentivize Principals

School principals are key to building stronger evaluation systems, experts say. Even under multi-rater models, they’ll continue to conduct a large share of classroom observations, and they’re the primary bridge in many schools between evaluation results and opportunities for teachers to improve their practice. It’s important, then, to ensure that principals are as fully invested as possible in the new systems.

One way to do that is to **give principals a greater say in hiring their teachers than exists in many school districts today, coupled with stricter accountability for teacher performance.** Another strategy is to make the demonstration of effective classroom observation skills a component of principals’ own evaluations. This would force principals to prioritize classrooms over the competing demands of school management. In this spirit, Massachusetts has begun making effective evaluation and feedback skills part of principal licensure.¹³¹

Another strategy would be **to divide principals’ jobs in half, having one person focus on school management and another on instruction,** under a co-leadership model of the sort used by the charter school network Uncommon Schools.

Incentivize Teachers

One way to get teachers to embrace more comprehensive evaluation systems is through professional rewards, using evaluation results as the foundation for performance pay and new roles and responsibilities for successful teachers. Teachers will support evaluations more fully if the results are tied to opportunities to improve their practice and advance professionally.

Finding Efficiencies

There's no doubt that the comprehensive teacher evaluation systems emerging in states and school districts today cost more than the simplistic systems they're replacing. The District of Columbia's multi-measure model costs about \$1,000 per teacher per year, beyond the funding required to build the school district's new evaluation infrastructure.¹³² With the Obama administration's federal incentive funding under its Race to the Top and NCLB waiver programs ending, finding ways to sustain the new evaluation systems is an increasing priority.

Yet strategies are emerging to reduce costs without compromising quality. An increasing number of districts, the District of Columbia among them, are **linking the frequency of evaluative observations to each teacher's prior results** and reducing the required number of observations for higher-rated teachers.

Other districts are **replacing in-person observations with the recording of lessons, to be used for training, scoring, and feedback.** The MET study found this practice led to "lower costs, greater ease of use, and better quality."¹³³

Repurposing Federal Funding

There are also opportunities to repurpose federal education aid to support the new evaluation systems. States and school districts could reasonably **use federal funding under Title II and, in some instances, Title I of the new Every Student Succeeds Act** to build out their teacher evaluation systems as a way of improving teacher quality and strengthening instruction. Traditionally, the majority of Title II monies has

been spent on classroom aides and mediocre professional development. Focusing the funding on comprehensive evaluation systems that help teachers improve their practice would surely produce greater returns.

Similarly, a Title II program that supports the state and local development of performance-based compensation systems in public education, now called the **Teacher and School Leader Incentive Program,** could logically be spent on sustaining new evaluation systems. Performance-based compensation requires knowing which teachers are performing and which aren't.

Ultimately, state and local education leaders should treat evaluation reform as an investment, not merely as a cost.

Additional Research

With many evaluation systems early in their evolution, and many policymakers forced to introduce reforms at a rapid pace, more research is needed on several fronts.

One priority would be to **try to salvage student learning objectives as measures of student achievement.** Though many of the objectives introduced to date are superficial and unreliable measures of student performance, SLOs have the potential to tie evaluations more closely to school- and classroom-level instructional aims than value-added measures. SLOs give teachers and principals a larger stake in the student-achievement side of evaluations, and, partly as a result, they're less controversial among educators.

There is also much work to be done to **help states and districts measure the impact of their new teacher-evaluation systems.** The smattering of research done so far suggests that comprehensive evaluations are raising the caliber of teachers in classrooms and contributing positively to student achievement. But there's much more to learn about the consequences of the reform movement on students, teachers, and public education generally. At this stage, says Tysza Gandha, "We don't even have the capacity to collect the necessary information to answer key research questions in many places."

ENDNOTES

¹ Thomas Toch and Robert Rothman, *Rush to Judgment: Teacher Evaluation in Public Education* (Washington, DC: Education Sector, 2008), http://educationpolicy.air.org/sites/default/files/publications/RushToJudgment_ES_Jan08.pdf.

² Daniel Weisberg et al., *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness*, 2nd ed. (Washington, DC: The New Teacher Project, 2009).

³ Steven Brill, “The Rubber Room: The Battle over New York City’s Worst Teachers,” *The New Yorker*, August 31, 2009, <http://www.newyorker.com/magazine/2009/08/31/the-rubber-room>.

⁴ A 2006 study of New York City public school teachers by researchers from Harvard, Dartmouth, and Columbia found that students of unlicensed teachers performed just as well as those whose teachers were fully certified. Thomas J. Kane, Jonah E. Rockoff, and Douglas O. Staiger, *What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City* (working paper no. w12155, Cambridge, MA: National Bureau of Economic Research, 2006), <http://ssrn.com/abstract=896463>. Other researchers, including Steven Rivkin of the University of Illinois—Chicago, Eric Hanushek of Stanford, and the late John Kain of the University of Texas—Dallas, found substantial differences in teachers’ impact on student test scores and concluded on the basis of statistical models that removing public education’s lowest-performing teachers would lift student achievement significantly. Steven G. Rivkin, Eric A. Hanushek, and John F. Kain, “Teachers, Schools, and Academic Achievement,” *Econometrica* 73, no. 2 (March 2005): 417-458.

⁵ William L. Sanders and Sandra P. Horn, “Research Findings from the Tennessee Value-Added Assessment System (TVAAS) Database: Implications for Educational Evaluation and Research,” *Journal of Personnel Evaluation in Education* 12, no. 3 (1998): 247-256.

⁶ Within two years of being named chancellor of the Washington, DC, school district in 2007, Michelle Rhee, the former president of The New Teacher Project, launched a sweeping plan to transform teaching in the nation’s capital into a performance-based profession, a plan built on a new, comprehensive teacher-evaluation system that quickly became a national model.

⁷ US Secretary of Education Arne Duncan and his advisors resolved to use the extraordinary circumstances of the nation’s fiscal crisis to leverage policy change in public education, making new, more rigorous evaluation systems a key criteria for \$4.35 billion in competitive Race to the Top grants to states and school systems under the American Recovery and Reinvestment Act, the \$787 billion stimulus package approved by Congress in 2009. The department declared states would be ineligible for the grants if they barred the use of student-achievement results in teacher evaluations. As a further catalyst, Duncan required in 2012 that states applying to the Department of Education for waivers to provisions of the Elementary and Secondary Education Act had to “develop, adopt, pilot, and implement” teacher (and principal) evaluation systems that stressed student achievement.

⁸ Kathryn M. Doherty and Sandi Jacobs, *State of the States 2015: Evaluating Teaching, Leading, and Learning* (Washington, DC: National Council on Teacher Quality, 2015), 20.

