

AUGUST 2014

Teacher Evaluations in an Era of Rapid Change:





FROM “UNSATISFACTORY” TO “NEEDS IMPROVEMENT”

Chad Aldeman and Carolyn Chuong



BELLWETHER
EDUCATION PARTNERS
IDEAS | PEOPLE | RESULTS

© 2014 Bellwether Education Partners

-  This report carries a Creative Commons license, which permits noncommercial re-use of content when proper attribution is provided. This means you are free to copy, display and distribute this work, or include content from this report in derivative works, under the following conditions:
-  Attribution. You must clearly attribute the work to Bellwether Education Partners, and provide a link back to the publication at <http://bellwethereducation.org/>.
-  Noncommercial. You may not use this work for commercial purposes without explicit prior permission from Bellwether Education Partners.
-  Share Alike. If you alter, transform, or build upon this work, you may distribute the resulting work only under a license identical to this one.

For the full legal code of this Creative Commons license, please visit www.creativecommons.org. If you have any questions about citing or reusing Bellwether Education Partners content, please contact us.

AUGUST 2014

Teacher Evaluations in an Era of Rapid Change:

FROM “UNSATISFACTORY” TO “NEEDS IMPROVEMENT”

Chad Aldeman and Carolyn Chuong



BELLWETHER
EDUCATION PARTNERS
IDEAS | PEOPLE | RESULTS

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	i
EXECUTIVE SUMMARY.....	iii
THE WIDGET EFFECT: A DEMAND FOR BETTER EVALUATIONS	1
DISTRICTS ARE STARTING TO EVALUATE TEACHERS AS PROFESSIONALS, RATHER THAN AS INTERCHANGEABLE WIDGETS.	3
SCHOOLS ARE PROVIDING TEACHERS WITH BETTER, TIMELIER FEEDBACK ON THEIR PRACTICE.	6
DESPITE STATE POLICY CHANGES, DISTRICTS STILL DON'T FACTOR STUDENT GROWTH INTO TEACHER EVALUATION RATINGS..	9
DISTRICTS HAVE WIDE DISCRETION EVEN UNDER "STATEWIDE" EVALUATION SYSTEMS ...	16
DISTRICTS CONTINUE TO IGNORE PERFORMANCE WHEN MAKING DECISIONS ABOUT TEACHERS.....	20
LESSONS FOR POLICYMAKERS.....	24
CONCLUSION	27
ENDNOTES	28
APPENDIX	31

ACKNOWLEDGEMENTS

The authors gratefully thank all those who offered generous feedback on earlier drafts of this paper, especially Larry Abramson, Tim Daly, Brad Jupp, and Bryan Hassel. Thanks also to our colleagues at Bellwether, in particular Sara Mead, Andrew Rotherham, and Ashley LiBetti Mitchel, for their comments and suggestions on earlier drafts. Thanks particularly to Leslie Kan for managing the editorial process of this paper, and to Sandy Fleishman and Noelle Grattan for copy-editing and design support. The views and analysis in this report are the responsibility of the authors alone.

This publication is based on research funded in part by the Bill & Melinda Gates Foundation. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation.

ABOUT THE AUTHORS

Chad Aldeman is an Associate Partner on the Policy and Thought Leadership team at Bellwether Education Partners. He can be reached at chad.aldeman@bellwethereducation.org. Carolyn Chuong is an Analyst at Bellwether Education Partners. She can be reached at carolyn.chuong@bellwethereducation.org.

ABOUT BELLWETHER EDUCATION PARTNERS



Bellwether Education Partners is a nonprofit dedicated to helping education organizations—in the public, private, and nonprofit sectors—become more effective in their work and achieve dramatic results, especially for high-need students. To do this, we provide a unique combination of exceptional thinking, talent, and hands-on strategic support.

EXECUTIVE SUMMARY

In a relatively brief spurt in 2010 and 2011, something remarkable happened to teacher evaluation. States began requiring school districts to use multilevel rating evaluation systems rather than black-or-white determinations that had long classified educators as either “satisfactory” or not. They required districts to incorporate student academic growth and high-quality evaluation rubrics into their ratings of teachers and principals. And they strengthened the potential consequences educators faced based on their evaluations.¹ Although it had long been known that teachers play a key role in increasing student achievement, states chose this moment to make significant changes.

But after this initial rush of reforms, progress stalled. The rollout of new evaluation systems slowed down as the action shifted from the policy realm to the more laborious implementation phase. Challenges have arisen in measuring student growth and providing educators the actionable feedback they need.

We set out to examine what can be learned so far from the now four-year-old effort to revamp teacher evaluations. To examine state implementation, we collected and synthesized data from the 17 states and the District of Columbia that have tracked and reported information on their evaluation efforts. This report sorts the data into five major lessons:

1. **Districts are starting to evaluate teachers as professionals, rather than as interchangeable widgets:** Districts have made substantial progress in differentiating between poor, fair, and great educator performance. Where once they saw only black or white—satisfactory or unsatisfactory performance—schools are starting to see nuance through the use of multi-tiered evaluation systems.
2. **Schools are providing teachers with better, timelier feedback on their practice:** With the use of higher-quality classroom observation rubrics, teachers are receiving more frequent observations and more detailed feedback on how they’re doing.
3. **Despite state policy changes, districts still don’t factor student growth into teacher evaluation ratings:** Many new evaluation systems fail to accurately reflect student academic progress and continue to mask poor educator performance.
4. **Districts have wide discretion even under “statewide” evaluation systems:** State policymakers, while providing broad guidelines and allowable uses, give districts ultimate discretion on how to implement evaluation policies. Local autonomy means that evaluation systems within the same state may look very different from one another.
5. **Districts continue to ignore performance when making decisions about teachers:** School districts across the country rarely use the results of educator evaluations to make consequential decisions around hiring, compensation, professional development, tenure, and dismissal. Fears that new evaluation systems would lead to mass dismissals of teachers do not appear to have any basis in reality.

New evaluation systems are just one part of sweeping changes in American schools. Over the next few years, 44 states and the District of Columbia will implement new college- and career-ready standards and assessments aligned to those standards. Because the number and extent of the changes are daunting, some states have already started amending or postponing their teacher evaluation systems. But evaluation reform is an effort worth making. Evidence from Cincinnati,² Washington, D.C.,³ and Denver⁴ suggests that comprehensive evaluation systems help teachers improve their practice, lead to improved recruitment and retention of high-quality educators, and, ultimately, boost student achievement. While caution ahead of large-scale change is understandable, policymakers serious about improving our nation’s schools shouldn’t roll back recent evaluation reforms before the new policies can even begin to take effect.

THE WIDGET EFFECT: A DEMAND FOR BETTER EVALUATIONS

The now-familiar “widget effect” describes schools’ practice of treating teachers like interchangeable parts. TNTP (formerly The New Teacher Project), a nonprofit that works to improve hiring practices in urban districts, coined the term in a 2009 report of the same name. After reviewing 12 evaluation systems across four states, they found three common problems:

1. Most districts rated school employees as simply satisfactory or not, with nothing above or between.
2. Only a very small percentage of teachers received unsatisfactory ratings.
3. Districts by and large did not use the evaluation results to make critical personnel decisions.⁵

A 2012 Education Sector report replicated and extended *The Widget Effect’s* findings to the entire state of Washington. Across the Evergreen State, districts’ failure to acknowledge and act on differences in performance extended beyond teachers to principals, superintendents, and school support staff (e.g., janitors and librarians). School districts in Washington identified only a miniscule number of employees as unsatisfactory: 0.92 percent of teachers, 1.42 percent of principals, 1.02 percent of superintendents, and 2.1 percent of school support staff. Across the state, 85 percent of schools failed to identify a single low-performing teacher. Nine out of 10 districts did not identify a single low-performing principal leading their schools.⁶

The Widget Effect catalyzed the federal government, states, and districts to push for evaluation systems that consider student growth, and to use evaluation results to reward great teachers, dismiss poor ones, and give struggling but promising teachers the support they need. In response, states made a number of changes to teacher and principal evaluations. According to the National Council on Teacher Quality, two-thirds of states adopted new ways to evaluate teachers between 2009 and 2012.⁷ As of 2014, 16 states were in the process of implementing new evaluation systems that include student growth. By 2015, more than 40 states plan to include some objective measure of student achievement in teacher evaluations. In 2009 no state required districts to consider student learning when deciding on teacher tenure, but by 2012, 16 states required districts to do so.⁸

In turn, advocates and policymakers pushed for more transparency on the evaluation systems themselves—what criteria were used, what the categories represented, and how educators were distributed along the performance spectrum. The U.S. Department of Education required all states—in exchange for their share of the \$53.6 billion State Fiscal Stabilization Fund (enacted as part of the 2009 American Recovery and Reinvestment Act) to collect information on district educator evaluation systems. Every single state took the money and agreed to track whether evaluations included student achievement or growth and the number and percentage of teachers within each performance level. States were also asked to report on how districts evaluated principals, and to make all this information public for each school.

State leaders and advocates have since clamored for more information about the implementation of evaluation systems. To date, we count 17 states and the District of Columbia that have released data that track the results of evaluation systems (see Appendix for the full list of states). Based on this data, we identified five major lessons on teacher evaluation reforms.

1. DISTRICTS ARE STARTING TO EVALUATE TEACHERS AS PROFESSIONALS, RATHER THAN AS INTERCHANGEABLE WIDGETS

The most dramatic finding from *The Widget Effect* was that districts were using binary, either/or evaluation systems that saw only black or white: Educators were either “satisfactory” or they weren’t, and the vast majority received a perfunctory “satisfactory” stamp of approval. That dichotomy is fading away under the new evaluation systems. States now have multi-tiered evaluation systems that place educators into four or five categories of performance.

Some observers of the new evaluation systems focus on the number of educators in the lowest category, pointing out that few educators are rated ineffective, just like under the old evaluation systems. These observers often use Georgia as an example; among 5,800 teachers participating in a pilot evaluation in 2012, less than 1 percent were rated as ineffective.

