

# For Good Measure?

Teacher Evaluation Policy in the ESSA Era



Kaitlin Pennington and Sara Mead



**BELLWETHER**  
EDUCATION PARTNERS  
IDEAS | PEOPLE | RESULTS



# Table of Contents

*Click on each title below to jump directly to the corresponding section.*

Executive Summary	<b>3</b>
Introduction	<b>6</b>
How We Got Here	<b>8</b>
The Good and Bad of Teacher Evaluation Reform	<b>14</b>
Risks Today	<b>20</b>
Overarching Lessons for Education Reform	<b>26</b>
Policy Recommendations	<b>28</b>
Conclusion	<b>31</b>
Endnotes	<b>32</b>
Acknowledgments	<b>36</b>

## Executive Summary

Over the past six years, there has been a lot of movement on teacher evaluation policy. From 2009 to 2015, 28 states enacted teacher evaluation laws requiring that objective measures of student achievement be included in teacher evaluations.<sup>1</sup> But many states and districts are now experiencing a backlash on teacher evaluation policies.

With the passage of the Every Student Succeeds Act (ESSA), which ushers in a new education policy era characterized by greater state flexibility and decreasing federal oversight, there are new risks for teacher evaluation policies. The coming years could well mark another era of transformation for how states and districts measure effective teaching.

As states and districts consider potential changes to their teacher evaluation systems and policies, this paper seeks to inform those efforts by reviewing the evolution of the teacher evaluation policy movement over the last several years, identifying positive outcomes of new systems and negative consequences, and describing risks that should be considered in a post-ESSA world. The risks to consider include:

- **Shifting focus to professional development without considering the design features and incentives needed to make evaluations useful for this purpose**

As teacher evaluation systems focus more on professional development, policymakers and practitioners need to consider if current systems are designed to properly address the systems' new intent, as well as the impact of this shift on incentives for both teachers and system leaders.

- **Impact of eliminating student achievement measures**

As some states and districts consider eliminating or scaling back the use of student growth measures in teacher evaluation systems, there is a risk that teacher evaluations will become less reliable and more expensive.

- **Loss of accountability**

If states and districts move away from both teacher and school accountability at the same time, there is a risk that neither schools nor teachers will be truly accountable for student achievement results in most schools. This could lead to a lack of progress in improving student outcomes, and exacerbate or slow progress in addressing inequities in outcomes for historically underserved student populations.

- **Lack of state and district capacity to use flexibility effectively**

ESSA reduces federal mandates and oversight related to both school accountability and teacher quality policies, and it gives states more flexibility to design their own policies and approaches. This creates opportunities for forward-looking states to innovate, but most states and districts continue to have limited capacity, time, and knowledge to do the work. This lack of capacity could lead to low-quality, poorly designed policies and systems or insufficient oversight or support for district implementation.

- **Equity risks and difficulty sharing lessons across states**

As states and districts change teacher evaluation policies with less federal and state guidance, teacher evaluation systems will end up in very different places, which could create equity risks. At the same time, greater variation could provide opportunities for learning if policymakers and practitioners are thoughtful about sharing best practices.

- **Ignoring the larger human capital ecosystem**

As states and districts amend teacher evaluation policies, they should be proactive about connecting these policies to other meaningful parts of the human capital system—a practice they've failed at in past iterations of teacher evaluation policy changes.

Recognizing these risks, this paper offers several policy recommendations:

- 1 Don't rush to action**

ESSA gives states flexibility to change their evaluation policies, but that doesn't mean they have to do so right now. It may be wise to wait until the dust settles on accountability and other issues.

- 2 Preserve a role for student achievement in teacher evaluation systems**

Student achievement measures remain a far more robust measure of teacher quality than many others available. Eliminating student achievement measures from teacher accountability systems is a mistake.

### **3 Consider the relationship between teacher evaluation and accountability systems**

As state policymakers make changes to school accountability and teacher evaluation systems, they should think carefully about the role that each system plays in encouraging continuous improvement in all schools and supporting educational equity for historically underserved students.

### **4 Invest in management and capacity to develop teachers**

Developing school leader capacity is crucial to supporting quality teaching yet schools and districts radically underinvest in leadership capacity. While there are a variety of models for expanding this capacity, all of them require the education field to take the culture of support more seriously and to invest much more in cultivating leaders who manage and support teachers.

### **5 Identify strategies to capture and learn from variation**

The next iteration in teacher evaluation policies will have much greater variation across states and among districts within states. This variation creates opportunities for learning—identifying both what works and what doesn't among different approaches—but only if there are structures and capacity in place to do so.

## Introduction

Five years ago, teacher evaluation was the hot education policy topic. Multiple factors came together to create a path to enact changes in teacher policy with few obstructions. As a result, 28 states enacted teacher evaluation laws requiring that objective measures of student achievement be included in teacher evaluations from 2009 to 2015.<sup>2</sup>

While many of these laws sought to make much-needed changes, the speed of change outpaced technical know-how and capacity in the field, and created collisions among various priorities. In 2012, Bellwether authored a report identifying four key tensions in the direction policies were taking: flexibility versus control, evaluation in an evolving system, purposes of evaluation, and evaluating teachers as professionals.<sup>3</sup> Some of these tensions played out as policy moved to implementation and are still at play today. Others did not, in part because state teacher evaluation laws did not necessarily do all the things they purported to do (or that critics often claimed), and also because states slowed implementation of teacher evaluation laws.

But our fundamental insights remain accurate. Even after policies have been enacted, challenged, and in some cases already amended or abandoned, the debate and conflict over teacher evaluation continues to focus on issues of politics and ideology. The main teacher evaluation conversation is still largely centered on whether it's appropriate to hold teachers accountable for student learning or, more broadly, whether to hold teachers accountable at all. Beyond ideological issues, public debate on teacher evaluation focuses

*The main teacher evaluation conversation is still largely centered on whether it's appropriate to hold teachers accountable for student learning or, more broadly, whether to hold teachers accountable at all.*

primarily on technical issues—implementation challenges, issues of teachers in non-tested grades and subjects, or the validity and reliability of valued-added measures. Key questions related to the purpose of teacher evaluations; the relationship between teacher evaluation, innovation, and other education reform efforts; and the appropriate balance between objective measures and managerial discretion remain largely unaddressed.

Now, as the recently reauthorized federal Elementary and Secondary Education Act ushers in a new era of education policy characterized by greater state flexibility and decreasing federal oversight, teacher evaluation policies may undergo another transformation. Spurred by implementation challenges, political pushback on reforms, and reduced federal emphasis on evaluations, state policymakers may retreat from these policies without learning from the successes and shortcomings of the most recent reforms.

Just as the rapid passage of teacher evaluation policies created risks five years ago, the movement away from them creates risks today. This paper seeks to outline those risks and offer advice on how to mitigate them.

## How We Got Here

Today’s teacher evaluation landscape must be understood in historical context. The push to reform teacher evaluations emerged in response to research showing that teachers are the greatest in-school factor affecting student achievement and that the impacts of individual teachers vary widely.<sup>4</sup> Despite this variation, however, the education system treated most teachers the same, or as widgets, with similar professional development plans, opportunities for growth, and compensation structures.<sup>5</sup> Research also showed that the major factors schools and districts did use to differentiate teachers—certification, years of experience, and postgraduate degrees—had little relationship to teacher quality.<sup>6</sup> Rather, the best predictor of how much a teacher’s students will learn in a given year is how much his or her students gained in previous years.<sup>7</sup> The recognition that existing systems failed to differentiate teachers based on performance led to calls to replace binary, checklist-style teacher evaluations—the norm in most districts—with new evaluation systems based on a teacher’s performance in leading students to academic achievement.<sup>8</sup>

The Obama administration’s Department of Education, led by Secretary of Education Arne Duncan, seized on emerging research to push for changes in teacher evaluation policies. The American Recovery and Reinvestment Act (ARRA)—the \$787 billion stimulus bill passed in July 2009—created an unprecedented opportunity to do so. The legislation included \$4.35 billion for the Race to the Top (RTT) program, a competitive program designed to incentivize states to improve their education policies and systems. ARRA authorized the U.S. Department of Education to make competitive grants to states that promised progress in four “assurance areas,” one of which was making improvements in teacher effectiveness.