⁹ *Ibid.*, 2. See also Matthew P. Steinberg and Morgaen L. Donaldson, “The New Educational Accountability: Understanding the Landscape of Teacher Evaluation in the Post-NCLB Era,” *Education Finance and Policy* (Forthcoming), 8.

¹⁰ Doherty and Jacobs, *State of the States 2015*, 2.

¹¹ “The Framework for Teaching,” The Danielson Group, accessed April 22, 2016, <https://www.danielsongroup.org/framework/>.

¹² Jim Hull, *Trends in Teacher Evaluation: How States Are Measuring Teacher Performance* (Alexandria, VA: The Center for Public Education, October 2013).

¹³ Taylor White, *Adding Eyes: The Rise, Rewards, and Risks of Multi-Rater Teacher Observation Systems* (issue brief, Stanford, CA: Carnegie Foundation for the Advancement of Teaching, December 2014), 2. In a 2013 study of new evaluation models in the state, the University of Michigan’s Institute for Social Research not surprisingly found that “the biggest gains in measurement reliability come when moving from one observation to about four observations.” See Brian Rowan et al., *Promoting High Quality Teacher Evaluations in Michigan: Lessons from a Pilot of Educator Effectiveness Tools* (Ann Arbor, MI: University of Michigan Institute for Social Research, December 2013), 16.

¹⁴ Steinberg and Donaldson, “The New Educational Accountability,” 12.

¹⁵ White, *Adding Eyes*, 2-13.

¹⁶ *Ibid.*, 2.

¹⁷ *Ibid.*, 5.

¹⁸ *Ibid.*, 12.

¹⁹ *Ibid.*, 2-10.

²⁰ *Ensuring Fair and Reliable Measures of Effective Teaching: Culminating Findings from the MET Project’s Three-Year Study* (The Bill & Melinda Gates Foundation, 2013). The study involved 3,000 teachers in Dallas, Denver, New York, and four other urban school districts.

²¹ Eric S. Taylor and John H. Tyler, “The Effect of Evaluation on Teacher Performance,” *American Economic Review* 2012 102, no. 7: 3628-3651.

²² Sanders and Horn, “Research Findings from the Tennessee Value-Added Assessment System (TVAAS) Database.”

²³ Steinberg and Donaldson, “The New Educational Accountability.”

²⁴ Tripod Education Partners, <http://tripoded.com/>

²⁵ Tennessee Department of Education, *Teacher and Administrator Evaluation in Tennessee: A Report on Year 3 Implementation* (Nashville, TN: Tennessee Department of Education, 2015), 34.

ENDNOTES *continued*

- ²⁶ Steinberg and Donaldson, “The New Educational Accountability,” 29.
- ²⁷ Jeff Schulz, Gunjan Sud, and Becky Crowe, *Lessons from the Field: The Role of Student Surveys in Teacher Evaluation and Development* (Washington, DC: Bellwether Education Partners, 2014), 8. See also The Aspen Institute, *Teacher Evaluation and Support Systems: A Roadmap for Improvement* (Washington, DC: The Aspen Institute, 2016), 11.
- ²⁸ Some school systems also use other measures, including “contributions to school culture,” peer surveys, and “professionalism.”
- ²⁹ *Ensuring Fair and Reliable Measures of Effective Teaching* (Bill & Melinda Gates Foundation, 5).
- ³⁰ Scholastic and The Bill & Melinda Gates Foundation, *Primary Sources: America’s Teachers on Teaching in an Era of Change*, 3rd ed., (The Bill & Melinda Gates Foundation, 2013), 4. <http://www.scholastic.com/primarysources/PrimarySources3rdEdition.pdf>.
- ³¹ “The Framework for Teaching,” The Danielson Group, accessed April 22, 2016, <https://www.danielsongroup.org>.
- ³² Tysza Gandha, interview with the author, April 2015.
- ³³ Heather Peske, interview with the author, April 2015.
- ³⁴ Noah Bookman, interview with the author, March 2015.
- ³⁵ Joanne Weiss, interview with the author, April 2015.
- ³⁶ Heather Kirkpatrick, interview with the author, May 2015.
- ³⁷ See Steven G. Rivkin, Eric A. Hanushek, and John F. Kain, “Teachers, Schools, and Academic Achievement,” *Econometrica* 73, no. 2 (March 2005): 417-458.
- ³⁸ Matthew A. Kraft and Allison F. Gilmour, “Revisiting the Widget Effect: Teacher Evaluation Reforms and the Distribution of Teacher Effectiveness,” Brown University Working Paper, February 2016.
- ³⁹ Toch and Rothman, *Rush to Judgment*, 3.
- ⁴⁰ Correspondence with Alden Wells, District of Columbia Public Schools, March 21, 2016.
- ⁴¹ Chad Aldeman and Carolyn Chuong, *Teacher Evaluations in an Era of Rapid Change: From ‘Unsatisfactory’ to ‘Needs Improvement’* (Washington, DC: Bellwether Education Partners, 2014) 20-21, http://bellwethereducation.org/sites/default/files/Bellwether_TeacherEval_Final_Web.pdf.
- ⁴² Melinda Adnot, Thomas Dee, Veronica Katz, and James Wyckoff, “Teacher Turnover, Teacher Quality, and Student Achievement in DCPS” (working paper no. 21922, National Bureau of Economic Research, January 2016), <http://www.nber.org/papers/w21922>. Wyckoff and different researchers, Susanna Loeb of Stanford and Luke Miller of the University of Virginia, reported in 2014 that evaluation reforms in New York City have encouraged weaker teachers to leave the profession there, as well. See Susanna Loeb, Luke C. Miller, and James Wyckoff, “Performance Screens for School Improvement: The Case of Teacher Tenure Reform in New York City,” *Educational Researcher* 44, no. 4 (May 2015): 199-212.
- ⁴³ John P. Papay, Eric S. Taylor, John H. Tyler, and Mary Laski, “Learning Job Skills from Colleagues at Work: Evidence from a Field Experiment Using Teacher Performance Data” (working paper 21986, National Bureau of Economic Research, 2016), <http://www.nber.org/papers/w21986>. See also Kaylan Connally and Melissa Tooley, *Beyond Ratings: Re-envisioning State Teacher Evaluation Systems as Tools for Professional Growth* (Washington, DC: New America, 2016) 25.
- ⁴⁴ *Ensuring Fair and Reliable Measures of Effective Teaching* (Bill & Melinda Gates Foundation, 21). The foundation has more recently given nine school districts and several charter management organizations some \$20 million to improve their professional development programs. And it has invested heavily in the development of new instructional materials to help teachers tackle the challenges of the Common Core State Standards.
- ⁴⁵ Doherty and Jacobs, *State of the States 2015*, 2.
- ⁴⁶ White, *Adding Eyes*, 2-13.
- ⁴⁷ Connally and Tooley, *Beyond Ratings*, 25. See also *Core Leadership: Teacher Leaders and Common Core Implementation in Tennessee*, The Aspen Institute, 2014.
- ⁴⁸ Taylor and Tyler, “The Effect of Evaluation on Teacher Performance,” 2012.
- ⁴⁹ Jess Wood, interview with the author, March 2015.
- ⁵⁰ Raj Chetty, John N. Friedman, and Jonah E. Rockoff, “Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood,” *American Economic Review* 104, no. 9 (September 2014): 2633-2679.
- ⁵¹ Dan Goldhaber et al., *Carnegie Knowledge Network Concluding Recommendations* (Stanford, CA: Carnegie Foundation for the Advancement of Teaching, 2015).
- ⁵² Scott Marion, interview with the author, May 2015.
- ⁵³ Peter Z. Schochet and Hanley S. Chiang, *Error Rates in Measuring Teacher and School Performance Based on Student Test Score Gains*, NCEE 2010-4004, (Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education, July 2010).
- ⁵⁴ Heather M. Buzick and Nathan D. Jones, “Using Test Scores from Students with Disabilities in Teacher Evaluation,” *Educational Measurement: Issues and Practice* 34, no. 3 (2015): 34.
- ⁵⁵ In a study of four urban school districts, Brookings Institution researchers found that only 22 percent of teachers were evaluated on test score gains. See Grover J. (Russ) Whitehurst, Matthew M. Chingos, and Katharine M. Lindquist, *Evaluating Teachers with Classroom Observations: Lessons from Four Districts* (Washington, DC: Brown Center on Education Policy at Brookings Institution, May 2014).
- ⁵⁶ David Smiley, “For Thousands of Florida Teachers, Evaluations Aren’t Making the Grade,” *Miami Herald*, November 29, 2013.
- ⁵⁷ Sandi Jacobs, interview with the author, May 2015.