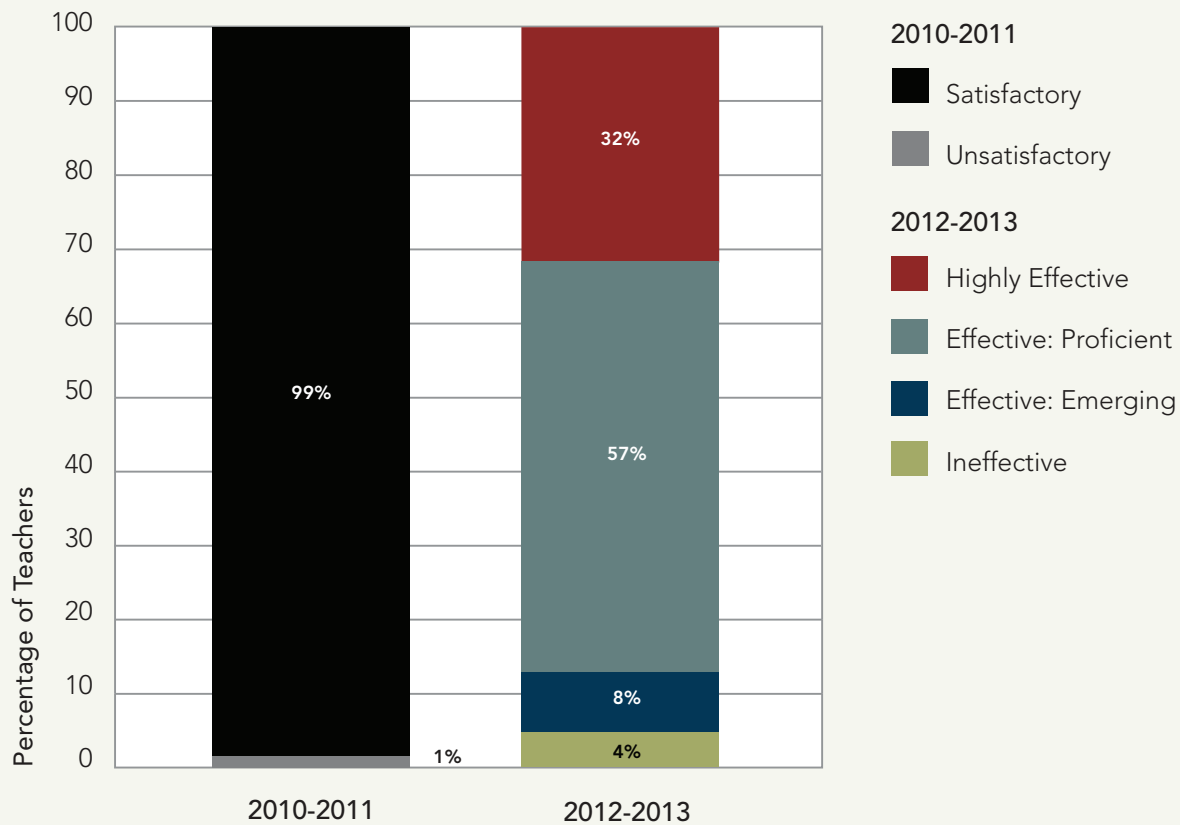
***The Widget Effect* found that districts were using binary, either/or evaluation systems that saw only black or white. That dichotomy is fading away under the new evaluation systems.**

But focusing only on the bottom tier misses two key issues. First, even though only a small percentage of teachers are identified as the lowest-performing, the absolute number can be significant. After Louisiana implemented a new evaluation system in 2012-13, 4 percent of teachers received the lowest performance rating, now called “Ineffective” (see Figure 1). Four percent may still not seem like much, but in a state with 50,000 educators, it represents 2,000 people put

on notice that they need to improve. Recent research from New York City suggests that merely identifying low-performing teachers and delaying decisions on their tenure can encourage weaker teachers to leave.⁹

Second, there are more than just two categories in the new evaluation systems. In Georgia, for example, rather than having nearly all teachers receiving the highest designation, districts participating in the state’s evaluation system pilot effort identified only one in five teachers with the highest rating of “Exemplary.”¹⁰ In Louisiana, instead of labeling 99 percent of teachers as “Satisfactory,” districts identified 32 percent of its teachers as top performers, or “Highly Effective,” under the new evaluation system.¹¹

FIGURE 1: LOUISIANA'S NEW EVALUATION SYSTEM, COMPASS, NOW CLASSIFIES TEACHERS ON A FOUR-TIER SCALE



Note: Percentages may not total 100% due to rounding.

Source: Louisiana Department of Education, “2012-2013 Compass Final Report,” <https://www.louisianabelieves.com/teaching/compass-final-report-2012-2013>.

The new terminology is a step in the right direction, too. Terms used in today's evaluation systems are much more descriptive. Under pre-reform systems, "Satisfactory" could be applied to teachers who were extraordinary or middling without any differentiation. Designations like "Exemplary" or "Highly Effective" used in many of the new systems today are much stronger terms reserved for true exceptionality.

Besides creating a broader spectrum for how educators are classified, evaluation systems that create greater differentiation can identify educators who need more targeted support. Louisiana has identified 8 percent of its teachers as "Effective: Emerging," which means they're not quite ineffective but they're not quite proficient either. These teachers aren't performing so poorly that they should be dismissed, but they do need more supervision and support than others. Such targeted support would have been nearly impossible under the state's old evaluation system.

Besides Louisiana, no state that we're aware of has released data that allow direct comparisons between old and new evaluation systems.¹² Other states and districts are making progress, although the pace of implementing meaningful reform varies.

2. SCHOOLS ARE PROVIDING TEACHERS WITH BETTER, TIMELIER FEEDBACK ON THEIR PRACTICE

With most of the attention on teacher evaluation reform focusing on student growth, it's easy to forget about the importance of actually observing teacher and principal performance. But qualitative data offer valuable feedback on teacher performance, helping educators make real-time changes to their instructional practices. A teacher's overall evaluation rating cannot help a teacher improve unless it is accompanied by formative, actionable feedback based on observations of practice.

As a practical matter, observations also have the added benefit of being universal for all teachers and principals. Objective measures of student growth are available only for educators teaching subjects that are assessed through state- or district-wide tests. A recent study from the Brookings Institution found that only about one-fifth of teachers across four urban districts were even eligible to have part of their evaluations based on student test scores.¹³ Classroom observations, on the other hand, can be carried out for teachers of all subjects and grades. Just as “student growth” can refer to a wide variety of measures and tools, classroom observations can vary in frequency and quality.

States are beginning to mandate that schools carry out classroom observations more frequently. Between 2009 and 2013, the number of states requiring annual evaluations for all teachers increased from 15 to 28.¹⁴ For example, during the two school years that New Jersey carried out its teacher evaluation pilot, the state significantly increased the number of classroom observations conducted so that all teachers could receive more frequent appraisals of their performance and so that struggling teachers could receive

additional support. During the first year of the pilot, teachers received 1.3 observations on average, but that number increased to an average of 3.0 during the second year. Just like employees in other sectors, teachers benefit from more regular feedback from their supervisors and increased opportunities to improve their work.

Just like employees in other sectors, teachers benefit from more regular feedback from their supervisors and increased opportunities to improve their work.

Providing high-quality feedback is more challenging than simply increasing the frequency of observations. Schools have considered a number of ways to observe teachers—from simple, locally-designed checklists to more extensive protocols that require formal training for observers. In recent years more schools are trending toward the use of externally created rubrics that guide

evaluators to carry out more detailed observations of teacher practice. That change is intended to increase the quantity and depth of feedback teachers receive and boost student performance. The Measures of Effective Teaching (MET) project, for instance, found a link between teacher scores on five high-quality observation instruments and increases in student achievement.¹⁵ Evaluation reforms in recent years have encouraged the widespread adoption or modification of these observation tools. Over the last few years, Arkansas, Delaware, Florida, Idaho, Illinois, New Jersey, New York, South Dakota, and Washington, as well as cities including Cincinnati, Los Angeles, and Pittsburgh, have adopted one of the instruments analyzed in the MET Project, Charlotte Danielson's Framework for Teaching, as their preferred teacher observation rubric.¹⁶ Some states and districts continue to use their own observation tools, but many more schools are adopting research-backed protocols to observe and support teachers.

The trickiest part for states and districts has been ensuring that schools implement these higher-quality systems fairly. The 2014 Brookings Institution report on four urban districts found that teachers whose students had higher initial performance at the start of the school year received better classroom observation scores.¹⁷ In other words, the classroom observation process is biased against teachers working with students who start the year behind. A study of Pittsburgh Public Schools also found that teacher observation ratings were higher for teachers serving gifted and talented students—and lower for those serving low-income and minority students.¹⁸ These findings are troubling because excellent teachers should not be penalized for serving students from high-needs backgrounds. Even if school districts shift to using better observation tools, as Pittsburgh

has, these tools will be meaningless in a system where observers consistently give teachers inflated evaluation ratings based on a student’s starting point. Districts must have processes in place—utilizing multiple raters for one teacher or hiring external observers, for instance—to ensure that these biases do not arise during classroom observations.

Despite these implementation challenges, however, teachers and principals report positive outcomes from these efforts to improve classroom observations. In a survey of Connecticut principals participating in a new pilot evaluation system, 70-80 percent reported spending more time observing teachers, talking with teachers after the observation, and developing written feedback.¹⁹ In Delaware, 85 percent of teachers said the feedback they received during the teacher evaluation pilot was “useful and applicable.”²⁰

3. DESPITE STATE POLICY CHANGES, DISTRICTS STILL DON'T FACTOR STUDENT GROWTH INTO TEACHER EVALUATION RATINGS

More frequent and rigorous feedback for teachers is unquestionably a good thing. But linking teacher and principal evaluation ratings more closely to the academic gains of their students faces more challenges. Many states now mandate that student growth be incorporated into evaluation ratings. Student growth measures indicate how much academic progress students make over a given time period. While raw student achievement metrics are biased—in favor of students from privileged backgrounds with more educational resources—student growth measures adjust for these incoming characteristics by focusing only on knowledge acquired over the course of a school year.

As more states and districts adopt teacher evaluation policies that include student growth, questions and concerns have arisen about whether these measures of student progress adequately reflect teacher performance. This backlash has led to three types of reactions:

REFUSAL: DO NOT INCORPORATE STUDENT GROWTH INTO EVALUATION SYSTEMS

Not all states have participated in the recent wave of evaluation reforms. As of September 2013, 40 states and the District of Columbia Public Schools require that objective measures of student achievement play a role in teacher evaluations.²¹ One of the remaining states, California, has lessons for reformers.