The “Great Teachers and Leaders” component of RTT—which accounted for more points than any other section of the application—required states to develop teacher and principal evaluation systems that evaluated all teachers and principals at least annually, included student achievement growth as a significant factor in teacher evaluations, differentiated multiple levels of teacher effectiveness in multiple categories, and used teacher evaluation results to inform key personnel decisions. States that prohibited the use of student achievement data in teacher evaluation were ineligible to win an RTT grant.<sup>9</sup>

*One study found that before RTT, states enacted about 10 percent of the policies that the competition encouraged states to adopt. After? Sixty-eight percent.*

RTT clearly stimulated state action on teacher evaluation policy and other education policy areas. One study found that before RTT, states enacted about 10 percent of the policies that the competition encouraged states to adopt. After? Sixty-eight percent.<sup>10</sup> By the time the first round of RTT applications was submitted in early 2010, 11 states had passed legislation to eliminate statutory barriers to using student achievement data in teacher evaluations, established new standards for school and district teacher evaluations, or created new state teacher evaluation systems. By 2012, 20 had done so. Even if states did not win the RTT money, these policies remained in place after they were enacted.

But RTT wasn’t the only factor driving changes in state teacher evaluation policies. Republican governors and state advocacy groups across the nation were also pushing state-level legislation to overhaul outdated evaluation policies. States like Florida and Tennessee had been at it for years. Florida had been requiring student growth to be included in teacher evaluations for more than a decade before the state passed legislation that required 50 percent student growth in 2011.<sup>11</sup> Similarly, Tennessee’s “First to the Top” legislation, enacted in 2010 in response to RTT, built on teacher evaluation system reforms that the state had already made in 2007.

Both RTT and many state-led teacher evaluation reforms borrowed heavily from the IMPACT evaluation system that Washington, DC Public Schools Chancellor Michelle Rhee had negotiated with the district’s teachers’ union in 2009 and officially signed into contract in 2010.<sup>12</sup> Research shows that IMPACT influenced the voluntary attrition of low-performing teachers and it improved the performance of high-performing ones.<sup>13</sup> (See DCPS sidebar for more.)

In 2012, the Obama administration further incentivized state action on evaluation through the Elementary and Secondary Education Act (ESEA) flexibility waiver process. After a failed attempt to reauthorize ESEA in 2011, the Obama administration initiated the flexibility waiver process to free states from meeting key provisions of ESEA — which hadn’t been updated since 2001. In exchange for flexibility, however, states were required to adopt college- and career-ready standards, focus on 15 percent of their most troubled schools, and create guidelines for teacher evaluations based in part on student performance.<sup>14</sup> Given frustrations with the outdated federal law and the rising number of schools identified for ESEA’s interventions, many states took the bait, furthering the reformed teacher evaluation domino effect.

Today, 43 states require that student growth and achievement be considered in teacher evaluations.<sup>15</sup> After rapid policy adoption from 2010 through 2013, the pace of teacher evaluation change slowed as states moved to implementation. The combination of teacher evaluations with other simultaneous reform efforts—such as the rollout of new standards and corresponding assessments—and unrealistic rollout timelines led to implementation challenges. In recognition of these challenges, Duncan offered waiver states flexibility to delay implementation of student growth or pilot its use in certain schools rather than adopting it statewide.<sup>16</sup>

The federal pressure to implement new teacher evaluations disappeared altogether in December 2015. When Congress and the Obama administration ultimately came together to reauthorize ESEA, the new law, the Every Student Succeeds Act, rolled back many of the policies that incentivized states and districts to change evaluation policies. Though ESSA requires states to have a definition of teacher ineffectiveness, states are not required to implement teacher and principal evaluation systems. In addition, ESSA prohibits the secretary of education from requiring teacher or principal evaluations or putting any parameters around how states define teacher effectiveness.

Even before ESSA, states were already slowing the pace of evaluations or moving away from recently enacted evaluation reforms. New York State offers one example: In 2010, the State Department of Education and teachers' unions came together to design and enact a reformed teacher evaluation system tied in part to student achievement. But the unions' support for new evaluations was short-lived, and union leaders' efforts to undermine evaluations and tie them to an anti-testing backlash<sup>17</sup> have contributed to successive years of legislation undoing previous reforms.<sup>18</sup> Local collective bargaining battles delayed implementation of new evaluations in the 2012–2013 school year.<sup>19</sup> The following year, state legislation prohibited factoring the results of new, Common Core-aligned assessments into teachers' evaluation scores.<sup>20</sup> Then, in early 2016, the Empire State's Board of Regents voted to exclude state tests from teacher evaluations until 2019.<sup>21</sup> Shortly after that, state officials announced a plan to overhaul its teacher and principal evaluations in spring 2017 in pursuit of a completely new system by 2019.<sup>22</sup> After receiving millions in federal dollars to develop and implement a new teacher evaluation system, New York is now poised to invest additional resources in yet another, different system.<sup>23</sup> While New York had its own unique dynamics, across the country both national teachers' unions have played a role in fostering anti-testing sentiment, which fueled an opt-out movement and created both technical and political challenges for performance-based teacher evaluations.<sup>24</sup>

Given results like this, the teacher evaluation drama could have been seen as a lot of fuss over nothing. New evaluation policies not only led to implementation trouble and political pushback but also may not have produced much change in the ratings teachers received. In almost every state where evaluation reform has been implemented, the vast majority of teachers end up with a rating equivalent to effective or higher. In the 2013-2014 school year, 92 percent of teachers in New York, 97 percent of teachers in New Jersey, and 99 percent of teachers in Delaware were rated effective or higher.<sup>25</sup> In the 2014-2015 school year, 88 percent of teachers in Indiana were rated in the state's top two categories and just under two percent were rated in the bottom two categories (approximately 10 percent of teachers were not rated).<sup>26</sup>

Was all the effort to enact and implement new teacher evaluation systems worth it? Answering that question requires a deeper look at the results of new teacher evaluation systems—both positive and negative.

## Sidebar 1

### The Evolution of DCPS IMPACT

Few teacher evaluation systems have received more attention—or garnered more controversy—than the District of Columbia Public Schools' (DCPS) IMPACT. Created in 2009, IMPACT was one of the first teacher evaluation systems in the country to make high-stakes decisions based on performance, including providing significant bonuses and salary increases to highly effective teachers as well as dismissing ineffective teachers.

Initially under IMPACT, teachers were evaluated by three main components: student achievement data as measured by value-added measures (approximately 15 percent of teachers) and student learning objectives (all teachers), instructional expertise as measured by observations conducted by school leaders and independent evaluators, and commitment to school community as measured by school-based metrics. Depending on teachers' overall evaluation score, they could be awarded a bonus of up to \$25,000 or be dismissed from the district.<sup>i</sup>

IMPACT has evolved since its initial implementation. In 2012, DCPS reduced the weight of value-added measures in evaluations for teachers in tested grades and subjects from 50 to 35 percent. The weight of student achievement remained at 50 percent, with the remaining 15 percent measured by student learning objectives. In addition, in both the 2014–2015 and 2015–2016 school years, state-administered student assessment scores were not included in evaluations at all as the district transitioned to Common Core-aligned assessments.

The shift to Common Core standards, and the higher expectations they set for teachers and students, drove DCPS to make broader changes to IMPACT. Starting in the 2016–2017 school year, DCPS made four major adjustments to IMPACT. First, student surveys are now added to teachers' overall evaluation score for teachers of third grade and

*Continued on next page*

above. Also, teachers are no longer observed by independent evaluators; instead, all formal observations are conducted by school leaders. In addition, the observation rubric used by principals is now reduced from nine to five indicators of exemplary instruction, or what DCPS calls “essential practices.” Lastly, after a two-year hiatus, state-administered student assessment scores are once again included for teachers who teach in tested grades and subjects.