ENDNOTES *continued*

- ⁵⁸ Five testing experts at the National Center for the Improvement of Educational Assessment concluded in a comprehensive 2014 study of state practices for evaluating teachers in non-tested subjects and grades that “evaluation procedures for this population [of teachers] has greatly lagged behind that of other teachers,” and that it is “extremely difficult” to come up with measures of this sort that are “rigorous and comparable across schools within districts.” Erika Hall et al., “State Practices Related to the Use of Student Achievement Measures in the Evaluation of Teachers in Non-Tested Subjects and Grades,” National Center for the Improvement of Educational Assessment, August 26, 2014, 3.
- ⁵⁹ Luke Kohlmoos, interview with the author, April 2015.
- ⁶⁰ Hall et al., “State Practices.”
- ⁶¹ Andy Baxter and Tysza Gandha, *Toward Trustworthy and Transformative Classroom Observations: Progress, Challenges and Lessons in SREB States* (Atlanta, GA: Southern Regional Education Board, 2015).
- ⁶² *Ibid.*, 12. Also, Scott Marion, interview with the author, March 2015.
- ⁶³ Heather Kirkpatrick, interview with the author, May 2015.
- ⁶⁴ The 2013 survey of teachers and principals by the University of Michigan’s Institute for Social Research found that the median number of days a year principals spent on evaluation in the state was 31. Rowan et al., “Promoting High Quality Teacher Evaluations in Michigan,” 4.
- ⁶⁵ Dana Markow, Lara Macia, Helen Lee, Harris Interactive, *The MetLife Survey of the American Teacher: Challenges for School Leadership: A Survey of Teachers and Principals* (New York: MetLife, 2013).
- ⁶⁶ Hardy Murphy et al., “Indiana Teacher Evaluation: At the Crossroads of Implementations,” Center on Education and Lifelong Learning, Indiana University, 2014, 24.
- ⁶⁷ The District of Columbia Public Schools reported that only a third of its principals in 2009-10 were still working in the school system at the start of the 2013-14 school year, through a combination of voluntary and involuntary attrition.
- ⁶⁸ Whitehurst, Chingos, and Lindquist, *Evaluating Teachers with Classroom Observations*, 2014.
- ⁶⁹ *Ibid.* Also, Baxter and Gandha, *Toward Trustworthy and Transformative Classroom Observations. And Ensuring Fair and Reliable Measures of Effective Teaching*,” Bill & Melinda Gates Foundation, 16.
- ⁷⁰ Thomas Kane, interview with the author, April 2015.
- ⁷¹ For example, 25 percent of Tennessee’s teachers earned value-added scores in 2011-12 that put them in the state’s lowest two rating categories (out of five), while only 2.5 percent of the state’s teachers received observation ratings in those categories. See *Teacher Evaluation in Tennessee: A Report on Year 1 Implementation* (Tennessee Department of Education, July 2012, 33).
- ⁷² Patrick McGuinn, *Evaluating Progress: State Education Agencies and the Implementation of New Teacher Evaluation Systems* (white paper #WP2015-09, University of Pennsylvania Consortium for Policy Research in Education, 2015), 6.
- ⁷³ The New Teacher Project, *The Mirage: Confronting the Hard Truth About Our Quest for Teacher Development* (Washington, DC: The New Teacher Project, 2015), http://tntp.org/assets/documents/TNTP-Mirage_2015.pdf.
- ⁷⁴ Murphy et al., “Indiana Teacher Evaluation,” 22. Also, in the 2013 national survey of 20,000 teachers by Scholastic and the Bill & Melinda Gates Foundation, only 17 percent of teachers said “classroom supports and resources have been identified to meet my needs,” and only 13 percent said that “professional learning/development opportunities have been customized to meet my needs.” See *Primary Sources*, <http://www.scholastic.com/primarysources/PrimarySources3rdEdition.pdf>. See also *Teachers Know Best: Teachers’ Views on Professional Development* (Bill & Melinda Gates Foundation, December 2014), <http://k12education.gatesfoundation.org/wp-content/uploads/2015/04/Gates-PDMarketResearch-Dec5.pdf>.
- ⁷⁵ See Duncan’s Gambit below.
- ⁷⁶ Stefanie K. Reinhorn, “Seeking Balance Between Assessment and Support: Teachers’ Experiences of Teacher Evaluation in Six High-Poverty Urban Schools,” (working paper, The Project on the Next Generation of Teachers, December 2013), 1.
- ⁷⁷ Murphy et al., “Indiana Teacher Evaluation,” 19.
- ⁷⁸ *The MetLife Survey of the American Teachers*.
- ⁷⁹ <http://www.gallup.com/poll/178997/teachers-favor-common-core-standards-not-testing.aspx>.
- ⁸⁰ Murphy et al., “Indiana Teacher Evaluation,” 15.
- ⁸¹ Heather Kirkpatrick, interview with the author, April 2015.
- ⁸² “NASSP Statement Rejects Value-Added Measurement in Teacher Evaluation,” (press release, National Association of Secondary School Principals, December 3, 2014).
- ⁸³ Personal correspondence with Steve Cantrell, April 21, 2015.
- ⁸⁴ “National Education Association Resolution D-21,” *2016 NEA Handbook*. http://www.nea.org/assets/docs/Resolutions_2016_NEA_Handbook.pdf.
- ⁸⁵ “National Education Association Resolution F8 and F9,” *2016 NEA Handbook*, http://www.nea.org/assets/docs/Resolutions_2016_NEA_Handbook.pdf.
- ⁸⁶ Randi Weingarten, “A New Path Forward: Four Approaches to Quality Teaching and Better Schools,” (as prepared for delivery, Washington, DC, January 12, 2010), http://www.aft.org/sites/default/files/wysiwyg/sp_weingarten011210.pdf.
- ⁸⁷ Stephen Sawchuk, “AFT’s Weingarten Backtracks on Using Value-Added Measures for Evaluations,” *Teacher Beat, Education Week*, January 7, 2014.
- ⁸⁸ Maggie Severns, “In New Education Bill, ‘Christmas’ Comes Early for Unions,” *Politico*, December 9, 2015.