California presents a cautionary tale of how policy changes may not manifest in changes in practice. Under a 1971 state law known as the Stull Act, all California districts are required to evaluate school employees based on students’ progress toward state academic standards.²² Many districts openly admit to defying the law. In 2011, California surveyed all of its districts on how they used their evaluation systems and whether student outcomes or growth were used to evaluate teacher and principal performance. Seven of the 10 largest school districts in California—San Diego, Fresno, Elk Grove, Santa Ana, San Francisco, San Bernardino, and Corona-Norco—reported that student outcomes or growth were not used at all to evaluate teacher performance.²³

DELAY: REQUIRE THAT “STUDENT GROWTH” BE A FACTOR IN EVALUATIONS WITHOUT DEFINING HOW IT WILL BE MEASURED OR INCORPORATED INTO RATINGS

Many states adopted new rules requiring teacher evaluations to include student growth but stopped short of defining how “student growth” should be measured. Massachusetts, for example, uses a matrix approach with student growth on one axis and professional practice on the other. Teachers receive their final evaluation ratings based on where they fall on these two categories, and the student growth component must include growth on the statewide assessment. But Massachusetts does not stipulate how much weight those assessments should have.²⁴

Other states promised to determine a process for including student growth in teacher and principal evaluations at some point in the future.²⁵ Many states have specified the overall weighting that student growth must carry in evaluations, but have not defined how it must be measured or how multiple measures of growth should add up to the total. Colorado, for instance, stipulated that 50 percent of a teacher’s evaluation be based on the sum total of student growth measures, and that student growth include results on statewide assessments. But it leaves local districts to determine the weighting for the assessment results within the overall growth component.²⁶

Many states have used no-stakes trial periods or pilots to test different ways to evaluate teacher performance and measure student growth. While implementing pilot evaluation systems may be a step in the right direction, states are unlikely to find results that are meaningfully different from the past until they include objective measures of student growth.

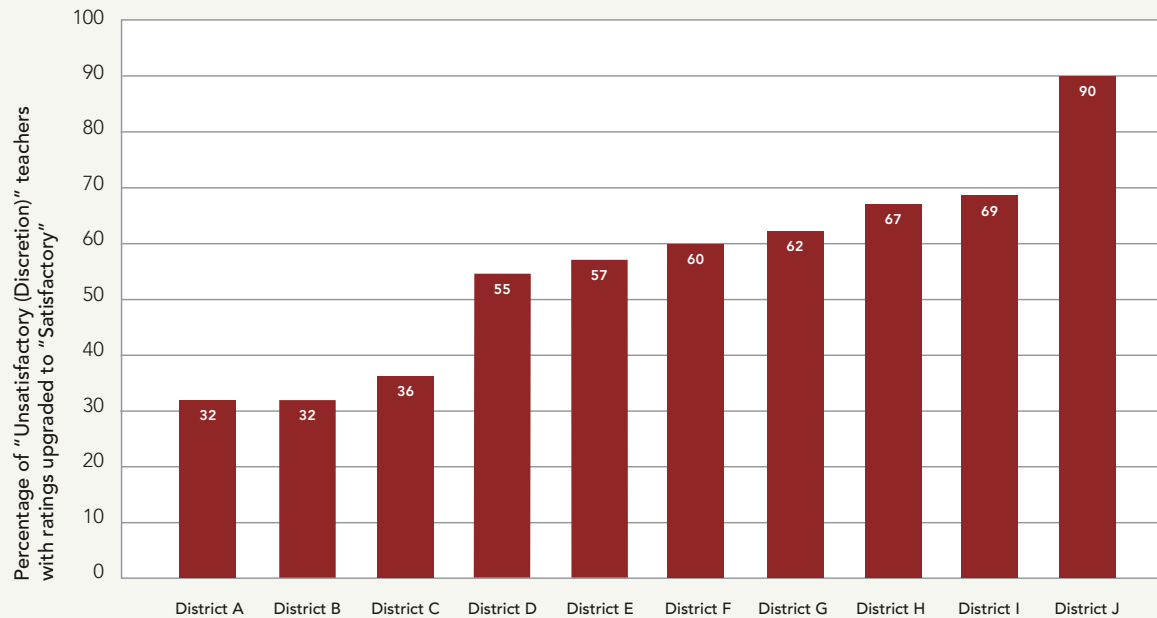
OBSCURE: GIVE LOCAL SCHOOL DISTRICTS FLEXIBILITY TO REVISE OR MAKE THEIR OWN DECISIONS, EVEN ABOUT OBJECTIVE DATA ON STUDENT GROWTH

The most common and potentially pernicious effect of the new evaluation systems is the appearance of change without actual substance. Many states adopted policies requiring student growth in teacher evaluations, and even stipulated that it play a “significant” or “predominant” factor in ratings. But the actual implementation of these reforms has been largely left to districts, and their efforts tend to downplay the impact of student growth. Even early-adopter states that paired extensive changes to their statewide evaluation systems with extensive training efforts, like Tennessee, have not seen dramatic changes in overall results. During the first year of Tennessee’s new evaluation system, 17 percent of teachers earned the lowest rating in student growth, but only 0.2 percent earned the lowest score on classroom observations.²⁷

Delaware provides another example. The state committed to using student growth and demonstrated the capacity to measure growth and include it in evaluation ratings. But schools and districts have another loophole. A provision built into Delaware’s evaluation law gives school administrators the discretion to upgrade a teacher’s student growth rating. The state created a specific performance category “Unsatisfactory (Discretion)” on the student growth component, whereby districts have the option to move those teachers to a higher rating despite their low scores. That means that even a quantitative metric—an indicator that is presumably completely objective—can still be revised based on someone’s subjective opinion.

All Delaware districts are taking advantage of this discretion to varying degrees. Statewide, 12 percent of teachers received an “Unsatisfactory (Discretion)” rating and were eligible for the upgrade in school year 2012-13, the first year of Delaware’s new evaluation system (see Figure 2). Delaware released anonymous results, but the extent of the revisions ranged from 32 percent in District A to 90 percent in District J. After these upgrades, 10 percent of teachers in Delaware received an “Unsatisfactory” rating, down from 17 percent based on the data alone (see Figure 3).

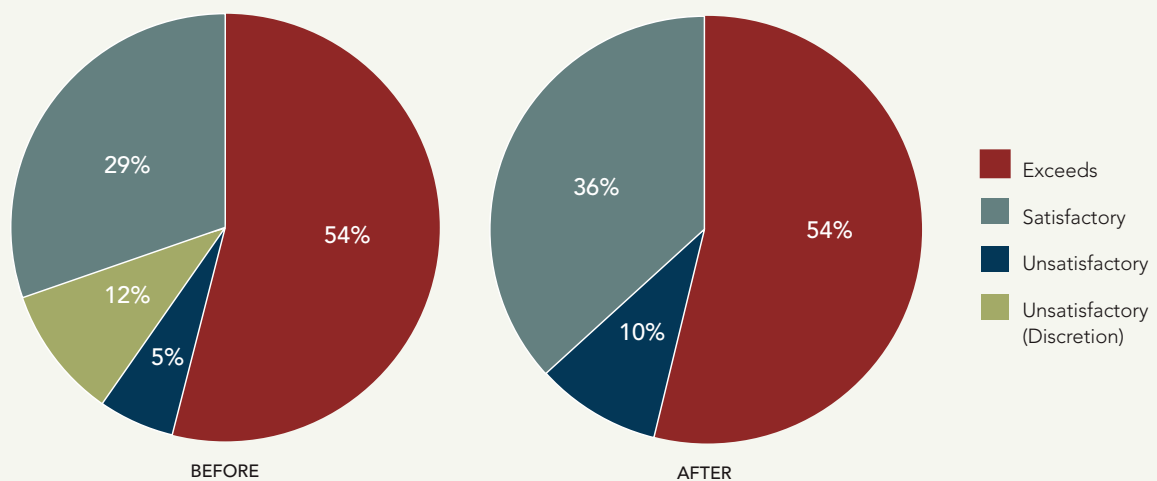
FIGURE 2: DELAWARE SCHOOL DISTRICTS SUBJECTIVELY UPGRADE TEACHER RATINGS FROM “UNSATISFACTORY (DISCRETION)” TO “SATISFACTORY”



Note: Districts with fewer than 10 “Unsatisfactory (Discretion)” teachers were excluded from figure.

Source: Delaware Department of Education, “Continuous Improvement: A Report on ‘Year One’ of the Revised DPAS-II Educator Evaluation System,” November 2013, http://www.doe.k12.de.us/tleu_files/DPAS_II_Year_One_Report_2013.pdf.

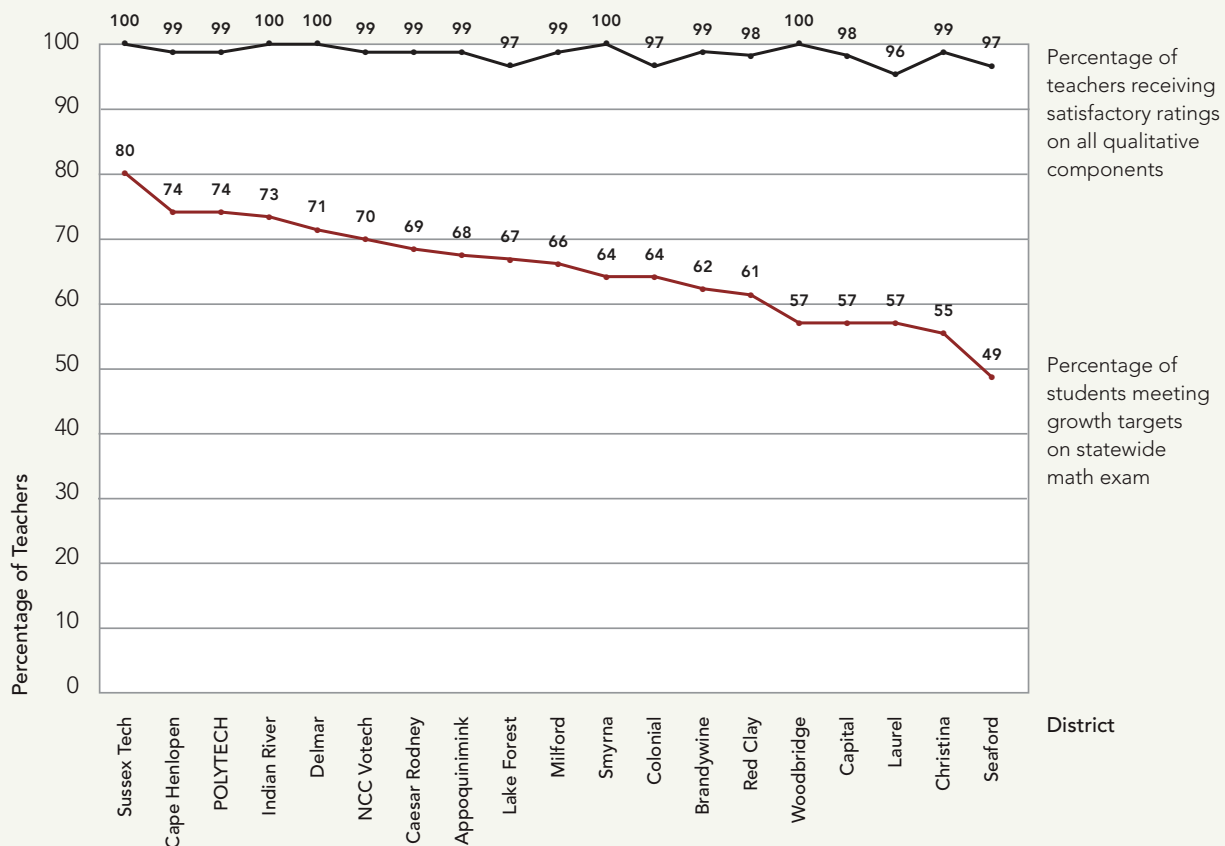
FIGURE 3: TEACHER EVALUATION RATINGS IN DELAWARE BEFORE (LEFT) AND AFTER (RIGHT) SCHOOL ADMINISTRATORS HAD THE OPTION TO UPGRADE RATINGS