In combination with changes to IMPACT, DCPS also rolled out a new professional learning system called Learning Together to Advance Our Practice, or LEAP. DCPS’ teachers, school leaders, and central office created LEAP to provide better professional learning opportunities for DCPS teachers so they would have the support necessary to teach to the more rigorous Common Core standards and assessments.

LEAP differs from traditional, one-time workshop-style approaches to professional development that research shows have little positive effect on teacher instruction.<sup>ii</sup> Under LEAP, DCPS teachers participate in weekly 90-minute group sessions with teachers of the same subject, led by a LEAP leader with a track record of success in that subject. LEAP provides support for teachers to develop their content knowledge and instructional mastery, as well as to collaborate on lesson plans and teaching methods in a group setting on a consistent basis.

Under LEAP, teachers are observed informally by LEAP leaders for periods of 15 minutes. These observations are not included in IMPACT evaluation ratings. Rather, LEAP observations are meant to offer an informal look at how teachers respond to feedback from the weekly group professional development time. After the observations are conducted, LEAP leaders debrief with teachers about the specific instructional practice teachers were working on and provide feedback.

Importantly, IMPACT and LEAP are two separate systems, but it is no coincidence that changes to IMPACT happened simultaneously with the introduction of LEAP. “We decided to let IMPACT be the measure of performance and that LEAP would be our primary conduit for development,” says Jason Kamras, chief of instructional practice at DCPS. “IMPACT helps us to do talent management things. It certainly informs LEAP. But LEAP is about the day-to-day work of building one’s practice and honing one’s craft so that we see outstanding instruction.”

DCPS has plans to measure both the quality and effectiveness of LEAP implementation across the district, and conduct an analysis of if and how LEAP leads to positive changes in teacher practice and student achievement. DCPS leaders will also have academic researchers study LEAP’s longer-term impact on teachers and students, and will publish the results.

A research study of IMPACT from the 2009–2010 through 2011–2012 school years shows that the teacher evaluation system produced positive results. The study found that IMPACT leads to significant instructional improvement of teachers. It also found that under IMPACT, low-performing teachers left the district at higher rates than high-performing teachers did, and that the teachers who replaced those who left the district led students to higher student achievement results. Specifically, these teachers drove student achievement gains by approximately four months of learning in both reading and math.<sup>iii</sup> The effects of IMPACT on teacher practice and quality may have translated to gains

*Continued on next page*

## Sidebar 1 *continued*

in student learning: Since the rollout and implementation of IMPACT, student achievement in DCPS has increased. According to 2015 data from the Trial Urban District Assessment (TUDA), DCPS is the fastest improving urban school district in the country.<sup>iv</sup>

D.C. has significant work to do, however, both to raise its overall student achievement to match national averages and to better develop its teachers' effectiveness. In its original iteration, IMPACT drove improvements in teacher effectiveness by changing the composition of the workforce. Now, with LEAP and the new iteration of IMPACT, DCPS is trying to accelerate development and growth among its existing teachers. The spotlight will continue to shine on the district as it embarks on this endeavor.

- i "Compensation: LIFT and IMPACTplus," DCPS, accessed October 16, 2016, <http://dcps.dc.gov/page/compensation-lift-and-impactplus>.
- ii "The Mirage: Confronting the Hard Truth about Our Quest for Teacher Development," TNTP, August 4, 2015, <http://tntp.org/publications/view/the-mirage-confronting-the-truth-about-our-quest-for-teacher-development>.
- iii Thomas Dee and James Wyckoff, "Incentives, Selection, and Teacher Performance: Evidence from IMPACT," *Journal of Policy Analysis and Management* 34, no. 2, (2015): 267–297, <http://cepa.stanford.edu/content/incentives-selection-and-teacher-performance-evidence-impact>.
- iv "DC Public Schools Continues Momentum as the Fastest Improving Urban School District in the Country," *District of Columbia Public Schools*, October 28, 2015, <http://dcps.dc.gov/release/dc-public-schools-continues-momentum-fastest-improving-urban-school-district-country>

# The Good and Bad of Teacher Evaluation Reform

*Reformed systems have given school staff opportunities to engage in meaningful conversations about instructional practice that were not possible with old, checklist-style evaluations.*

## The Good

As states and districts prepare to make more changes to teacher evaluation systems, it's worth taking time to reflect on what we've learned from the most recent round of reforms. While some data points suggest that the impact of teacher evaluation has been minimal, other data suggest that new policies have produced real benefits.

One major benefit is that reformed systems have given school staff opportunities to engage in meaningful conversations about instructional practice that were not possible with old, checklist-style evaluations. A pilot study in Chicago found that new teacher evaluation systems helped principals and teachers articulate effective teaching practice and communicate about the kinds of instruction that matters most for student achievement.<sup>27</sup> Similarly, Tennessee educators report having a clearer understanding of what constitutes effective teaching and that regular and specific feedback leads to increased self-reflection and focus on instructional improvement among teachers.<sup>28</sup>

These open lines of communication ushered in opportunities to break down decades-old barriers between teachers and principals to begin the conversation about effective instructional practice. While this continues to be a work in progress in most schools and districts, some are experiencing success. 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.<sup>29</sup>

Research on the impact of teacher evaluation on the quality of the teacher workforce is limited, but there is some evidence that reformed district teacher evaluation systems helped improve overall teacher quality. A recent study of a Chicago teacher evaluation pilot found that the evaluation system influenced low-performing teachers to leave the profession, while retaining effective ones.<sup>30</sup> Similarly, a study of IMPACT, DC Public Schools' teacher evaluation system, found that IMPACT influenced the voluntary attrition of low-performing teachers and improved the performance of high-performing ones.<sup>31</sup>

*It's true that very few teachers are rated ineffective in the new systems. But focusing on ineffective teachers alone ignores the greater variation in ratings in new systems.*

It's true that very few teachers are rated ineffective in the new systems. But focusing on ineffective teachers alone ignores the greater variation in ratings in new systems. In contrast to previous binary systems, the new systems produce more nuanced ratings, leading to more meaningful differentiation of teacher performance. At the high end of the performance continuum, some states are beginning to differentiate between "effective" and "highly effective" teachers.<sup>32</sup> In New Jersey, for example, 73.9 percent of teachers were rated "effective" in the 2013–2014 school year and 23.4 percent rated "highly effective." On the lower end of the performance continuum, states have added ratings between "ineffective" and "effective."<sup>33</sup> A recent study of teacher performance ratings across 19 states found that the median percentage of teachers rated below "proficient" was 2.7 percent, up from just a fraction of a percentage five years earlier.<sup>34</sup> This type of differentiation is important because it allows education leaders to identify teachers who need extra support and provide them with professional development targeted to specific gaps in their knowledge and skills. Differentiation of highly effective teachers also enables systems to provide specific retention incentives or leadership opportunities to help keep these teachers in the classroom or enable them to have broader schoolwide impacts.

Do the new ratings reflect the distribution of teacher performance on reliable measures of student learning? It's too soon to tell. However, most observers should agree that the variation in teacher performance is broader than what a binary effective/ineffective distinction can capture, and that a system that rates close to 100 percent of teachers effective is probably not accurate. The act of differentiating ratings then is a step toward identifying a distribution of teaching effectiveness that better reflects the level of instruction students receive throughout the nation's classrooms.