ENDNOTES *continued*

- ⁸⁹ Denise Jewell Gee, “NYSUT President Urges Parents to Opt Out of State Tests,” *The Buffalo News*, March 30, 2015, <http://schoolzone.buffalonews.com/2015/03/30/nysut-president-urges-parents-to-opt-out-of-state-tests/>.
- ⁹⁰ Stephen Sawchuk, “Teacher Evaluation Heads to the Courts,” *Education Week*, Oct 7, 2015, <http://www.edweek.org/ew/section/multimedia/teacher-evaluation-heads-to-the-courts.html>.
- ⁹¹ *Ibid.*
- ⁹² Marla Ucelli-Kashyap, interview with the author, April 2015.
- ⁹³ Rob Weil, interview with the author, April 2015.
- ⁹⁴ McGuinn, *Evaluating Progress*, http://www.cpre.org/sites/default/files/policybrief/2061_pbm McGuinn2.pdf.
- ⁹⁵ Charlotte Danielson, correspondence with the author, April 2015.
- ⁹⁶ *TNTP Core Teaching Rubric: A Tool for Conducting Common Core-Aligned Classroom Observations* (The New Teacher Project, February 18, 2014), <http://tntp.org/publications/view/tntp-core-teaching-rubric-a-tool-for-conducting-classroom-observations>. The new rubric measures teachers’ content knowledge as well as their instructional skills.
- ⁹⁷ Michelle Hudacsko, interview with the author, April 2016.
- ⁹⁸ Baxter and Gandha, “Toward Trustworthy and Transformative Classroom Observations,” 7.
- ⁹⁹ Grace Tatter, “Tennessee’s Teacher Evaluation System Improving, State Report Says,” *Chalkbeat*, April 9, 2015, <http://tn.chalkbeat.org/2015/04/09/tennessees-teacher-evaluation-system-improving-state-report-says/#.VnmyxP3SnCZ>. See also The Aspen Institute, *Teacher Evaluation and Support Systems*, 10.
- ¹⁰⁰ Calculations by the author based on District of Columbia Public Schools. The \$1,000 figure does not count the time principals spend evaluating teachers (seemingly an instruction-related activity principals should be doing) or time spent working with teachers of non-tested grades and subjects to capture their students’ achievement through SLOs.
- ¹⁰¹ Taylor and Tyler, 2012, “*The Effect of Evaluation on Teacher Performance*,” 26-28. They found that the city’s evaluation system produced higher-achieving teachers and that students of those teachers would enjoy higher lifelong earnings as a result of learning more under those teachers.
- ¹⁰² White, *Adding Eyes*, 3.
- ¹⁰³ *Ibid.*
- ¹⁰⁴ Michelle Hudacsko, interview with the author, April 2016.
- ¹⁰⁵ Tom Kane, interview with the author, May 2015.
- ¹⁰⁶ Douglas N. Harris, *Value-Added Measures in Education What Every Educator Needs to Know* (Cambridge, MA: Harvard Education Press, 2011).
- ¹⁰⁷ Julie Marsh, interview with the author, May 2015.
- ¹⁰⁸ Heather Peske, interview with author, April 2015. Catherine Brown, Lisette Partelow, and Annette Konoske-Graf, *Educator Evaluation: A Case Study of Massachusetts’ Approach* (Washington, DC: Center for American Progress, March 16, 2016).
- ¹⁰⁹ See Morgaen L. Donaldson and John P. Papay, “An Idea Whose Time Had Come: Negotiating Teacher Evaluation Reform in New Haven, Connecticut,” *American Journal of Education* 122, no. 1 (November 2015): 39-70.
- ¹¹⁰ “Harvard Center’s Best Foot Forward Project Shares Results on the Use of Video in Classroom Observations,” Center for Education Policy Research, Harvard University, October 5, 2015, <http://cepr.harvard.edu/news/harvard-center%E2%80%99s-best-foot-forward-project-shares-results-use-video-classroom-observations>.
- ¹¹¹ The Aspen Institute, *Teacher Evaluation and Support Systems*, 6.
- ¹¹² See *Classroom Observations: Measuring Teachers’ Professional Practices*, Education First, November 2014, 3. How feedback is given is, itself, important. “If teachers don’t sense that their core abilities are under indictment, they are more likely to see [feedback conversations] as an opportunity for growth,” writes Jeannie Myung and Krissia Martinez in a report on effective feedback for the Carnegie Foundation for the Advancement of Teaching. See Jeannie Myung and Krissia Martinez, *Strategies for Enhancing the Impact of Post-Observation Feedback for Teachers* (Stanford, CA: Carnegie Foundation for the Advancement of Teaching, July 2013), 5, <http://www.carnegiefoundation.org/resources/publications/strategies-enhancing-impact-post-observation-feedback-teachers/>.
- ¹¹³ Jason Kamras, *LEAP/IMPACT Webinar for Teachers*,” February 25, 2016, <https://dcnet.webex.com/dcnet/ldr.php?RCID=4d9f382c22351d9e402830a45f842086>.
- ¹¹⁴ Steve Cantrell, interview with the author, April 2015.
- ¹¹⁵ Tom Kane, interview with the author, May 2015.
- ¹¹⁶ Organisation for Economic Co-operation and Development, “*Supporting Teacher Professionalism: Insights from TALIS 2013*,” (powerpoint, 50), <http://www.slideshare.net/OECD/EDU/supporting-teacher-professionalism-insights-from-talis-2013>.
- ¹¹⁷ Joan Brasher, “Tennessean: TN Teachers Happier with Evaluations; Testing a Burden,” *Vanderbilt University: Research News at Vanderbilt*, August 27, 2015, <http://news.vanderbilt.edu/2015/08/tennessean-tn-teachers-happier-with-evaluations-testing-a-burden/>. See also *Tennessee Educator Survey Report* (Tennessee Department of Education: Division of Data and Research, 2015), http://tn.gov/assets/entities/education/attachments/data_survey_report_2015.pdf.
- ¹¹⁸ Thomas Toch, *In the Name of Excellence: The Struggle to Reform the Nation’s Schools, Why It’s Failing, and What Should Be Done* (Oxford University Press, 1991), 143.