Source: Delaware Department of Education, “Continuous Improvement: A Report on ‘Year One’ of the Revised DPAS-II Educator Evaluation System,” November 2013, http://www.doe.k12.de.us/tleu_files/DPAS_II_Year_One_Report_2013.pdf.

The reason this practice is so troubling is that the public also can't have much faith that other, non-student-growth data effectively differentiates educator performance. In the first year of the state's new evaluation system, Delaware school districts gave nearly all teachers satisfactory ratings on all qualitative components of the evaluation system (which could be why teachers report such favorable perceptions of the new evaluations). But student growth results present a very different picture of teacher and student performance in these same districts (see Figure 4). Uniformly high ratings on classroom observations, regardless of how much students learn, suggest a continued disconnect between how much students grow and the effectiveness of their teachers.

FIGURE 4: IN DELAWARE, QUALITATIVE EVALUATION RATINGS HAVE LITTLE CONNECTION TO STUDENT GROWTH



Source: Delaware Department of Education, "Continuous Improvement: A Report on 'Year One' of the Revised DPAS-II Educator Evaluation System," November 2013, http://www.doe.k12.de.us/tleu_files/DPAS_II_Year_One_Report_2013.pdf.

Florida provides another example of how local implementation decisions weaken the intent of state law. From the outside, Florida appears to have a tough, “one-size-fits-all” system. For teachers whose students take a statewide assessment, the state requires that student growth on the state test contribute 50 percent of a teacher’s evaluation rating, one

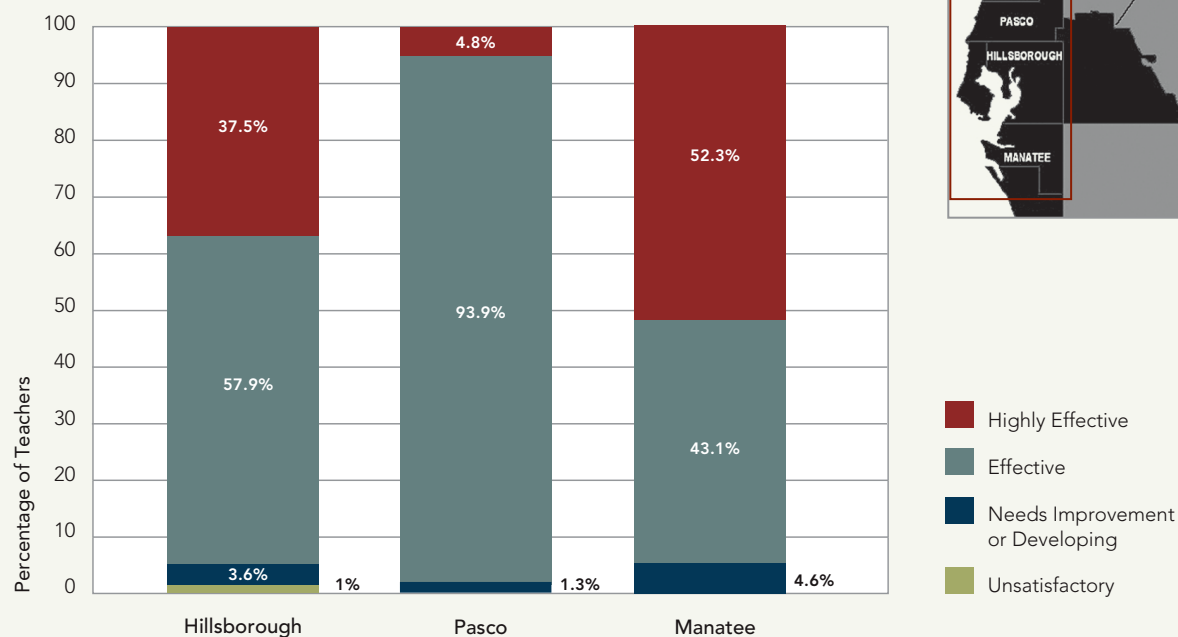
From the outside, Florida appears to have a tough, “one-size-fits-all” system. But each district gets to decide what to do with this information and how to turn the scores into an overall student growth rating.

of the highest percentages in the country. The state Department of Education compares students with their peers statewide, calculates a “value-added” score for each teacher based on how well his or her students performed compared with their peers, and provides the results back to school districts. But each district gets to decide what to do with this information and how to turn the scores into an overall student growth rating. Even with the infusion of objective data where, by *definition*, some teachers are more effective than others,

Florida school districts rated 98 percent of teachers as “Highly Effective” or “Effective” in school year 2012-13.²⁸

These choices manifest in very different outcomes for teachers. Evaluation ratings across school districts, even neighboring districts, can look very different (see Figure 5). Pasco County, for example, rated nearly 94 percent of its teachers as “Effective,” but only about 5 percent as “Highly Effective.” Its neighbor to the south, Hillsborough County, had more than eight times as many “Highly Effective” teachers.

FIGURE 5: THREE NEIGHBORING COUNTIES IN FLORIDA HAVE EXTREMELY DIFFERENT DISTRIBUTIONS IN TEACHER RATINGS



Note: “Developing” describes teachers in their first three years of teaching who would otherwise receive a “Needs Improvement” rating. For our analysis, we chose to collapse the two categories.

Source: Florida Department of Education, <http://www.fldoe.org/profdev/pdf/EduEvalRatingsMarch14.pdf>.

Even with policymakers requiring student growth to be included in teacher and principal evaluation ratings, it’s becoming clear that many states and districts aren’t embedding student growth in evaluation ratings in any meaningful way. The “widget effect” continues in all but a few places, and the ability of teachers and principals to improve student growth outcomes has little bearing on their evaluation outcomes. As a result, in many places there is still no clear connection between the results of educator evaluations and the academic achievement of students within the same school.

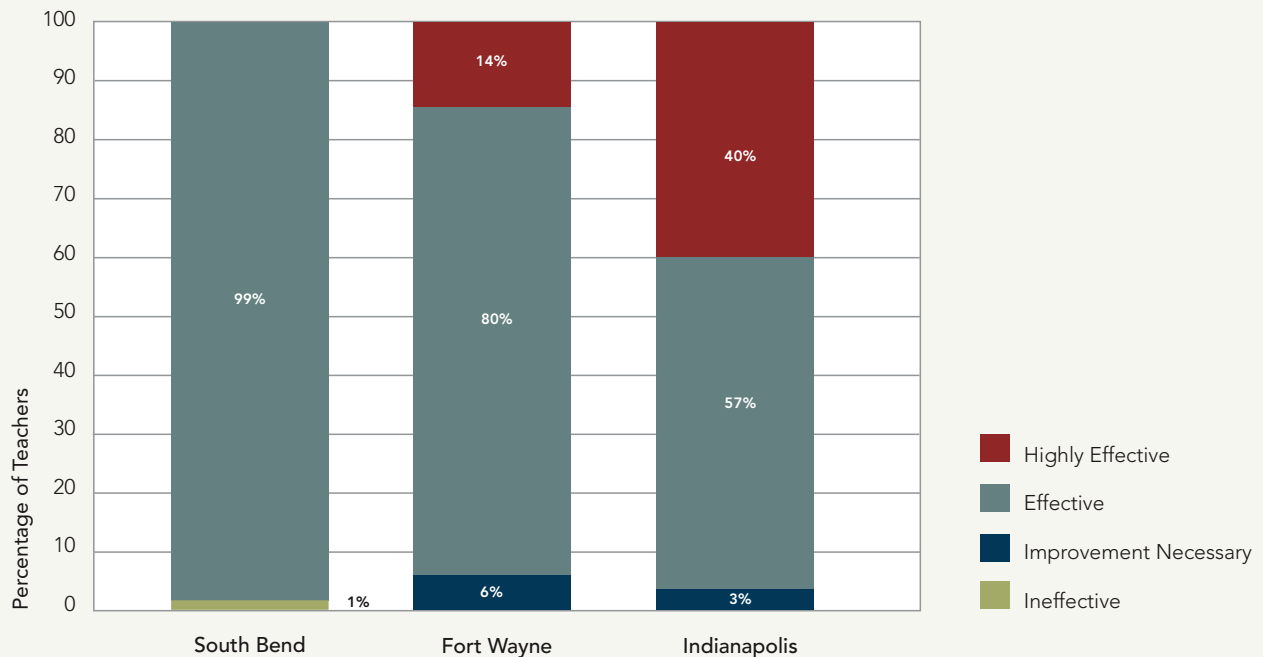
4. DISTRICTS HAVE WIDE DISCRETION EVEN UNDER “STATEWIDE” EVALUATION SYSTEMS

As evaluation reforms gained traction over the past few years, state and local advocates began to express concerns about a “one-size-fits-all” approach to educator evaluations.²⁹ But the truth is that the vast majority of states set minimum requirements for educator evaluations and leave local school districts responsible for the ultimate success of those efforts. Evaluation reform has not meant the end of local discretion.