Another clear benefit of new teacher evaluation systems is that they've encouraged states to extend data system capacity to link teacher and student data. Many states collect data on students and teachers, but until recently, matching teachers and students by course at the state level was not common practice. The 2002 iteration of the ESEA—No Child Left Behind—had annual testing requirements in grades three through eight that generated abundant student achievement data, but only a few states linked these data back to students' teachers. Research on teacher variance led to the desire to develop more robust teacher evaluation systems—which required linking teacher and student data. At the

same time, the U.S. Department of Education's Statewide Longitudinal Data Systems Grant Program provided funds to states to strengthen their K-12 longitudinal data systems. By 2013, 45 states had the capacity to link teacher and student data.<sup>35</sup> These linkages are crucial not just for teacher evaluation but also for building our understanding of the connection between student academic growth and teacher training, qualifications, and practice.<sup>36</sup>

## The Bad

*The speed with which states and districts moved on teacher evaluation reform created obstacles, leaving key stakeholders unhappy and important tenets of the reforms vulnerable to attack.*

On the other hand, the speed with which states and districts moved on teacher evaluation reform created obstacles, leaving key stakeholders unhappy and important tenets of the reforms vulnerable to attack. One glaring concern is that new teacher evaluation policies have increased the polarization around using outcomes-based measures as part of teachers' evaluation. Successful teachers should lead students to academic results, and education leaders should have a way to track student results back to teachers in a way that feels safe and fair to teachers.

But the story is not that simple—many teachers currently view performance-based teacher evaluation systems as a mechanism to harm them. This is true, even though few teachers' jobs have been threatened due to changes in evaluation systems. In New York, for example, as of late 2015, only one tenured teacher had been fired through the evaluation and dismissal process.<sup>37</sup> Despite this, several recent surveys show that teachers are largely skeptical about the use of student achievement in evaluation.<sup>38</sup> There seems to be a fundamental disagreement between some policymakers and practitioners about whether student results provide a meaningful reflection of teacher quality. This is particularly an issue when practitioners do not trust state tests or feel that they measure only a narrow part of students' learning and development. Discomfort with use of student achievement in teacher evaluation systems has driven opposition to the new systems even though all of these systems use multiple measures to evaluate the impact of teachers' practice, with student achievement as just one component. And as teachers and teachers' unions have criticized the use of student results in teacher evaluation systems, some states have reduced the weight of these measures in teacher evaluations or eliminated them altogether.

In 2012, when states were moving rapidly to enact and implement teacher evaluations, Bellwether's Sara Mead, Andy Rotherham, and Rachael Brown argued that overly prescriptive or poorly designed evaluation systems might create a barrier to educational innovation:

*Unless state policies provide for additional flexibility around teacher evaluation in blended learning and other innovative approaches to schooling, we will miss out on the opportunity to develop these new forms of evaluation. More troubling, teacher evaluation requirements could actually become a barrier to the expansion of blended learning models that de-link*



*student learning from individual teachers, or to the development of new models combining in-person teaching with technology and other delivery mechanisms to personalize student learning experiences. Charter school authorizers may be unwilling to approve schools using new blended models if those schools cannot explain how they will comply with state teacher evaluation laws.<sup>39</sup>*

It's not clear that this played out in practice. Some leaders implementing innovative models report that evaluation pressures have made it more difficult to build teacher buy-in for new instructional approaches. When teachers know they will be evaluated based on classroom observations and student growth, they may prefer to stick with instructional approaches that they are comfortable implementing and know will result in positive ratings rather than risk new strategies they have less experience with. Yet there is little hard evidence that evaluation has delayed the spread of personalized learning models. Indeed, with encouragement from federal initiatives such as the Race to the Top-District competition<sup>40</sup>, adoption of state laws that promote competency-based learning and school progression, and emergence of new organizations supporting implementation of personalized learning models, more districts and schools have adopted elements of personalized learning over the past five years. Given that evaluation requirements have been implemented slowly in many states, that districts in most states have considerable flexibility in how they design evaluation systems, and that relatively few teachers have faced negative consequences from evaluation, the potential negative impacts may have been overstated.

*If personalized learning models are to truly transform students' learning experiences, they must also fundamentally transform how teachers do their work.*

Still, it is cause for concern that, even as state and federal policies created incentives for districts and schools to implement more personalized learning approaches, these policies have not encouraged districts to adopt new approaches to evaluation adapted to personalized learning contexts. And districts implementing personalized learning have not developed new models of assessing teacher performance that reflect the different ways teachers work in new instructional models.<sup>41</sup> If personalized learning models are to truly transform students' learning experiences, they must also fundamentally transform how teachers do their work. And that should also mean changing how that work is evaluated.

## Achievement First Teacher Career Pathway and Evaluation System

Achievement First is a charter management organization (CMO) that operates a network of 32 public charter schools in New Haven, Bridgeport, and Hartford, Connecticut; Brooklyn, New York; and Providence, Rhode Island, all of which serve 11,600 students in grades K–12. Achievement First began developing its teacher evaluation system in 2007 in response to teachers voicing a desire for pathways to grow and develop as professionals without becoming principals or school leaders.

The design of Achievement First’s evaluation system reflected this focus on teacher development and growth. Teachers, principals, and network leaders collaborated to create the Teacher Career Pathway. They started by outlining each stage of a teacher’s trajectory: intern, new teacher, teacher, distinguished teacher, and master teacher. After that, they defined what excellent instruction would look like along those steps, which they would measure against an in-house created observation rubric. The career pathway stages then informed Achievement First’s approaches to both teacher evaluation and professional development. Each stage of the career pathway is accompanied by increased compensation and leadership opportunities, with distinguished and master teachers receiving additional professional development and network-wide recognition.

Achievement First’s teacher evaluation system is based on four elements:

- Student academic growth as measured by external or internal assessments (as applicable).
- Student character development as measured through student and parent surveys.
- Quality instruction as measured by classroom observations rated against an in-house created rubric.
- Core values and contributions to team achievement as measured through principal and peer surveys.

Student growth and quality instruction each count for 35 percent of teachers’ evaluation scores, and the final 30 percent is made up of a combination of student, family, peer, and leader surveys.

As Achievement First leaders developed the network’s teacher evaluation system, teachers also expressed a desire for aligned professional development opportunities so they could receive support in advancing along the career pathway. Achievement First invests significant resources in teacher professional development. It employs teacher coaches who work directly with teachers on specific needs that help them improve in their practice. In addition, each Achievement First teacher receives weekly non-evaluative observations from his or her coach, followed by a 30- to 45-minute debrief. School leadership teams also lead weekly professional development sessions at school sites, where teachers focus on the instructional skills they need to create strong classroom cultures and intellectually prepare to teach upcoming lessons and units. There are also several structured network-wide learning days throughout the year, where teachers dive deep into content knowledge and best practices. Teachers and school and network leaders also collaborate to analyze student data in order to plan instruction that meets every student’s needs and ensures that all children are on a path to success.

*Continued on next page*

## Sidebar 2 *continued*

The focus on professional development as part of the teacher evaluation system at Achievement First seems to have paid off. A recent TNTP study of the effectiveness of teacher professional development found that seven out of 10 Achievement First teachers studied made substantial growth in their practice, compared to an average of only three out of 10 teachers in other districts included in the study.<sup>i</sup> When compared to surrounding schools, Achievement First teachers also made a greater impact on their students' learning from year to year. And overall test scores in math and reading are higher across the Achievement First network than in surrounding schools.

Despite these successes, Achievement First leaders recognize they still have room to improve the teacher evaluation system. After several years of implementation, Achievement First leaders took a step back and did a listening tour of leaders and teachers across the network to see how they could strengthen the evaluation system and align professional development. These leaders found that teachers didn't fully understand how the teacher evaluation system tied into their weekly coaching and that, due to this, in some cases teachers were surprised by the evaluation scores they received. School leaders also noted that the teacher evaluation system took more time to complete than they would have preferred, and some teachers shared that they did not understand the value of having an external reviewer conduct observations of their practice.