ENDNOTES *continued*

- ¹¹⁹ Matt Miller, “First, Kill All the School Boards,” *The Atlantic*, January/February 2008. <http://www.theatlantic.com/magazine/archive/2008/01/first-kill-all-the-school-boards/306579/>.
- ¹²⁰ Baxter and Gandha, “Toward Trustworthy and Transformative Classroom Observations,” 7.
- ¹²¹ Aldeman and Chuong, *Teacher Evaluations in an Era of Rapid Change*, 2014.
- ¹²² Tysza Gandha, interview with the author, March 2015.
- ¹²³ Tom Kane, interview with the author, May 2015.
- ¹²⁴ Sandi Jacobs, interview with the author, March 2015.
- ¹²⁵ Luke Kohlmoos, interview with the author, March 2015.
- ¹²⁶ *Ensuring Fair and Reliable Measures of Effective Teaching*, Bill & Melinda Gates Foundation, 12.
- ¹²⁷ Whitehurst, Chingos, and Lindquist, *Evaluating Teachers with Classroom Observations*.
- ¹²⁸ Tom Kane, interview with the author, May 2015.
- ¹²⁹ Douglas N. Harris, *Value-Added Measures in Education*.
- ¹³⁰ Scott Thompson, interview with the author, March 2014.
- ¹³¹ The Aspen Institute, *Teacher Evaluation and Support Systems*, 6.
- ¹³² Author’s calculations based on District of Columbia Public Schools data.
- ¹³³ *Ensuring Fair and Reliable Measures of Effective Teaching*, Bill & Melinda Gates Foundation, 20.

BIBLIOGRAPHY

- Adnot, Melinda, Thomas Dee, Veronica Katz, and James Wyckoff. "Teacher Turnover, Teacher Quality, and Student Achievement in DCPS." National Bureau of Economic Research. Working Paper no. 21922 (January 2016).
- Aldeman, Chad, and Carolyn Chuong. *Teacher Evaluations in an Era of Rapid Change: From 'Unsatisfactory' to 'Needs Improvement.'* Washington, DC: Bellwether Education Partners, 2014. http://bellwethereducation.org/sites/default/files/Bellwether_TeacherEval_Final_Web.pdf.
- American Institutes for Research. *Measuring Principal Performance: How Rigorous Are Commonly Used Principal Performance Assessment Instruments?* Issue brief. January 2012. Quality School Leadership.
- _____. *Measuring School Climate for Gauging Principal Performance: A Review of the Validity and Reliability of Publicly Accessible Measures.* Issue brief. April 2012. Quality School Leadership.
- _____. *The Ripple Effect: A Synthesis of Research on Principal Influence to Inform Performance Evaluation Design.* Issue brief. May 2012. Quality School Leadership.
- Barnum, Matt. "The War Over Evaluating Teachers—Where it Went Right and How it Went Wrong." *The 74*, November 3, 2015. <https://www.the74million.org/article/the-war-over-evaluating-teachers-where-it-went-right-and-how-it-went-wrong>.
- Brasher, Joan. "Tennessean: TN teachers happier with evaluations; testing a burden." Vanderbilt University: Research News at Vanderbilt, August 27, 2015. <http://news.vanderbilt.edu/2015/08/tennessean-tn-teachers-happier-with-evaluations-testing-a-burden/>.
- Brown, Catherine, Lisette Partelow, and Annette Konoske-Graf. *Educator Evaluation: A Case Study of Massachusetts' Approach.* Washington, DC: Center for American Progress, 2016.
- Buzick, Heather M., and Nathan D. Jones. "Using Test Scores from Students with Disabilities in Teacher Evaluation." *Educational Measurement: Issues and Practice* 34, no. 3 (2015): 28-38.
- Camera, Lauren. "States Didn't Spend Big On New Teacher Evaluations." *U.S. News & World Report*, October 28, 2015. <http://www.usnews.com/news/articles/2015-12-21/states-didnt-invest-federal-education-dollars-on-new-teacher-evaluations>.
- Cerrone, Chris. "Education reform must move in a different direction." *The Buffalo News*, January 10, 2016. <http://www.buffalonews.com/opinion/viewpoints/education-reform-must-move-in-a-different-direction-20160110>.
- Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review* 104, no. 9 (September 2014): 2593-2632.
- _____. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review* 104, no. 9 (September 2014): 2633-2679.
- Cincinnati Public Schools. "Teacher Evaluation." <http://www.cps-k12.org/about-cps/employment/tes>.
- Colorado Department of Education. *Colorado State Model Educator Evaluation System: Practical Ideas for Evaluating Teachers of the Arts.* Colorado Department of Education, 2015. <https://www.cde.state.co.us/educatoreffectiveness/practicalideasthearts>.
- Connecticut State Department of Education. *A Guide to the BEST Program for Beginning Teachers, 2006–2007.* Hartford, CT: Connecticut State Department of Education, 2007.
- _____. *Portfolio Performance Results, Five Year Report, 1999–2004.* Hartford, CT: Connecticut State Department of Education, 2005.
- Connally, Kaylan and Melissa Tooley. *Beyond Ratings: Re-envisioning State Teacher Evaluation Systems as Tools for Professional Growth.* Washington, DC: New America, 2016.
- Council of Chief State School Officers. *Principles for Teacher Support and Evaluation Systems.* Washington, DC: Council of Chief State School Officers, 2016.
- Culver, Christina E., and Kathleen T. Hayes. *Teacher evaluations and local flexibility: Burden or benefit?* School Improvement Network, November 2013. <http://www.schoolimprovement.com/pdf/teacher-evaluations-and-local-flexibility.pdf>.
- Curtis, Rachel, and Ross Wiener. *Means to an End: A Guide to Developing Teacher Evaluation Systems that Support Growth and Development.* Washington, DC: The Aspen Institute, 2012.
- Danielson, Charlotte. *Enhancing Professional Practice: A Framework for Teaching*, 2nd ed. Alexandria, VA: Association for Supervision and Curriculum Development, 2007.
- Danielson, Charlotte, and Thomas L. McGreal. *Teacher Evaluation to Enhance Professional Practice.* Alexandria, VA: Association for Supervision and Curriculum Development, 2007.
- Darling-Hammond, Linda, and Cynthia D. Prince. *Executive Summary: Strengthening Teacher Quality in High-Need Schools: Policy and Practice.* Washington, DC: Council of Chief State School Officers, 2007.
- Dee, Thomas S., and James Wyckoff. "Incentives, Selection and Teacher Performance: Evidence from IMPACT." *Journal of Policy Analysis and Management* 34, no. 2 (Spring 2015): 267-97. <http://onlinelibrary.wiley.com/doi/10.1002/pam.21818/abstract>.
- Delaware Department of Education. "Commendations & Recommendations: A Report on Educator Evaluation in Delaware." PowerPoint. November 2015. http://www.doe.k12.de.us/cms/lib09/DE01922744/Centricity/Domain/355/Educator_Evaluation_Commendation_Expectations_Report_November_2015.pdf.
- Doherty, Kathryn M., and Sandi Jacobs. *State of the States 2013: Connect the Dots: Using evaluations of teacher effectiveness to inform policy and practice.* Washington, DC: National Council on Teacher Quality, 2013. http://www.nctq.org/dmsView/State_of_the_States_2013_Using_Teacher_Evaluations_NCTQ_Report.