Within “statewide” evaluation systems, districts have much more room for local discretion than commonly appreciated. Districts use this flexibility to determine how various required components are graded, scored, and compiled into overall educator evaluation ratings. Indiana provides one such example. Although the state passed teacher evaluation reforms in 2011 and developed a statewide evaluation model, districts have the option to develop their own system that varies how classroom observations and student growth factor into a teacher’s final rating. For example, Indiana state law now requires that “objective measures of student achievement and growth significantly inform the evaluation” of all teachers,³⁰ but it does not require a specific percentage devoted to student growth. A recent analysis of six Indiana school districts found that the student growth component constituted anywhere from zero to 40 percent of a teacher’s overall evaluation.³¹

Figure 6 shows the evaluation results for three of the largest school districts in the Hoosier State after they were required to begin using a four-tier evaluation system. Although South Bend is technically in compliance with state law in having four evaluation tiers, the district has effectively maintained its old habits. South Bend determined that none of its teachers deserved the highest rating of “Highly Effective” and none warranted “Improvement Necessary”—but 99 percent merited “Effective” ratings. Two other large districts, Fort Wayne and Indianapolis, actually identified lower percentages of educators as “Ineffective” than did South Bend (0.2 and 0.5 percent versus 1.1 percent, respectively), but they at least made use of all four performance categories.

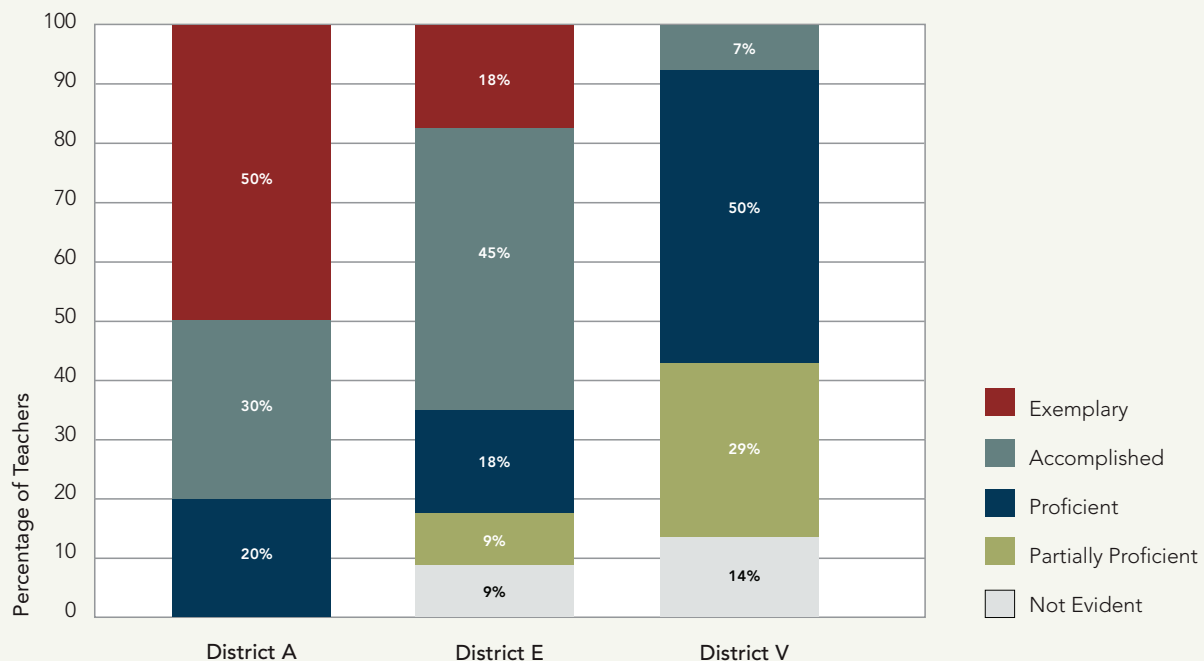
FIGURE 6: THREE INDIANA DISTRICTS ARE MAKING MIXED PROGRESS TOWARD DIFFERENTIATION



Source: Indiana Department of Education, Staff Performance Evaluation Results 2012-2013, <http://www.doe.in.gov/evaluations>.

Colorado provides another example of a state with large variation in teacher ratings across school districts. Reformers praised Colorado for passing a strong evaluation law mandating that 50 percent of all teacher evaluations be based on student growth. The state is still rolling out the new system and, three years later, it has not yet articulated how it will measure growth, but districts have already begun piloting the other components of the new system. Like Delaware, Colorado has released only anonymous data so far. But the data suggest that District A believes it has a very different teacher workforce than District E or V. District A did not identify a single teacher as “Partially Proficient” or “Not Evident,” whereas District V identified over 40 percent of its teachers in these two categories and none as “Exemplary” (see Figure 7).

FIGURE 7: DURING THE PILOT YEAR OF COLORADO’S NEW EVALUATION SYSTEM, DISTRICTS SHOWED WIDE VARIATION IN EVALUATION RATINGS



Note: Percentages may not total 100% due to rounding.

Source: Colorado Department of Education, “Colorado State Model Evaluation System for Teachers: 2012-2013 Pilot Year,” http://www.cde.state.co.us/sites/default/files/2012-13%20Teacher%20Pilot%20Report_0.pdf.

To take these results at face value, one must believe that some districts employ teachers that are orders of magnitude better than others. While teacher quality likely varies from district to district, it's hard to believe that the differences are this large.

To take these results at face value, one must believe that some districts employ teachers that are orders of magnitude better than others. While teacher quality likely varies from district to district, it's hard to believe that the differences are this large. An alternative explanation is simply that districts hold their teachers to vastly different standards even while complying with a “statewide” evaluation system. Most likely, it means some districts have easier grading scales.

It's reasonable to worry that statewide evaluation systems could eventually become overly prescriptive and squash appropriate local autonomy. But the data indicate

these worries are premature. Most districts have wide flexibility in how they implement the new evaluation systems, and they're using that flexibility to produce widely divergent results, for good and bad.

5. DISTRICTS CONTINUE TO IGNORE PERFORMANCE WHEN MAKING DECISIONS ABOUT TEACHERS

Most schools and districts still make consequential decisions based on inertia—granting tenure because someone’s been there the correct number of years—or credentials—basing compensation or layoff decisions solely on seniority. By standardizing policies for every teacher and school leader, despite differences in performance, schools and districts continue to make objective but ultimately uninformed decisions around compensation, retention, and dismissal. Given the many delays in rolling out evaluation reforms, it’s not surprising that districts have held off on using these results for consequential decision making.

Tennessee, for instance, became the first state to implement a comprehensive statewide evaluation system in 2011-12, but the state doesn’t require districts to make any personnel decisions based on the results. In its approved No Child Left Behind waiver request, Tennessee didn’t make any promises around consequential decision making. It noted that districts *may* dismiss teachers who were classified within the two lowest performance categories.³²

But potential dismissals are not the same thing as actual dismissals. On the assumption that firing teachers is newsworthy, we searched Tennessee newspapers to find stories of teachers fired for poor performance in 2013. The sum total: 60 teachers dismissed in Nashville, slightly more than 1 percent of its teaching workforce;³³ 97 teachers in Memphis, about 1.3 percent of its teachers;³⁴ and four teachers in Knox County, 0.1 percent of its 3,927 teachers.³⁵ Based on our search, we found no other examples of districts using the flexibility afforded in Tennessee state law.

Nationwide, it's rare to find examples of districts dismissing teachers for poor performance. A recent analysis of New York City schools found that the district fired a total of 12 teachers, out of 75,000 citywide, from 1997 to 2007.³⁶ This means that only *0.016 percent* of teachers were dismissed over this 10-year period, a very low percentage given the numerous challenges within the NYC school system. The entire state of New Jersey dismissed 23 teachers for poor performance between 2012 and 2014 (out of more than 100,000 classroom teachers statewide, this represents less than .02 percent).³⁷ After the New Haven, Connecticut school district implemented a new evaluation system in 2010, it dismissed 62 teachers over the next two years, an annual dismissal rate of 1.7 percent.³⁸

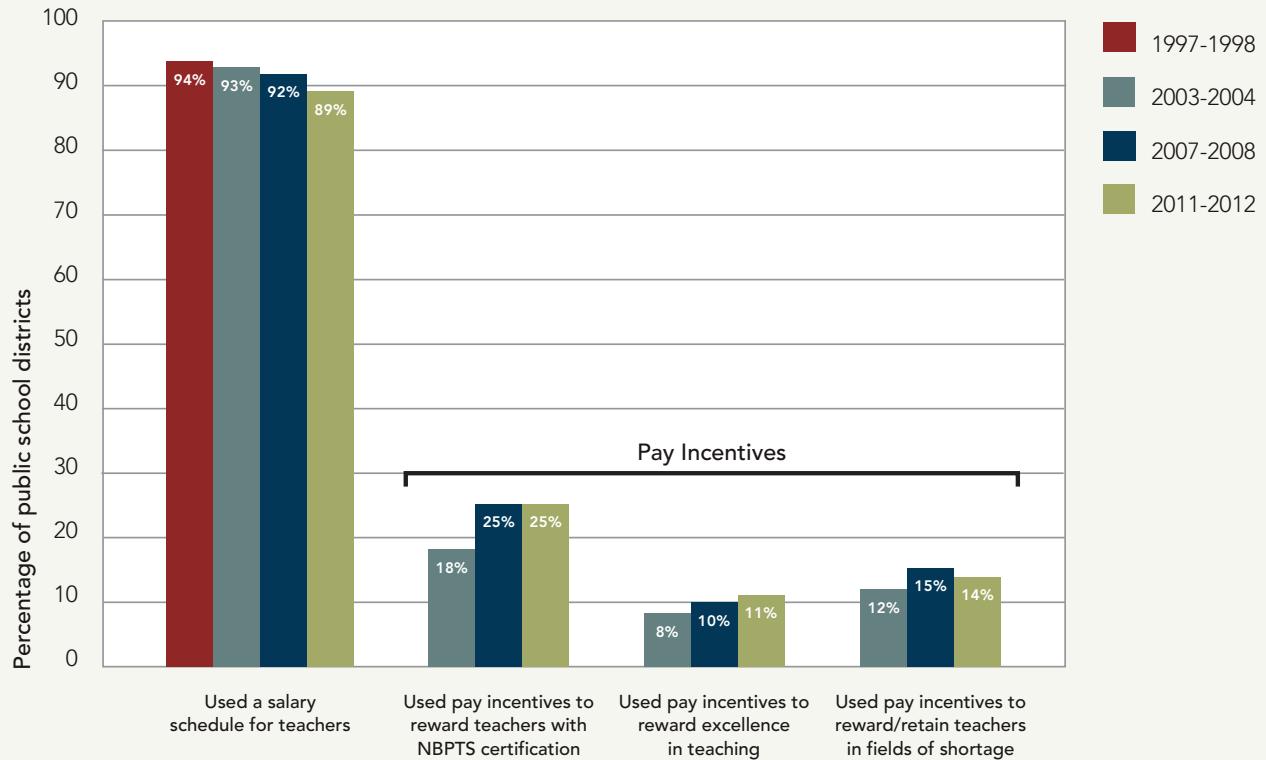
Washington, D.C. is one of the few places in the country known for systematically identifying and dismissing its low-performing teachers and principals. It made national news when the district fired 241 teachers for poor performance in 2010—CNN, *The New York Times*, PBS, the *Wall Street Journal*, and *The Washington Post* all published stories.³⁹ It no longer makes national news, but *The Washington Post* has continued to publish news stories as the district has continued to fire approximately 3 to 4 percent of its teaching workforce for poor performance each year.