Based on this feedback, Achievement First made several changes to the teacher evaluation system beginning in the 2016–2017 school year. To bring more clarity to what excellent instruction looks like and to simplify the evaluation process, the teacher evaluation rubric was condensed from 30 pages to five pages. Achievement First leaders also better aligned the rubric to the skills and knowledge teachers should display throughout the different parts of the school year, or what Achievement First calls the “arc of the year.” Each arc is about six weeks long. Throughout the year, coaches work with teachers on the parts of the evaluation rubric that align to the arcs.

Achievement First has also eliminated the use of outside observers, and instead now uses coaches' and school leaders' observations in teacher evaluations. In addition, Achievement First is working with Panorama Education to improve the student survey piece of the teacher evaluation system. The Panorama Education survey is a nationally normed survey, so the Achievement First survey data can be compared with not only other network schools but also more than 6,000 schools nationwide.

Achievement First continually reassesses whether its teacher evaluation system is achieving the desired result of improved instructional practice. The CMO's experience illustrates the importance of listening to leaders' and teachers' experiences throughout implementation, paying attention to lessons learned over time, and making changes in response to those lessons. This, in turn, requires creating space for teacher evaluation systems to evolve to meet the needs of the system and its teachers.

i “The Mirage: Confronting the Hard Truth about Our Quest for Teacher Development,” *TNTP*, August 4, 2015, <http://tntp.org/publications/view/the-mirage-confronting-the-truth-about-our-quest-for-teacher-development>

## Risks Today

The flexibility that ESSA affords states and districts around teacher evaluations creates an environment ripe with new risks. Supporters and critics of new teacher evaluation systems should be aware of these risks when considering changes in teacher evaluation policies.

### **Shifting Focus to Professional Development Without Considering the Design Features and Incentives Needed to Make Evaluations Useful for This Purpose**

With the passage of ESSA and the move toward local control, many policymakers and advocates have voiced support for shifting the focus of teacher evaluation systems from making employment decisions to influencing teacher professional development opportunities. A report by the Washington, DC-based think tank, New America, set forth ideas for “re-envisioning state teacher evaluation systems as tools for professional growth.”<sup>42</sup> The Aspen Institute created a “roadmap” for teacher evaluation improvement, with the first recommendation being “prioritize principal and evaluator training and certification with a focus on professional growth.”<sup>43</sup> And the Council of Chief State School Officers released a report on teacher evaluation systems, renaming them “Teacher Support and Evaluation Systems.”<sup>44</sup> Current evaluation systems built around a small number of observations and little feedback are not likely to contribute much to teacher professional growth, however. Instead, prioritizing professional development would require a completely revised teacher evaluation system created and implemented with a focus on professional development at the forefront.

Moreover, it's not clear that we have the research or knowledge base to know how to create evaluation systems that support professional development that leads to improved practice and student learning. Research suggests that most current teacher professional development systems do little to improve teachers' practice.<sup>45</sup> And without teacher evaluations or other objective ways to assess teacher performance, it will be difficult to know if professional development affects the quality of instruction.

If teacher evaluations are no longer used to inform employment decisions, advocates and policymakers must consider how the loss of accountability "teeth" will affect practitioners' implementation of and response to these systems. Experience in systems that are leading thinking about how teacher evaluations contribute to educators' professional development shows that building these systems requires significant, intentional investments in design and capacity (see *Achievement First sidebar on page 19*). It's not clear that school systems will feel the urgency to make these investments in the absence of external pressure.

*If teacher evaluations are no longer used to inform employment decisions, advocates and policymakers must consider how the loss of accountability "teeth" will affect practitioners' implementation of and response to these systems.*

### **Impact of Eliminating Student Achievement Measures**

Given increased flexibility in ESSA and political pushback, new teacher evaluation laws and policies have already begun to experience change. In the 2016 legislative session, legislators in states from Alaska to Georgia introduced bills to decrease the weight of, or eliminate, student achievement as a state-mandated factor in teacher evaluations.<sup>46</sup> And large school districts, such as Los Angeles Unified School District, are making changes to the use of student outcomes in teacher evaluations.<sup>47</sup>

Because many of the policies that require the use of student achievement are written into state law, changes in these policies will happen over time, not immediately. While it's unlikely that all states will eliminate the role of student growth in their evaluation policies, even a few states reversing course will make it difficult for advocates to protect measures of student achievement in teacher evaluation policy. The risk here is that outcomes-based measures, such as the amount of impact teachers have on student learning, are far better measures of teacher quality than the inputs-based measures the system used before teacher evaluation reform. Losing student learning measures could negatively affect the systems' ability to identify effective teaching or ensure equitable access to quality teaching for underserved students.

In place of student learning measures, states may choose to place greater weight on other measures, such as teacher observations. However, studies show that classroom observations alone are unreliable at identifying effective teachers' practice.<sup>48</sup> Not only that, but observations are also expensive and logistically difficult to implement.<sup>49</sup> One study of three districts found that observations were by far the most expensive component in current teacher evaluation systems. In Memphis City Schools (now Shelby County School District), for example, activities related to classroom observation were the largest source of expenditure at 82 percent, student surveys represented 17 percent, and the activities to produce student growth measures connected to teacher performance accounted for just one percent.<sup>50</sup>

Some districts and states are reducing observation costs by differentiating how often teachers are observed. For example, a district may observe high-performing or veteran teachers less frequently than lower-performing or novice teachers. Twenty-seven states require multiple observations for some teachers compared to 11 states that require multiple observations of all teachers.<sup>51</sup> While this strategy might save money, it will inevitably devalue the evaluations of higher-performing teachers for professional development purposes. This move contradicts the idea that the main focus of teacher evaluation systems in a post-ESSA world should be teacher professional development.

Instead of eliminating the use of student growth in teacher evaluation systems, some experts have proposed different ways of including the metric. Researcher Doug Harris suggested using student achievement metrics only as an initial screen to identify top and bottom performers for purposes of career advancement at the top and remediation at the bottom.<sup>52</sup> And Tom Kane advised using student achievement results in the evaluation of non-tenured teachers while reducing the role of achievement results for tenured teachers.<sup>53</sup> Both proposals keep the use of student achievement as part of the conversation for some teachers but at the same time water down the metric and rely too heavily on other parts of the system—such as the tenure process—to identify teachers who consistently lead students to academic success. These recommendations may offer an appealing option for state and district policymakers who are facing pressure to deemphasize student achievement but want to keep it as one part of the teacher evaluation system. They fall short, however, in grounding the teaching profession in outcomes-based measures that research shows are the best predictors of teacher effectiveness.

## Loss of Accountability

The effort to improve teacher evaluation systems was also partly about changing the locus of accountability for student improvement. Policy changes in recent years shifted the emphasis of accountability from schools to individual teachers. No Child Left Behind expected all schools to make “adequate yearly progress” but did not establish accountability metrics for individual teachers. Rather, the law assumed that imposing accountability pressures on schools would create incentives for teachers and principals to improve instruction. ESEA waivers, in contrast, allowed states to focus intervention efforts on just 15 percent of the lowest-performing schools, effectively exempting the majority of schools from accountability pressure. At the same time, however, it reached all schools through the expectation that all teachers would lead students to academic growth.

Reasonable people can disagree about the advisability or fairness of shifting accountability for academic growth down from the school to the teacher level. ESEA waivers focused on teachers because of research documenting the enormous influence teachers have on student learning. Research also shows that there is wide variation in effectiveness among teachers even within the same school.<sup>54</sup> Critics of teacher evaluations, however, argue that teachers

*Reasonable people can disagree about the advisability or fairness of shifting accountability for academic growth down from the school to the teacher level.*

do their work in the context of larger school and system cultures, and that factors beyond teachers' control have a great deal of influence on children's learning results. This argument is not limited to those who question the value of standardized tests and accountability more generally. Some have argued that, rather than mandating specific teacher evaluation policies, a better approach to ensuring quality teaching may be to hold school leaders accountable for school performance while giving them greater flexibility and control over hiring, dismissal, and professional development.<sup>55</sup> This combination of accountability and flexibility is not the reality in most schools today, however. Moreover, placing responsibility for teacher evaluation in the hands of individual leaders could exacerbate inequities for teachers and students across schools. The lack of a common definition of teacher effectiveness could also make it more difficult for schools to learn from one another or collaborate on strategies to improve instruction or student achievement.