BIBLIOGRAPHY *continued*

- _____. *State of the States 2015: Evaluating Teaching, Leading, and Learning*. Washington, DC: National Council on Teacher Quality, 2015. <http://www.nctq.org/dmsView/StateofStates2015>.
- Donaldson, Morgaen L. *So Long, Lake Wobegon? Using Teacher Evaluation to Raise Teacher Quality*. Washington, DC: Center for American Progress, 2009.
- Donaldson, Morgaen L., and John P. Papay. "An Idea Whose Time Had Come: Negotiating Teacher Evaluation Reform in New Haven Connecticut." *American Journal of Education* 122, no.1 (November 2015): 35-70.
- Education First. *Classroom Observations: Measuring Teachers' Professional Practices*. Education First, November 2014.
- Emma, Caitlin. "Duncan: States get delay on test results in teacher evals." *PoliticoPro*, August 21, 2014. <https://www.politicopro.com/story/education/?id=37586>.
- Ferguson, Ronald, F. "Can student surveys measure teaching quality?" *Phi Delta Kappan* 94, no. 3 (2012): 24-28.
- Gandha, Tysza. *Arkansas TESS & LEADS Focus Group Report*. Atlanta, GA: Southern Regional Education Board, 2015. <http://edboard.arkansas.gov/AttachmentViewer.aspx?AttachmentID=6297&ItemID=4328>.
- Getting Teacher Evaluation Right: A Brief for Policymakers*. Issue Brief. 2011. American Education Research Association and National Academy of Education. <https://edpolicy.stanford.edu/sites/default/files/publications/getting-teacher-evaluation-right-challenge-policy-makers.pdf>.
- Gorski, Eric. "Colorado's pick for education commissioner on Common Core, teacher evaluations and turnaround." *Chalkbeat Colorado*, December 15, 2015. <http://co.chalkbeat.org/2015/12/15/colorados-pick-for-education-commissioner-on-common-core-teacher-evaluations-and-turnaround/#.VpAp3JMrU>.
- Hamill, Sean D. *Forging a New Partnership: The Story of Teacher Union and School District Collaboration in Pittsburgh*. Washington, DC: The Aspen Institute, 2011.
- Headden, Susan. *Inside IMPACT: DC's Model Teacher Evaluation System*. Washington, DC: Education Sector, 2011. http://educationpolicy.air.org/sites/default/files/publications/IMPACT_Report_RELEASE.pdf.
- Heneman, H.G. III, Anthony Milanowski, Steven M. Kimball, and Allan Odden. *Standards-Based Teacher Evaluation as a Foundation for Knowledge- and Skill-Based Pay*. CPRE Policy Brief RB-45. May 2006. Philadelphia, PA: Consortium for Policy Research in Education.
- Hershberg, Ted. "Value-Added Assessment and Systemic Reform: A Response to America's Human Capital Development Challenge." Paper prepared for the Aspen Institute's Congressional Institute, Cancun, Mexico, February 22-27, 2005.
- Kane, Thomas J., J.E. Rockoff, and Douglas O. Staiger. *What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City*. Cambridge, MA: National Bureau of Economic Research, 2006.
- Karlin, Rick. "With moratorium in place, teachers offer evaluation design ideas." *Times Union*, December 16, 2015. <http://www.timesunion.com/local/article/With-moratorium-in-place-teachers-offer-6701087.php>.
- Kraft, Matthew A., and Allison F. Gilmour. "Revisiting the Widget Effect: Teacher Evaluation Reforms and the Distribution of Teacher Effectiveness." Brown University Working Paper, February 2016.
- LeBuhn, Mac. *Culture of Countenance: Teachers, Observers and the Effort to Reform Teacher Evaluations*. Democrats for Education Reform, May 2013.
- Long, Daniel A., and Jessica K. Beaver. *Delaware Performance Appraisal System Second Edition (DPAS-II): Evaluation Report*. Prepared for the Delaware Department of Education by Research for Action, November 2015. <http://www.doe.k12.de.us/cms/lib09/DE01922744/Centricity/Domain/375/RFA%20Evaluation%20of%20DPAS-II%2011.1.2015.pdf>.
- Louisiana Principals' Teaching & Learning Guidebook: A Path to High-Quality Instruction in Every Classroom*. Louisiana Department of Education, 2015. <http://www.louisianabelieves.com/docs/default-source/teacher-toolbox-resources/2015-louisiana-principals%27-teaching-learning-guidebook.pdf?sfvrsn=2>.
- Maryland State Department of Education. "Presentation to the Maryland State Board of Education: Descriptive Analysis of School Year 2014-2015 Teacher and Principal Effectiveness Ratings." PowerPoint. October 27, 2015. <http://marylandpublicschools.org/MSDE/programs/tpe/docs/Analysis2014-15TeacherPrincipalEffectivenessRatings.pdf>.
- McCaffrey, Daniel F., and J.R. Lockwood. "Missing Data in Value-Added Modeling of Teacher Effects." *The Annals of Applied Statistics* 5, no. 2A (2011): 773-97.
- McGuinn, Patrick. *Evaluating Progress: State Education Agencies and the Implementation of New Teacher Evaluation Systems*. White Paper (#WP2015-09). Philadelphia: Consortium for Policy Research in Education, University of Pennsylvania, 2015. http://www.cpre.org/sites/default/files/policybrief/2061_pbmccguinn2.pdf.
- McKay, Sarah, and Elena Silva. *Improving Observer Training: The Trends and Challenges*. Issue brief. September 2015. Carnegie Foundation for the Advancement of Teaching. http://cdn.carnegiefoundation.org/wp-content/uploads/2015/09/BRIEF_Improving_Observer_Training.pdf.
- Moore, Lynn. "Without guidelines for Michigan teacher evaluations, everyone is awesome." *mLive Michigan*, December 17, 2015. http://www.mlive.com/news/index.ssf/2015/12/unlikely_allies_behind_teacher.html.
- Morello, Rachel. "Teachers See Evaluation System Less Favorably Than Administrators." *Indiana State Impact*, December 1, 2014. <http://indianapublicmedia.org/stateimpact/2014/12/01/indiana-educators-differ-views-teacher-evaluation-system/>.
- National Institute for Excellence in Teaching. *Creating a Successful Performance Compensation System for Educators*. Washington, DC: National Institute for Excellence in Teaching, 2007.