Districts too often fail to identify their truly excellent teachers and principals, devise ways to provide them with extra compensation or flexibility, and implement strategies to ensure that the best teachers work with the students who need them most.

Dismissals receive an inordinate amount of attention because they're rare, easy to count, and the ultimate judgment on someone's performance. The amount of attention given to dismissals is unfortunate because there are many other actions that should be driven by performance before dismissal becomes part of the conversation. It's just as important to identify and act on excellence as it is on poor performance. Districts too often fail to identify their truly excellent teachers and principals, devise ways to provide them with extra compensation or flexibility, and implement strategies to ensure that the best teachers work with the students who need them most.

While educators don't enter the profession for financial reasons, money is one of the few incentives districts have at their disposal. And very few districts use cash to reward or retain their best employees. According to the most recent national data, one-fourth of school districts are rewarding teachers who completed a voluntary, peer-reviewed process developed by the National Board for Professional Teaching Standards (NBPTS) with an extra financial incentive. As of 2012, only 11 percent of districts rewarded teachers specifically for excellent performance, and only 14 percent of districts offered extra compensation to reward or retain teachers in hard-to-staff shortage areas (see Figure 8). These numbers have barely budged over the last 15 years.

FIGURE 8: NINE OUT OF 10 TEACHERS ARE STILL ON A SALARY SCHEDULE, AND IT REMAINS UNCOMMON FOR TEACHERS TO RECEIVE PAY INCENTIVES



Note: The data on pay incentives in the Schools and Staffing Survey only go back to 2003-04.

Source: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey (SASS), <https://nces.ed.gov/surveys/sass/>.

It’s also useful to not only look at *whether* teachers receive financial incentives, but also at the *level* of incentive offered. Teachers in Fort Wayne, Indiana receiving an “Effective” or “Highly Effective” rating can receive a base salary increase of \$1,100. About 84 percent of teachers qualified in 2013. This is a step up from the prior contract, which offered no financial incentive based on teaching performance, but those same teachers would be better off earning a master’s degree, which would increase a teacher’s salary by \$4,000 regardless of effectiveness.⁴⁰ Given that research studies have found that teachers with master’s degrees are not necessarily more effective than those without one, one must wonder why Fort Wayne is willing to pay teachers almost four times as much for a credential rather than for actual classroom performance.⁴¹

Likewise, instead of taking proactive steps to influence whether a teacher stays or leaves, districts often leave staffing decisions up to the teachers themselves. In most districts, principals have little control over who works in their school, and a recent survey found that districts and school principals did not make strategic efforts to retain their best employees.⁴²

Failure to differentiate high- and low-performers also hurts students. According to research from the University of Washington's Center for Education Data & Research, using seniority as the sole factor in making layoff decisions forces districts to pink-slip more teachers. Because less-experienced teachers earn lower salaries, a district has to lay off about 10 percent more teachers to achieve the same cost reductions as an across-the-board cut.⁴³ Because a policy that relies on seniority and ignores performance will force districts to lay off both high- and low-performing employees, rather than only low-performing ones, the overall result is a less effective teaching workforce. It seems ludicrous to purposely dismiss a great teacher while retaining poor ones, but some school policies on tenure and layoffs do just that.

LESSONS FOR POLICYMAKERS

Policy makers should take four things away from the early implementation efforts.

TRACK DATA ON THE COMPONENTS, SUMMATIVE RESULTS, AND USES OF THE NEW EVALUATION SYSTEMS

Without information on how districts are evaluating instructional practice, assigning evaluation ratings, and using these results to make personnel decisions, state leaders must rely on anecdotal evidence that may not fully capture what’s happening in schools. Instead, states should collect and publicly report summative evaluation ratings at the school and district level (but not for individual educators), the component elements that make up those ratings, and how the ratings are being used to drive personnel decisions. In addition, states should conduct surveys and focus groups of educators to collect best practices and gather feedback on the evaluation system itself. And states should monitor district processes, such as whether they are meeting deadlines and acting in a timely fashion. Publicly reporting such information can act as a form of accountability. Local news coverage and engaged families and community members may ultimately be able to shape ongoing implementation efforts.

Tracking the results of the evaluation systems need not be expensive. The Office of Management and Budget estimated that it would cost \$7.5 million a year to get the data from every single school and district in the country and release it to the public.⁴⁴ When asked as part of their State Fiscal Stabilization Fund applications how much it would cost to provide this information, the state of California said it would cost \$93,750⁴⁵ and the Texas Department of Education estimated a total of \$38,000 to collect and provide the data.⁴⁶ For such important information, these are very modest sums in states with school budgets in the tens of billions of dollars.

States also have an important role to play in overseeing and improving implementation efforts. State education agencies are the best-positioned body to conduct quality audits to assess district implementation efforts and serve as an outside voice to ensure that all educators are given timely and actionable feedback on their practice. As states gather more data over time, they will be better able to assess the relationship between classroom practices and student growth. With a more nuanced understanding of what matters in the classroom, states will be able to continue improving their evaluation systems.

WORK CLOSELY WITH DISTRICTS TO UNDERSTAND THE CAUSES OF MEASURED OUTCOMES AND ANY VARIATIONS

Under new evaluation systems, decisions about what percentage of teachers receive which ratings are ultimately judgment calls by schools and districts. States should resist the temptation to “solve” these inequities or set quotas for unsatisfactory evaluations. Disparities in evaluation ratings may be as much a function of local philosophies as of actual performance. Especially in the early stages of implementation, states should be wary of jumping to conclusions if they see wide variations across districts. Instead, they should work to ensure that district evaluations are consistently rigorous across schools and classrooms. Introducing smart timelines for action, multiple evaluation measures including student growth, requirements for data quality, and a policy to use confidence intervals in the case of student growth measures could all protect districts and educators that set ambitious goals. In cases of inflated evaluation ratings, states should fight the urge to respond by imposing ever-tighter policies and, as Andrew J. Rotherham, Sara Mead, and Rachael Brown warned in a recent report, “States should not mistake processes and systems as substitutes for cultural change.”⁴⁷

DON'T HALT OR WEAKEN IMPLEMENTATION BEFORE REFORMS HAVE A CHANCE TO TAKE EFFECT

Instead of waiting for objective evidence on how their new teacher evaluation systems are playing out, some states—such as Indiana,⁴⁸ Hawaii,⁴⁹ New Jersey,⁵⁰ Ohio,⁵¹ and Maryland⁵²—have preemptively decreased the weighting for test-based student growth. Other states have delayed full implementation of their evaluation systems as they navigate new assessments aligned to the Common Core standards.⁵³ Research suggests that reducing the weighting of student growth, while politically appealing, will weaken the ability of evaluation ratings to predict who will be an effective teacher in the future.⁵⁴

The U.S. Department of Education (USED) is guilty of prematurely backing away from its promises as well. In 2011, USED offered states flexibility from No Child Left Behind in exchange for, among other things, adopting new teacher and principal evaluation systems. Those systems were originally supposed to be in place by 2014-15, but the federal government delayed the timeline so that states don’t need to make decisions based on their evaluation systems until 2018. Instead of waiting to see how things would play out, USED preemptively extended its deadlines four years into the future—and that’s assuming timelines aren’t pushed back again.

USED has even backed away from its earlier insistence on good data. Although its plan to collect teacher and principal evaluation data through the State Fiscal Stabilization Fund was met with great fanfare, USED quietly waived the requirement for the 42 states and the District of Columbia that received a comprehensive waiver to NCLB.⁵⁵ USED also failed to enforce the requirement for the remaining eight states. Of these eight states, California and Vermont were the only two to follow through with their commitments and report the data. Six states—Illinois, Iowa, Montana, Nebraska, North Dakota, and Wyoming—never fulfilled their promise and have faced no repercussions.

EVALUATION REFORM CAN CO-EXIST WITH OTHER CHANGES

With so many reforms in education taking place simultaneously, it’s easy to understand why educators feel overwhelmed. Schools and districts are being bombarded with reminders to implement Common Core standards, utilize technology in the classroom in innovative ways, and prepare students for the realities of a 21st century workforce. Some observers have called for a “moratorium” on consequences attached to educator evaluation results as students and teachers adjust to these new circumstances. While that’s an understandable impulse, not all states or districts may need it. For example, when New York administered new, tougher assessments in 2013, student scores plunged. The percentage of students deemed “proficient” by the state fell from 55 percent in reading and 65 percent in math to 31 percent in both subjects.⁵⁶ Despite lower student scores, teacher scores on the official state-provided student growth measure stayed nearly identical.⁵⁷ In other words, rather than freezing evaluation reforms and choosing to make slow, incremental changes, it is possible to make simultaneous large-scale changes in education.

CONCLUSION

In recent years states have made sweeping changes to the way teachers and principals are evaluated. Schools and districts shifted from seeing educator performance in terms of black or white—whether an educator was “satisfactory” or not—to differentiating excellence from mediocrity and mediocrity from ineffectiveness. Teachers and principals are receiving more frequent, better feedback than ever before. The results of those evaluations, however, are still too often divorced from what happens to students and how much they learn. And districts rarely make consequential decisions about the adults working in their schools based on their on-the-job performance.

Shifting to a system that values performance has not been easy, and reform fatigue has begun to set in. Policymakers delayed or quietly modified evaluation systems before the systems even had a chance to effect change. States made significant improvements on paper to their evaluation policies; changing the culture in schools will be much harder but will have an even greater impact.