*In a post-ESSA world, there is a risk that neither schools nor teachers will be truly accountable for student achievement results in most schools.*

Now under ESSA, states may decide to hold neither schools nor teachers accountable for student learning. ESSA explicitly requires states to identify only five percent of low-performing schools (states may identify more if they choose) and does not require that states and districts evaluate teachers for their contributions to student growth. Therefore, in a post-ESSA world, there is a risk that neither schools nor teachers will be truly accountable for student achievement results in most schools. Moving away from both teacher and school accountability at the same time could significantly reduce urgency to improve outcomes for chronically underserved students and subgroups—ultimately leading to increased educational inequity and backsliding on recent progress narrowing achievement gaps.

### **Lack of State and District Capacity to Use Flexibility Effectively**

When new teacher evaluation reform efforts were being rolled out, numerous observers voiced concerns that prescriptive policies would stifle districts' ability to customize the systems to meet unique needs. As Rick Hess and Linda Darling Hammond wrote in 2011, "To existing mandates, they [Democratic reformers] would add heavy-handed, unproven teacher-evaluation requirements that could stifle innovative teaching and school design."<sup>56</sup> But these concerns did not bear out in practice. In fact, most districts have had the ability to shape teacher evaluation systems. Only nine states created statewide systems, while 30 states encouraged locally designed systems that meet certain requirements and characteristics.<sup>57</sup> The problem is that districts have limited capacity, time, and knowledge to do this work.

Indiana is a prime example of a state that provides a lot of flexibility and room at the local level to innovate with teacher evaluation systems, but it has ended up with mixed results. In 2015 and 2016, Indiana districts implemented more than 200 different evaluation systems. Researchers found that the majority of those systems scored in the medium range of what makes for "high-quality teacher evaluation plans." Only 31 districts' evaluation plans got high marks, with more—33—falling in the low marks category. Depending on where a

teacher is employed in Indiana, she will have a vastly different evaluation experience than her peers elsewhere in the state. And, most likely, the system that evaluates her will not be considered high quality.<sup>58</sup>

Indiana's teacher evaluation landscape is a hodgepodge of systems that look different from one district to the next and may or may not meet best practices criteria. This happened even with federal oversight. In the post-ESSA world, with oversight of teacher evaluations left to capacity-strapped states and districts, it's likely that many states' teacher evaluation systems will look like Indiana's, and some may even be more incoherent.

## Equity Risks and Difficulty Sharing Lessons Across States

Even though many state and district teacher evaluation reforms used guidance from RTT and ESEA waivers to develop policy, states and districts are in very different places in the design and implementation of those systems. Some teacher evaluation systems have been fully implemented for multiple years, and state and district leaders are merely refining and leveraging those evaluation results to influence other aspects of teaching and learning. Other places haven't even begun implementation. In a post-ESSA world, one sure bet is that states and districts will end up in very different places when it comes to teacher evaluation policies and implementation.

This variation could, in turn, exacerbate inequities in access to quality teaching for low-income and historically underserved student groups. Districts with low capacity and proportionally greater challenges are less likely to have the bandwidth necessary to make strong reforms. That's true in all environments, but especially when there's little pressure on them to do so.

At the same time, however, greater variation could provide opportunities for learning if policymakers and practitioners are thoughtful. To mitigate the risks and take advantage of the learning opportunities, policymakers, advocates, and education leaders must work to create mechanisms to capture lessons from different approaches and experiences. The absence of such measures could leave the field without opportunities to learn from past mistakes and productively evolve over time.

## Ignoring the Larger Human Capital Ecosystem

Teacher evaluation alone is neither a substitute nor a sufficient catalyst in itself for a more comprehensive rethinking of human capital systems more generally. There has been tremendous movement on connecting teacher evaluations to employment decisions, but states have been less successful linking these policies to other meaningful parts of the human capital system. Only nine states use evaluations to determine licensure, and just seven states directly tie teacher compensation to teacher evaluation results.<sup>59</sup>

*To mitigate the risks and take advantage of the learning opportunities, policymakers, advocates, and education leaders must work to create mechanisms to capture lessons from different approaches and experiences.*



Advocates and policymakers often assumed that teacher evaluations would provide information that could be used to drive change in other parts of the system, but the data produced by the current evaluation regimes haven't been useful for that purpose. Take teacher preparation: The lack of differentiation among teachers also limits an ability to differentiate among teacher preparation providers. As our colleagues Chad Aldeman and Ashley LiBetti Mitchel observed, if evaluations better assessed teacher quality, then we could more precisely differentiate teachers and, by extension, the programs that trained them.<sup>60</sup>

The increased identification of highly effective teachers does create opportunities to use evaluation systems as the foundation for strategies to retain high-performing teachers or cultivate them as leaders in their schools. But states and districts have devoted relatively little energy and resources to developing retention incentives or career advancement opportunities for high-performing teachers. Moreover, it's not clear that the levers states and districts have tried—such as compensation—actually address what drives teachers on a day-to-day basis. Using evaluations to support and retain high-performing teachers requires a more sophisticated understanding of incentives for teachers, as well as the levers through which state and district policies could affect those incentives.

Another problem is the failure of new teacher evaluation systems to invest in developing the leadership capacity needed to make evaluation policies work. States and districts used a variety of structures to implement new teacher evaluations, but many put significant new demand on principals to observe and evaluate teachers—without necessarily building their capacity or changing their roles to create more time to carry out these new responsibilities. This mistake exacerbated principals' reflections that the job has become too complex.<sup>61</sup> States and districts must think carefully about how to develop leadership capacity to observe teacher practice and support teacher development. This may mean rethinking principals' roles or distributing responsibilities that have traditionally been held by one leader across a larger leadership team.

Dismantling current teacher evaluation systems will not eliminate the problem of addressing the larger human capital ecosystem. New teacher evaluation systems must be able to connect the disparate parts of the human capital pipeline, from teacher preparation all the way through to compensation. If new teacher evaluation systems cannot be charged with this task, policymakers need to conceive of a different way of connecting the disparate parts of the system. This would require a radical rethinking of current human capital practices that few seem to have the energy for, given other demands. At the same time, policymakers must realize that teacher evaluation systems—and, indeed, any education policy reform—are only as good as the leaders charged with implementing them. Improving teacher quality requires building leadership capacity at all levels of the system, as well as designing systems to operate in ways that provide space and autonomy for leaders to be effective.

*Dismantling current teacher evaluation systems will not eliminate the problem of addressing the larger human capital ecosystem.*

## Overarching Lessons for Education Reform

Another critical risk is failing to learn from experience with teacher evaluations in recent years and mistakes that contributed to their current state. The below lessons are not just relevant to teacher evaluation but also can inform education reformers in thinking about how to shift the focus in the post-ESSA era from broad federally driven and multi-state policy movements to thinking about how to make positive impacts in unique contexts across states.

### Clearly Identify—and Communicate—the Theory of Action Behind Proposed Reforms

Part of the reason that pushback on teacher evaluations has been effective is that advocates did not clearly define a theory of action; they argued that teachers are important and that new evaluations would lead to improvement in the quality of teaching but were not always clear about the mechanisms through which they would do so. Some messengers, such as Secretary Arne Duncan, emphasized the use of evaluation for professional development in speeches to teachers. Others, such as New Jersey Governor Chris Christie, pitched teacher evaluations as a way to get rid of underperforming teachers by focusing on their link to teacher tenure.<sup>62</sup> Duncan and Christie could have been highlighting two facets of the same system, and neither would have been wrong, but the combination of competing messages and the rhetoric used by some teacher evaluation supporters enabled opponents to define teacher evaluation as primarily a punitive measure targeting teachers.