BIBLIOGRAPHY *continued*

Neason, Alexandria, and Meredith Kolodner. “DC’s lessons for New York on teacher evaluations.” *Politico New York*, April 9, 2015. <http://www.capitalnewyork.com/article/albany/2015/04/8565615/dcs-lessons-new-york-teacher-evaluations>.

Neufeld, Sara. “Will New Teacher Evaluations Help or Hurt Chicago’s Schools?” *The Atlantic*, April 30, 2013. <http://www.theatlantic.com/national/archive/2013/04/will-new-teacher-evaluations-help-or-hurt-chicago-schools/275415/>.

North Carolina State Board of Education. “Teacher Evaluation School Year 2014-15.” PowerPoint. Public Schools of North Carolina, January 6, 2016. https://eboard.eboardsolutions.com/meetings/TempFolder/Meetings/PowerPoint%20Presentation_49122aqiadv1xyqzuzrizmsfgaiwx.pdf.

OECD. *Teachers for the 21st Century: Using Evaluations to Improve Teaching*. OECD Publishing, 2013. <http://www.oecd.org/site/eduistp13/TS2013%20Background%20Report.pdf>.

Papay, John P., Andrew Bacher-Hicks, Lindsay C. Page, and William H. Marinell. “The Challenge of Teacher Retention in Urban Schools: Evidence of Variation from a Cross-Site Analysis.” Working Paper (May 18, 2015). Available at SSRN: <http://ssrn.com/abstract=2607776>.

Papay, John P., Eric S. Taylor, John H. Tyler, and Mary Laski. “Learning Job Skills from Colleagues at Work: Evidence from a Field Experiment Using Teacher Performance Data.” National Bureau of Economic Research. Working Paper no. 21986 (February 2016).

Parrish, Frances. “Teacher evaluation changes a possibility in the future.” *Independent Mail*, December 28, 2015. <http://www.independentmail.com/news/teacher-evaluation-changes-a-possibility-in-future-27fb57b7-681b-0278-e053-010007f4438-363674491.html>.

Pennington, Kaitlin. “The Peril of Teacher Evaluation Policy under ESSA.” *Ahead of the Heard*, December 4, 2015. Bellwether Education Partners. <http://aheadoftheheard.org/the-peril-of-teacher-evaluation-policy-under-essa/>.

Phillips, Vicki, and Randi Weingarten. “Six Steps to Effective Teacher Development and Evaluation.” *New Republic*, March 25, 2013. <https://newrepublic.com/article/112746/gates-foundation-sponsored-effective-teaching>.

Podgursky, Michael J., and Matthew G. Springer. *Teacher Performance Pay: A Review*. (Nashville, TN: National Center on Performance Incentives, 2006).

Porter, Eduardo. “Grading Teachers by the Test.” *The New York Times*, March 24, 2015. <http://nyti.ms/1BhHQJ>.

Pratt, Tony. *Making Every Observation Meaningful: Addressing Lack of Variation in Teacher Evaluation Ratings*. Policy Brief. Nashville, TN: Tennessee Department of Education, Office of Research and Policy, November 2014.

Robles, Yesenia. “Evolution of teacher evaluations is leading performance pay reforms.” *The Denver Post*, July 28, 2015. http://www.denverpost.com/news/ci_28547009/evolution-teacher-evaluations-is-leading-performance-pay-reforms.

Roen, Paul. “Carving a Place for Blended Learning in the Era of Teacher Evaluation.” *Getting Smart*, July 13, 2013. <http://gettingsmart.com/2013/07/carving-a-place-for-blended-learning-in-the-era-of-teacher-evaluation/>.

Sawchuk, Stephen. “ESSA Loosens Reins on Teacher Evaluations, Qualifications.” *Education Week*, January 5, 2016. <http://www.edweek.org/ew/articles/2016/01/06/essa-loosens-reins-on-teacher-evaluations-qualifications.html>.

_____. “Florida Releases ‘Value Added’ Data on Teachers.” *Education Week*, February 24, 2014. http://blogs.edweek.org/edweek/teacherbeat/2014/02/florida_releases_teacher_value.html.

_____. “Teacher Evaluation Heads to the Courts.” *Education Week*, October 6, 2015. <http://www.edweek.org/ew/section/multimedia/teacher-evaluation-heads-to-the-courts.html>.

_____. “Teacher Evaluation Lawsuit in Tennessee Nixed by Federal Judge.” *Education Week*, February 22, 2016. http://blogs.edweek.org/edweek/teacherbeat/2016/02/teacher-evaluation_lawsuit_in_.html.

_____. “New York Panel Eliminates Tests From Teacher Evaluations, For Now.” *Education Week*, December 15, 2015. http://blogs.edweek.org/edweek/teacherbeat/2015/12/new_york_panel_eliminate_test.html.

Schulz, Jeff, Gunjan Sud, and Becky Crowe. *Lessons from the Field: The Role of Student Surveys in Teacher Evaluation and Development*. Washington, DC: Bellwether Education Partners, 2014. http://bellwethereducation.org/sites/default/files/Bellwether_StudentSurvey.pdf.

Schochet, Peter Z., and Hanley S. Chiang. *Error Rates in Measuring Teacher and School Performance Based on Student Test Score Gains*. NCEE 2010-4004. US Department of Education, July 2010. <http://ies.ed.gov/ncee/pubs/20104004/pdf/20104004.pdf>.

Shearer, Lee. “Mandated tests, evaluations driving Georgia teachers from profession, survey shows.” *The Florida Times-Union*, January 8, 2016. <http://jacksonville.com/news/georgia/2016-01-08/story/mandated-tests-evaluations-driving-georgia-teachers-profession-survey>.

Silva, Elena. *Teachers At Work: Improving Teacher Quality Through School Design*. Education Sector Reports, October 2009.

Slotnik, William J., Daniel Bugler, and Guodong Liang. *Change in Practice in Maryland: Student Learning Objectives and Teacher and Principal Evaluation*. Washington, DC: Mid-Atlantic Comprehensive Center, 2015. <http://marylandpublicschools.org/tpe/TPEReport2015.pdf>.