ENDNOTES

¹ NCTQ, “State of the States 2012: Teacher Effectiveness Policies,” 2012, http://www.nctq.org/dmsView/State_of_the_States_2012_Teacher_Effectiveness_Policies_NCTQ_Report.

² See Thomas J. Kane et al., Education Next, “Evaluating Teacher Effectiveness: Can classroom observations identify practices that raise achievement?,” 2011, http://educationnext.org/files/ednext_20113_research_kane.pdf; and Eric S. Taylor and John H. Tyler, Education Next, “Can Teacher Evaluation Improve Teaching?,” 2012, http://educationnext.org/files/ednext_20124_taylortyler.pdf.

³ Thomas Dee and James Wyckoff, Stanford University Center for Education Policy Analysis, “Incentives, Selection, and Teacher Performance: Evidence from Impact,” NBER Working Paper Series, October 2013, <http://cepa.stanford.edu/sites/default/files/w19529.pdf>.

⁴ Dan Goldhaber and Joe Walch, University of Washington Center for Education Data & Research, “Strategic Pay Reform: A Student Outcomes-Based Evaluation of Denver’s ProComp Teacher Pay Initiative,” 2011, [http://www.cedr.us/papers/working/WP%202011-3%20Procomp%20Strategic%20Compensation%20\(9-28\).pdf](http://www.cedr.us/papers/working/WP%202011-3%20Procomp%20Strategic%20Compensation%20(9-28).pdf).

⁵ David Weisberg et al., TNTP, *The Widget Effect*, http://tntp.org/assets/documents/TheWidgetEffect_2nd_ed.pdf.

⁶ Based on data availability, we focus exclusively on teachers in much of this paper. Our preference would be to look at how districts evaluate and support *all* educators, including school leaders and support staff.

⁷ NCTQ, “State of the States 2012: Teacher Effectiveness Policies,” 2012.

⁸ Kathryn M. Doherty and Sandi Jacobs, “Connect the Dots: Using Evaluations of Teacher Effectiveness to Improve Policy and Practice,” NCTQ State of the States, October 2013, http://www.nctq.org/dmsView/State_of_the_States_2013_Using_Teacher_Evaluations_NCTQ_Report.

⁹ Susanna Loeb, Luke C. Miller, and James Wyckoff, “Performance Screens for School Improvement: The Case of Teacher Tenure Reform in New York City,” May 2014, <http://cepa.stanford.edu/sites/default/files/NYCTenure%20brief%20FINAL.pdf>.

¹⁰ Georgia Department of Education, “Overview to the 2012 TKES/LKES Pilot Evaluation Report,” December 2012, http://www.gadoe.org/School-Improvement/Teacher-and-Leader-Effectiveness/Documents/Pilot%20Report_Overview%20and%20Report%20Combined%201-10-13.pdf.

¹¹ Louisiana Department of Education, “2012-2013 Compass Final Report,” <https://www.louisianabelieves.com/teaching/compass-final-report-2012-2013>.

¹² We excluded District of Columbia Public Schools because it’s not a state and used a four-tier evaluation system even before it introduced a new, more rigorous evaluation system called IMPACT in 2009-10. In the first year under IMPACT, teacher evaluation ratings shifted dramatically. See page 3 here: <http://dcps.dc.gov/DCPS/Files/downloads/ABOUT%20DCPS/Press/2010-11%20IMPACT%20Results%20Overview.pdf>.

¹³ Grover J. Whitehurst, Matthew M. Chingos, and Katharine Lindquist, Brookings Institution, “Evaluating Teachers with Classroom Observations: Lessons Learned in Four Districts,” May 2014, <http://www.brookings.edu/~media/research/files/reports/2014/05/13%20teacher%20evaluation/evaluating%20teachers%20with%20classroom%20observations.pdf>.

¹⁴ Doherty and Jacobs, “Connect the Dots: Using Evaluations of Teacher Effectiveness to Improve Policy and Practice,” 2013.

¹⁵ Bill & Melinda Gates Foundation, “Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains,” January 2012, http://www.metproject.org/downloads/MET_Gathering_Feedback_Research_Paper.pdf.

¹⁶ Teachscape, “Charlotte Danielson’s Framework for Teaching (FFT) Chosen as a Teaching Evaluation Rubric for New York State,” July 2011, <http://www.teachscape.com/about/press-and-news/2011/charlotte-danielson%E2%80%99s-framework-for-teaching-fft-chosen-as-a-teaching-evaluation-rubric-for-new-york-state.html>. Lauren Sartain et al., The University of Chicago Consortium on Chicago School Research, “Rethinking Teacher Evaluation in Chicago: Lessons Learned from Classroom Observations, Principal-Teacher Conferences, and District Implementation,” November 2011, <http://ccsr.uchicago.edu/publications/rethinking-teacher-evaluation-chicago-lessons-learned-classroom-observations-principal>.

¹⁷ Whitehurst, Chingos, and Lindquist, “Evaluating Teachers with Classroom Observations: Lessons Learned in Four Districts,” 2014.

¹⁸ Duncan Chaplin et al., Mathematica Policy Research, “Professional practice, student surveys, and value-added: Multiple measures of teacher effectiveness in Pittsburgh Public Schools,” July 2014, http://ies.ed.gov/ncee/edlabs/regions/midatlantic/pdf/REL_2014024.pdf.

¹⁹ Morgaen L. Donaldson et al., UConn Center for Education Policy Analysis, “An Evaluation of the Pilot Implementation of Connecticut’s System for Educator Evaluation and Development,” January 2014, http://www.connecticutseed.org/wp-content/uploads/2014/01/Neag_Final_SEED_Report_1-1-2014.pdf.

²⁰ Delaware Department of Education, “Continuous Improvement: A Report on ‘Year One’ of the Revised DPAS-II Educator Evaluation System,” November 2013, http://www.doe.k12.de.us/tleu_files/DPAS_II_Year_One_Report_2013.pdf.

- ²¹ Doherty and Jacobs, “Connect the Dots: Using Evaluations of Teacher Effectiveness to Improve Policy and Practice,” 2013.
- ²² See Section 44662 of California’s Education Code here: <http://www.leginfo.ca.gov/cgi-bin/displaycode?section=edc&gro up=44001-45000&file=44660-44665>.
- ²³ See California Department of Education, Great Teacher and Principals Survey, <http://www2.cde.ca.gov/greatteachers/search.aspx>. We identified California’s 10 largest districts by referring to the state Department of Education’s report found at: <http://www.cde.ca.gov/ds/sd/cb/ceflargesmalldist.asp>.
- ²⁴ Massachusetts Department of Elementary and Secondary Education, “Massachusetts Model System for Educator Evaluation,” August 2012, <http://www.doe.mass.edu/eval/model/PartVII.pdf>.
- ²⁵ See U.S. Department of Education, Utah ESEA Flexibility Waiver, <http://www2.ed.gov/policy/eseaflex/approved-requests/ut.pdf>; and Arkansas Department of Education, Rules Governing the Teacher Excellence and Support System, Section 6.03, July 2014, http://www.arkansased.org/public/userfiles/rules/Current/TESS_EMERGENCY_RULE.pdf.
- ²⁶ Colorado Department of Education, “Measures of Student Learning: Approaches for Selecting and Using Multiple Measures in Educator Evaluations,” August 2013, <http://www.cde.state.co.us/sites/default/files/MeasuresofStudentLearningFINAL081413.pdf>.
- ²⁷ Tennessee Department of Education, “Teacher Evaluation in Tennessee: A Report on Year 1 Implementation,” July 2012, http://www.tn.gov/education/teaching/docs/yr_1_tchr_eval_rpt.pdf.
- ²⁸ Florida Department of Education, <http://www.fldoe.org/profdev/pdf/EduEvalRatingsMarch14.pdf>.
- ²⁹ See, for example, Tom Precious, *The Buffalo News*, “New union president sees a new NYSUT coming,” April 2014, http://blogs.buffalonews.com/politics_now/2014/04/new-nysut-president-a-new-nysut-is-coming.html; Liz Bowie, *The Baltimore Sun*, “State seeks to delay using test scores to evaluate teachers,” February 2014, http://articles.baltimoresun.com/2014-02-25/news/bs-md-teacher-evaluation-extension-20140225_1_common-core-maryland-education-association-state-school-board; Lisa Gartner, *Tampa Bay Times*, “State takes notice as Pinellas tries a new way to evaluate teachers,” April 2014, <http://www.tampabay.com/news/education/k12/state-takes-notice-as-pinellas-tries-a-new-way-to-evaluate-teachers/2173765>.
- ³⁰ See Indiana Code 20-28-11.5-4, <https://iga.in.gov/legislative/laws/2014/ic/titles/020/articles/028/chapters/11.5/>.
- ³¹ Stand for Children Indiana, “Teacher Evaluation in Indiana: A Review of PL 90 and Its Implementation,” May 2014, <http://stand.org/sites/default/files/Indiana/Teacher%20Evaluations%20in%20Indiana%20-%20June%202014.pdf>.
- ³² U.S. Department of Education, Tennessee ESEA Flexibility Waiver, January 2012, <http://www2.ed.gov/policy/eseaflex/approved-requests/tn.pdf>.
- ³³ The dismissal figures come from Andrea Zelinski, Nashville City Paper, “Metro Schools preparing to fire low-rated teachers,” July 2013, <http://nashvillecitypaper.com/content/city-news/metro-schools-preparing-fire-low-rated-teachers>. Nashville Public Schools employed 5,167 certificated teachers in school year 2013-14. See Metropolitan Nashville Public Schools, “2013-14 Facts,” <http://www.mnps.org/AssetFactory.aspx?did=81681>.
- ³⁴ Dismissal figures come from The Associated Press, “Memphis schools plan on cutting at least 97 teachers,” May 2013, <http://www.knoxnews.com/news/2013/may/16/memphis-schools-plan-cutting-at-least-97/>. The total number of teachers in the district comes from the Shelby County Board of Education, “Fiscal Year 2014-15 Proposed Budget,” <http://www.scsk12.org/uf/finance/2014/fy15.pdf>.
- ³⁵ The dismissal figures come from Trevor TenBrink, Education Action Group Foundation, “New evaluations exposing weak instructors in Knox County, Tennessee,” August 2013, <http://eagnews.org/new-tennessee-evaluations-exposing-struggling-teachers-in-knox-county/>. The total number of teachers comes from Knox County Schools, <http://knoxschools.org/Page/5504>.
- ³⁶ Katharine B. Stevens, *New York Daily News*, “Firing teachers: mission impossible,” February 2014, <http://www.nydailynews.com/opinion/firing-teachers-mission-impossible-article-1.1615003>.
- ³⁷ Diane D’Amico, *Press of Atlantic City*, “Both sides say New Jersey’s new tenure hearings working,” May 2014, http://m.pressofatlanticcity.com/education/both-sides-say-new-jersey-s-new-tenure-hearings-working/article_1df51e38-dc84-11e3-a884-0019bb2963f4.html?mode=jqm.
- ³⁸ Devin Leonard, *Businessweek*, “New Haven Shows How You Fix Public Schools,” May 2013, <http://www.businessweek.com/articles/2013-05-09/new-haven-shows-how-you-fix-public-schools#p1>.
- ³⁹ See, for example, Sally Holland, CNN, “D.C. school system fires 241 teachers,” July 2010, <http://www.cnn.com/2010/US/07/23/district.of.columbia.teachers.fired/>; Tamar Lewin, *The New York Times*, “School Chief Dismisses 241 Teachers in Washington,” July 2010, http://www.nytimes.com/2010/07/24/education/24teachers.html?_r=0; VDIC, PBS News, “D.C. Schools Chief Rhee Fires 241 Teachers Using New Evaluation System,” July 2010, <http://www.pbs.org/newshour/rundown/dc-schools-chief-rhee-fires-241-teachers-using-new-evaluation-system/>; Stephanie Banchero, *Wall Street Journal*, “Teachers Lose Jobs Over Test Scores,” July 2010, <http://online.wsj.com/news/articles/SB10001424052748704249004575385500484438266>; Bill Turque, *The Washington Post*, “Rhee dismisses 241 D.C. teachers; union vows to contest firings,” July 2010, <http://www.washingtonpost.com/wp-dyn/content/article/2010/07/23/AR2010072303093.html>.
- ⁴⁰ See pages 19-20 of the Master Contract Between the Board of School Trustees of Fort Wayne Community Schools and Fort Wayne Education Association, Inc., July 1, 2013-June 30, 2015: <http://fwcs.k12.in.us/files/FWEAContract.pdf>.