Some state teacher evaluation laws also jumbled together policies on teacher evaluation, tenure, dismissal, layoffs, and other issues without considering whether the timeline for the implementation of those efforts may conflict or how communication of those efforts may affect various stakeholder groups. Going forward, states and districts need to think about how teacher policies work together and the purposes they are trying to achieve through evaluations, and then thoughtfully design systems with multiple and evolving purposes in mind. They also need to more proactively think through their full theory of action underlying teacher evaluations and other complementary reforms, and how they do or do not work together to improve student outcomes.

*Continued on next page*

## Consider the Intersection of Policies

Teacher evaluation implementation was not happening in a vacuum. It bumped up against Common Core State Standards implementation. Changes in assessments created technical challenges for teacher evaluation implementation timelines and sparked backlash from teachers' unions and parents.

These events created a firestorm around teacher evaluation. Even though it turned out that changes in standards and assessments didn't have much impact on teacher evaluation ratings,<sup>63</sup> the narrative moved beyond the reality. Consideration of the intersection of other reforms and teacher evaluation policy from the beginning may have mitigated some of these issues.

## Culture Eats Policy for Breakfast

In retrospect, future education historians will likely view RTT and ESEA waivers as an Obama administration experiment to see if the culture of education systems could be changed through policy. New teacher evaluation policies sought to move school systems toward more performance-based cultures. They reflected an assumption that policies could, by defining and measuring what good practice looked like, clarify to teachers what the district and state expected them to do in their classrooms every day and incentivize teachers to develop their own capacity to implement good practice. They also assumed that specific protocols for teacher evaluations would overcome a culture of reluctance to give hard feedback about performance—rather than that evaluators would find ways to bend protocols to avoid having hard conversations.

Policymakers may have underestimated the ability of policy to change culture. Changes in evaluation policy alone did not transform a culture that was resistant to frank discussions about teacher performance.

## Policies Must be Adaptable

Many new teacher evaluation requirements have been put into legislation rather than regulation. Legislation certainly protects the changes, but it also locks in the current state of thinking, making it difficult to adapt to new demands. New teacher evaluation reforms were happening at a particularly odd time of political agreement, which lawmakers and advocates recognized and took advantage of. But now, with the passage of ESSA and more state flexibility, there will likely be many battles over teacher evaluation laws in statehouses across the nation. Instead of iterating on a flexible regulation, policymakers will spend double the time or more fighting to reverse or amend legislation—further stalling implementation of new systems. Politics create incentives to enact policies in forms that stick, but in education, a culture of continuous improvement requires openness to failure, learning, and flexibility.

## Policy Recommendations

Given these lessons from the most recent round of teacher evaluation reforms, how should policymakers act today to make the most of current teacher evaluation policies and fix issues that need to be addressed without undermining progress or creating new problems? Because of the wide variation in current state contexts, and our own mixed record predicting the future of teacher evaluation policies, we are reluctant to offer prescriptive recommendations. The following recommendations, however, can help state policymakers avoid mistakes in the post-ESSA landscape—and aid forward-looking districts, advocates, and philanthropic funders in taking advantage of opportunities for innovation and learning that this new landscape creates.

### 1 Don't rush to action

ESSA gives states flexibility to change their evaluation policies, but that doesn't mean they have to do so right now. Especially because the dire predictions of teacher evaluation reform critics—mass firing, scores of exiting teachers, demoralization of the profession—didn't come to pass.<sup>64</sup> There are other good reasons for states to move slowly in enacting major changes in teacher evaluation policies: If the rapid pace of legislation contributed to implementation problems in the past, there's no reason to compound the error by rushing to enact legislation now. In many states, new evaluation policies have barely been implemented, leaving little time to learn from implementation to inform future policies. Under ESSA, states are making other major changes—particularly to their school accountability systems—that may have implications for the design of new teacher evaluations. And simultaneously changing several major state policies could create

new problematic interactions between teacher evaluation and other policy initiatives. Moreover, the same ESSA flexibility that allows states to change their evaluation policies also means they don't have the same tight timeline to make these changes that they do for revisions to accountability systems. If states have flexibility to change their teacher evaluation policies at any time, it may be wise to wait until the dust is settled on accountability and other issues.

## **2 Preserve a role for student achievement in teacher evaluation systems**

In light of political pushback, some states and districts are moving to eliminate student achievement measures from teacher accountability systems. This is a mistake. Despite the political narrative, student achievement measures remain a far more robust measure of teacher quality than many others available. Moreover, including student achievement in teacher evaluation systems, even if as only one of multiple measures, sends a clear signal that quality teaching is defined in part by impact on student learning. States may well want to reduce the role of student achievement measures in their evaluation systems—but they should not eliminate them entirely. Even if states do choose to end policies that require the use of student achievement in teacher evaluations, they should not create barriers for districts or schools that want to continue to use these measures in their evaluation systems. Moreover, states should continue to provide districts the data needed to do so.

## **3 Consider the relationship between teacher evaluation and accountability systems**

As noted above, ESEA waivers replaced accountability for all schools with accountability for all teachers. Under ESSA, states have much more flexibility in both school accountability and teacher evaluations, and can choose to focus school identification and accountability efforts on only a small subset of schools. As state policymakers make changes to school accountability and teacher evaluation systems, they should think carefully about the role that each system plays in encouraging continuous improvement in all schools and supporting educational equity for historically underserved students. In addition, as states make changes to the metrics used in school and teacher accountability systems, they should ensure that the measures for which teachers are accountable align with those for which schools are accountable to avoid creating conflicting or misaligned incentives for different actors in the system. If, as we argue above, student achievement and growth measures continue to play a role in teacher evaluation systems, it is crucial that states ensure that student assessments are of the highest quality, valid, and reliable.

#### **4 Invest in management and capacity to develop teachers**

Research shows that school leaders have tremendous influence on teacher quality through the hiring decisions they make, expectations they set, the feedback they provide to teachers, and the school culture they create. Regardless of how state and district policies related to teacher evaluation evolve, developing school leader capacity will be crucial to supporting quality teaching. The current school structures and budgeting radically underinvest in leadership capacity to supervise, develop, and support teachers. While there are a variety of models for expanding this capacity, all of them will require the education field to take the culture of support more seriously and to invest much more in cultivating leaders who manage and support teachers. This shift is necessary for reasons that extend beyond teacher evaluations. As federal policies move away from mandating specific interventions in low-performing schools, school and district leaders play an even more crucial role in acting on information from state accountability designations, and developing and implementing strategies to improve performance. Further, federal policies that provide states and districts with more flexibility to design their own policies and systems increase the importance of strong leadership at all levels of the system.

#### **5 Identify strategies to capture and learn from variation**

The next iteration in teacher evaluation policies will have much greater variation across states and among districts within states. This variation creates opportunities for learning—identifying both what works and what doesn't among different approaches—but only if there are structures and capacity in place to do so. Federal policymakers and philanthropic funders can support this learning by funding independent research on the effects of state and district evaluation practices, as well as by funding consortia of states and districts to share what they are doing and learn from one another. States that move away from requiring specific teacher evaluation policies should establish systems to track key features of district evaluation systems—even if these features are not required. States, districts, and intermediaries within states can work with researchers to allow them to have access to information and share that information back with districts in ways that help them learn from what they are doing and improve their practice over time.

## Conclusion

As states and districts embark on making new changes to teacher evaluation policies, they should consider lessons from the reform efforts of the recent past. In many cases, a complete overhaul may not be necessary. In fact, there are key attributes of recent reforms that can and should be kept in teacher evaluation systems in order to preserve the systems as vehicles for assessing effective teaching that ultimately leads to student academic achievement. However, new changes provide an opportunity to consider teacher evaluation as part of a larger effort to attract, retain, and leverage teacher talent in a way that may have been overlooked in recent reforms. Importantly, as more variation occurs across states and districts and less policy oversight is directed at teacher evaluation systems, local policy leaders and practitioners should consider ways to share lessons learned so that that systems can continue to improve over time.