Solochek, Jeffrey S. “Florida education department increases emphasis on value-added model in teacher evaluations.” *Tampa Bay Times*, January 4, 2016. <http://www.tampabay.com/blogs/gradebook/florida-education-department-increases-emphasis-on-value-added-model-in/2259922>.

BIBLIOGRAPHY *continued*

- Sporte, Susan E., W. David Stevens, Kaleen Healey, Jennie Jiang, and Holly Hart. "Teacher Evaluation in Practice: Implementing Chicago's REACH Students." Chicago: University of Chicago, Consortium on Chicago School Research, 2013. https://consortium.uchicago.edu/sites/default/files/publications/REACH_Report_0.pdf.
- Stecher, Brian, Mike Garet, Deborah Holtzman, and Laura Hamilton. "Implementing measures of teacher effectiveness." *Phi Delta Kappan* 94, no. 3 (2012): 39-43.
- Steinberg, Matthew P., and Morgaen L. Donaldson. "The New Educational Accountability: Understanding the Landscape of Teacher Evaluation in the Post-NCLB Era." *Education Finance and Policy*. (Forthcoming). http://cepa.uconn.edu/wp-content/uploads/sites/399/2014/02/The-New-Educational-Accountability_policy-brief_8-19-14.pdf.
- Surowiecki, James. "Better All the Time." *The New Yorker*, November 10, 2014. <http://www.newyorker.com/magazine/2014/11/10/better-time>.
- Tatter, Grace. "As Tennessee finishes its Race to the Top, teachers caught in the middle of competing changes." *Chalkbeat Tennessee*, December 15, 2015. http://tn.chalkbeat.org/2015/12/15/as-tennessee-finishes-its-race-to-the-top-teachers-caught-in-the-middle-of-competing-changes/#.VpAqCZMrI_U.
- _____. "Tennessee's teacher evaluation system improving, state report says." *Chalkbeat Tennessee*, April 9, 2015. <http://tn.chalkbeat.org/2015/04/09/tennessees-teacher-evaluation-system-improving-state-report-says/#.VnmyxP3SnCZ>.
- Taylor, Eric S., and John H. Tyler. "The Effect of Evaluation on Teacher Performance," *American Economic Review* 2012, 102(7): 3628-3651.
- Taylor, Kate. "New York Regents Vote to Exclude State Tests in Teacher Evaluations." *The New York Times*, December 14, 2015. http://www.nytimes.com/2015/12/15/nyregion/new-york-regents-vote-to-exclude-state-tests-in-teacher-evaluations.html?_r=0.
- Tennessee Department of Education. *Teacher and Administrator Evaluation in Tennessee: A Report on Year 3 Implementation*. Nashville, TN: Tennessee Department of Education, 2015. http://team-tn.org/wp-content/uploads/2013/08/rpt_teacher_evaluation_year_31.pdf.
- _____. "Tennessee Educator Survey." 2015. <http://tndoe.azurewebsites.net/>.
- "Test scores in teacher evaluations to move forward in Nevada." *CBS Las Vegas*, December 28, 2015. <http://lasvegas.cbslocal.com/2015/12/28/teacher-evaluation-test-scores-moving-forward-in-nevada/>.
- The Aspen Institute. *Teacher Evaluation and Support Systems: A Roadmap for Improvement*. Washington, DC: The Aspen Institute, 2016.
- The New Teacher Project. *The Irreplaceables: Understanding the Real Retention Crisis in America's Urban Schools*. Washington, DC: The New Teacher Project, 2012.
- _____. *The Mirage: Confronting the Hard Truth About Our Quest for Teacher Development*. Washington, DC: The New Teacher Project, 2015. http://tntp.org/assets/documents/TNTP-Mirage_2015.pdf.
- The Teaching Commission. *Teaching at Risk: A Call to Action*. New York: The Teaching Commission, 2004.
- _____. *Teaching at Risk: Progress and Potholes*. New York: The Teaching Commission, 2006.
- Thomsen, Jennifer. "Teacher performance plays growing role in employment decisions." Education Commission of the States. *Teacher Tenure: Trends in State Laws*. May 2014. <http://www.ecs.org/clearinghouse/01/12/42/11242.pdf>.
- Toch, Thomas, and Robert Rothman. *Rush to Judgment: Teacher Evaluation in Public Education*. Washington, DC: Education Sector, 2008.
- Tyler, John H. *Designing High Quality Evaluation Systems for High School Teachers: Challenges and Potential Solutions*. Washington, DC: Center for American Progress, 2011.
- Watts, Lisa, Joy Singleton-Stevens, and Mark Teoh. *Lessons from the Leading Edge: Teachers' Views on the Impact of Evaluation Reform*. Teach Plus, June 2013. http://www.teachplus.org/sites/default/files/publication/pdf/lessons_from_the_leading_edge.pdf.
- Weiner, Ross, and Ariel Jacobs. *Designing and Implementing Teacher Performance Management Systems: Pitfalls and Possibilities*. Washington, DC: The Aspen Institute, 2011.
- Weisberg, Daniel, Susan Sexton, Jennifer Mulhern, and David Keeling. *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness*. 2nd ed. Washington, DC: The New Teacher Project, 2009.
- White, Taylor. *Adding Eyes: The Rise, Rewards, and Risks of Multi-Rater Teacher Observation Systems*. Issue brief. December 2014. Carnegie Foundation for the Advancement of Teaching. http://cdn.carnegiefoundation.org/wp-content/uploads/2014/12/BRIEF_Multi-rater_evaluation_Dec2014.pdf.
- Wilson, Mark, P.J. Hallman, Ray Pecheone, and Pamela Moss. "Using Student Achievement Test Scores as Evidence of External Validity for Indicators of Teacher Quality: Connecticut's Beginning Educator Support and Training Program." Unpublished paper, October 2007.
- Youngs, Peter. "How Elementary Principals' Beliefs and Actions Influence New Teachers' Experiences." *Education Administration Quarterly* 43, no. 1 (2007): 101-37.

Acknowledgements

This report was generously funded by the Joyce Foundation.

I am immensely grateful to the many educators and education policymakers who contributed information and their considerable insights to this report and who, in several instances, generously read draft sections of the report.

Thanks also to Katherine Parham for her extensive research assistance, and to Jackie Arthur, Molly Breen, Kristen Garabedian, and Lisa Hansel for their many editorial contributions.

The conclusions in the report, however, are the author's alone, as are any errors of fact or interpretation.

© 2016 Thomas Toch

The non-commercial use, reproduction, and distribution of this report is permitted.

Center on the Future of
AMERICAN
EDUCATION
at Georgetown University

GRADING THE GRADERS:
A REPORT ON TEACHER EVALUATION
REFORM IN PUBLIC EDUCATION