⁴¹ See Douglas N. Harris and Tim R. Sass, Urban Institute National Center for Analysis of Longitudinal Data in Education Research, “Teacher Training, Teacher Quality, and Student Achievement,” March 2007, <http://files.eric.ed.gov/fulltext/ED509656.pdf>; Steven G. Rivkin, Eric A. Hanushek, and John F. Kain, “Teachers, Schools, and Academic Achievement,” *Econometrica* 73, No. 2, March 2005, <http://www.econ.ucsb.edu/~jon/Econ230C/HanushekRivkin.pdf>.

⁴² TNTP, “The Irreplaceables: Understanding the Real Retention Crisis in America’s Urban Schools,” July 2012, http://tntp.org/assets/documents/TNTP_Irreplaceables_2012.pdf.

⁴³ Dan Goldhaber and Roddy Theobald, University of Washington Center for Education Data & Research, “Managing the Teacher Workforce in Austere Times: The Implications of Teacher Layoffs,” July 2010, [http://www.cedr.us/papers/working/CEDR%20WP%202011-1.2%20Teacher%20Layoffs%20\(6-15-2011\).pdf](http://www.cedr.us/papers/working/CEDR%20WP%202011-1.2%20Teacher%20Layoffs%20(6-15-2011).pdf).

⁴⁴ See Federal Register Volume 76 Number 185, September 2011, <http://www.gpo.gov/fdsys/pkg/FR-2011-09-23/pdf/2011-24407.pdf>.

⁴⁵ State of California, Application for Funding for Phase II of the Education Fund under the State Fiscal Stabilization Fund Program, <http://www2.ed.gov/programs/statestabilization/stateapps/phase-ii/ca-phase-2-approved.pdf>.

⁴⁶ State of Texas, Application for Funding for Phase II of the Education Fund under the State Fiscal Stabilization Fund Program, <http://www2.ed.gov/programs/statestabilization/stateapps/phase-ii/tx-phase-2-approved.pdf>.

⁴⁷ Sara Mead, Andrew J. Rotherham, and Rachael Brown, “The Hangover: Thinking about the Unintended Consequences of the Nation’s Teacher Evaluation Binge,” American Enterprise Institute’s Teacher Quality 2.0 series, September 2012, <http://bellwethereducation.org/sites/default/files/legacy/2012/09/Teacher-Quality-Mead-Rotherham-Brown.pdf>.

⁴⁸ Indiana Department of Education, “RISE Evaluation and Development System: Evaluation and Teacher Handbook Version 2.5,” July 2013, http://www.nctq.org/docs/rise_2_5_handbook_final.pdf.

⁴⁹ See Hawaii’s request to amend its approved Race to the Top grant at U.S. Department of Education Office of the Deputy Secretary, July 2013, <http://www2.ed.gov/programs/racetothetop/amendments/hawaii-13.pdf>.

⁵⁰ John Mooney, *NJ Spotlight*, “Christie sticks to middle of the road in arriving at PARCC decision,” July 2014, <http://www.njspotlight.com/stories/14/07/14/christie-sticks-to-middle-of-the-road-in-arriving-at-parcc-decision/>.

⁵¹ Patrick O’Donnell, *Cleveland Plain Dealer*, “Change in how teachers are graded in Ohio is a compromise in a big debate,” June 2014, http://www.cleveland.com/metro/index.ssf/2014/06/change_in_how_teachers_are_gra.html.

⁵² See Maryland’s request to amend its approved Race to the Top grant at U.S. Department of Education Office of the Deputy Secretary, December 2012, <http://www2.ed.gov/programs/racetothetop/amendments/maryland-10.pdf>.

⁵³ See, for example, Connecticut: <http://courantblogs.com/capitol-watch/malloys-delays-teacher-evals-sets-up-common-core-task-force/> and Georgia: <http://www2.ed.gov/programs/racetothetop/amendments/georgia-12.pdf>.

⁵⁴ Bill & Melinda Gates Foundation, “Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains,” 2012.

⁵⁵ See Federal Register Volume 77 Number 20, January 2012, <http://www.gpo.gov/fdsys/pkg/FR-2012-01-31/pdf/2012-2125.pdf>.

⁵⁶ New York City Department of Education, “2013 New York State Common Core Test Results: New York City Grades 3-8,” http://www.nyc.gov/html/om/pdf/2013/2013_math_ela_deck.pdf.

⁵⁷ Anika Anand, Chalkbeat New York, “In a low-stakes year, NYC teachers excel on growth scores,” September 2013, http://ny.chalkbeat.org/2013/09/05/in-a-low-stakes-year-nyc-teachers-excel-on-growth-scores/#.U5_L5fldWYh.

APPENDIX

State	Included Educators	Available Years	Pilot or All Districts	School, District, or State-level data	Includes Student Growth or SLOs?	Data Source	Additional Data Source
California	teachers, principals	2010, 2011	all districts	district-level	varies by district	Great Teachers and Principals Survey	
Colorado	teachers, principals	SY 2012-13	pilot with 26 districts (22 included in report)	district-level	not for pilot year, but will be included in later years	Colorado State Model Evaluation System for Teachers: 2012-2013 Pilot Report	Colorado State Model Evaluation System for Principals: 2012-2013 Pilot Report
Connecticut	teachers	SY 2012-13	pilot with 14 districts (12 included in report)	district-level	yes	An Evaluation of the Pilot Implementation of Connecticut's System for Educator Evaluation and Development	
Washington, D.C.*	teachers	SY 2009-10, 2010-11, 2011-12, 2012-13	entire district	district-level	yes	2010-2011 IMPACT Results	Retention of and Access to Effective Teachers in D.C. Public Schools
Delaware	teachers	SY 2012-13	all districts	district-level	yes	A Report on "Year One" of the revised DPAS-II Educator Evaluation System	
Florida	teachers, principals	SY 2011-12, SY 2012-13	all districts	school-level	yes	District Performance Evaluation Systems	
Georgia	teachers, principals	Pilot from Jan.-May 2012	pilot with 26 districts	pilot-level (aggregate data for participating districts)	not for pilot year, but will be included in later years	Overview to the 2012 TKES/LKES Pilot Evaluation Report	
Indiana	teachers	SY 2012-13	all districts	school-level	yes	Staff Performance Evaluation Results 2012-2013	

*Although the District of Columbia Public Schools is a district, not a state, it has released several years of data on its four-tiered evaluation system known as IMPACT.

APPENDIX

State	Included Educators	Available Years	Pilot or All Districts	School, District, or State-level data	Includes Student Growth or SLOs?	Data Source	Additional Data Source
Louisiana	teachers, principals	SY 2012-13	all districts	school-level	yes	2012-2013 Compass Reports	
Maine	teachers, principals	2010, 2011, 2012, 2013	all districts	school-level	NA	Maine Educator Quality Report	
Michigan	teachers, principals	SY 2011-12	all districts	school-level	yes	Educator Effectiveness Ratings and Factors Report	
New Jersey	teachers, principals	SY 2011-12, 2012-13	pilot with 10 districts in 11-12, 30 districts in 12-13	pilot-level (aggregate data for participating districts)	not for pilot year, but will be included in later years	Evaluation Pilot Advisory Committee: Final Report 2013	2011-2012 Evaluation Pilot Advisory Committee Interim Report
New York	teachers, principals	SY 2012-13	all districts, excluding NYC	state-level	yes	Composite Scores 2012-13: Preliminary APPR Results	
North Carolina	teachers, principals	SY 2011-12, 2012-13	all districts	school-level	yes, beginning SY 2012-2013	North Carolina Educator Effectiveness Data	
Rhode Island	teachers, principals	SY 2012-13	all districts	school-level	yes	Year 1 Report on Educator Evaluations	
Tennessee	teachers	SY 2011-12, 2012-13	all districts	state-level	yes	Teacher Evaluation in Tennessee: A Report on Year 1 Implementation	Teacher Evaluation in Tennessee: A Report on Year 2 Implementation
Texas	teachers, principals	SY 2010-11	all districts	district-level	varies by district		
Vermont	teachers, principals	SY 2010-11	all districts	state-level	varies by district	State of Vermont Teacher and Principal Evaluation Survey Results	State of Vermont Principal Evaluation Survey

Note: This Appendix includes all data that was available through July 2014.