## Endnotes

- 1 Kathryn M. Dougherty and Sandi Jacobs, "State of the States 2015: Evaluating, Teaching, Leading, and Learning," *National Council on Teacher Equality*, November 2015, <http://www.nctq.org/dmsView/StateofStates2015>.
- 2 Ibid., <http://www.nctq.org/dmsView/StateofStates2015>.
- 3 Sara Mead, Andrew Rotherham, and Rachael Brown, "The Hangover: Thinking about the Unintended Consequences of the Nation's Teacher Evaluation Binge," *Bellwether Education Partners*, September 2012, [http://bellwethereducation.org/sites/default/files/TeacherQuality\\_TheHangover\\_0.pdf](http://bellwethereducation.org/sites/default/files/TeacherQuality_TheHangover_0.pdf).
- 4 "Teachers matter: Understanding Teachers' Impact on Student Achievement," *Rand Education*, accessed October 8, 2016, <http://www.rand.org/education/projects/measuring-teacher-effectiveness/teachers-matter.html>.
- 5 "The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness," *The New Teacher Project (TNTP)*, last modified June 8, 2009, <http://tntp.org/publications/view/evaluation-and-development/the-widget-effect-failure-to-act-on-differences-in-teacher-effectiveness>.
- 6 Steven G. Rivkin, Eric A. Hanushek, and John F. Kain, "Teachers, Schools, and Academic Achievement," *Econometrica* 73, no. 2 (2005): <https://econ.ucsb.edu/~jon/Econ230C/HanushekRivkin.pdf>
- 7 "Ensuring Fair and Reliable Measures of Effective Teaching," *Education Week*, accessed October 8, 2016, <http://www.edweek.org/media/17teach-met1.pdf>.
- 8 "Partners in Reform," *US Department of Education*, last modified July 2, 2009, <http://www2.ed.gov/news/speeches/2009/07/07022009.html>.
- 9 "Race to the Top Program Executive Summary," *US Department of Education*, November 2009, <https://www2.ed.gov/programs/racetothetop/executive-summary.pdf>.
- 10 William G. Howell, "Results of President Obama's Race to the Top," *EducationNext* 15, no. 4 (2015): <http://educationnext.org/results-president-obama-race-to-the-top-reform/>.
- 11 Dougherty and Jacobs, "State of the States 2015," <http://www.nctq.org/dmsView/StateofStates2015>.
- 12 Bill Turque, "D.C. Teachers, Rhee Appear Close to Contract; Both Sides Might Yield Some Ground," *Washington Post*, last modified September 11, 2009, <http://www.washingtonpost.com/wp-dyn/content/article/2009/09/10/AR2009091004312.html>; Mike Debonis, "DCPS Teacher Contract to Be Unveiled; Big Raises Funded by \$65M in Private Money," *Washington City Paper*, last modified April 6, 2010, <http://www.washingtoncitypaper.com/news/city-desk/blog/13061329/dcps-teacher-contract-to-be-unveiled-big-raises-funded-by-65m-in-private-money>.
- 13 Dan Weisberg, "The Positive Impact of IMPACT," *TNTP*, last modified October 17, 2013, <http://tntp.org/blog/post/the-positive-impact-of-impact>; Thomas Dee and James Wyckoff, "Incentives, Selection, and Teacher Performance: Evidence from IMPACT," *Journal of Policy Analysis and Management* 34, no. 2 (2015): <http://cepa.stanford.edu/content/incentives-selection-and-teacher-performance-evidence-impact>.
- 14 Alyson Klein, "Obama Administration Sets Rules for NCLB Waivers," *Education Week*, last modified September 22, 2011, [http://blogs.edweek.org/edweek/campaign-k-12/2011/09/obama\\_administration\\_sets\\_rule.html](http://blogs.edweek.org/edweek/campaign-k-12/2011/09/obama_administration_sets_rule.html).
- 15 Dougherty and Jacobs, "State of the States 2015," <http://www.nctq.org/dmsView/StateofStates2015>.
- 16 Alyson Klein, "Arne Duncan Allows Waiver States Extra Time on Teacher Evaluation," *Education Week*, last modified June 18, 2013, [http://blogs.edweek.org/edweek/campaign-k-12/2013/06/arne\\_duncan\\_allows\\_waiver\\_stat.html](http://blogs.edweek.org/edweek/campaign-k-12/2013/06/arne_duncan_allows_waiver_stat.html); Lauren Camera, "Waiver States Can Seek a Year's Delay in Using Tests for Teacher Evaluations," *Education Week*, last modified August 21, 2014, [http://blogs.edweek.org/edweek/campaign-k-12/2014/08/ed\\_announcement.html](http://blogs.edweek.org/edweek/campaign-k-12/2014/08/ed_announcement.html).
- 17 Rick Karlin, "Going on the Offense, NYSUT's Magee Calls for Test Boycott," *Times Union*, last modified March 30, 2015, <http://blog.timesunion.com/capitol/archives/231494/going-on-the-offense-nysuts-magee-calls-for-test-boycott/>.
- 18 Jennifer Medina, "Agreement Will Alter Teacher Evaluations," *New York Times*, last modified May 10, 2010, [http://www.nytimes.com/2010/05/11/nyregion/11teacher.html?\\_r=0](http://www.nytimes.com/2010/05/11/nyregion/11teacher.html?_r=0); Monica Disare, "Union Pushing More 'Authentic' Assessments in Teacher Evaluations," *Chalkbeat*, last modified September 20, 2016, <http://www.chalkbeat.org/posts/ny/2016/09/20/union-pushing-more-authentic-assessments-in-teacher-evaluations/#.V-GmSPkrLIU>.



- 19 Monica Disare, "New York's Teacher Evaluations Aren't Going Anywhere. In Fact, They're Getting a Makeover that Nobody Planned," *Chalkbeat*, last modified June 10, 2016, <http://www.chalkbeat.org/posts/ny/2016/06/10/new-york-states-teacher-evaluations-arent-going-anywhere-in-fact-theyre-getting-a-makeover-that-nobody-planned/#.V17LKrsrLIV>; Medina, "Agreement," [http://www.nytimes.com/2010/05/11/nyregion/11teacher.html?\\_r=1](http://www.nytimes.com/2010/05/11/nyregion/11teacher.html?_r=1); "Teacher Evaluation Policy in New York: Where Is New York in Implementing Teacher Effectiveness Policies?" *National Council on Teacher Equality*, accessed October 8, 2016, [http://www.nctq.org/dmsView/Evaluation\\_Timeline\\_Brief\\_New\\_York](http://www.nctq.org/dmsView/Evaluation_Timeline_Brief_New_York).
- 20 "Teacher Evaluation Policy in New York," [http://www.nctq.org/dmsView/Evaluation\\_Timeline\\_Brief\\_New\\_York](http://www.nctq.org/dmsView/Evaluation_Timeline_Brief_New_York).
- 21 Monica Disare, "In Big Shift, Regents Vote to Exclude State Tests from Teacher Evals until 2019," *Chalkbeat*, last modified December 14, 2015, <http://www.chalkbeat.org/posts/ny/2015/12/14/breaking-in-big-shift-regents-vote-to-exclude-state-tests-from-teacher-evals-until-2019/#.VyDtJvkrLIV>.
- 22 "New York Will Have a New Teacher Evaluation System in 2019, Officials Say," *Chalkbeat*, last modified April 18, 2016, <http://www.chalkbeat.org/posts/ny/2016/04/18/new-york-will-have-an-new-teacher-evaluation-system-in-2019-officials-say/#.VyDtjfrLIX>.
- 23 Disare, "In Big Shift," <http://www.chalkbeat.org/posts/ny/2015/12/14/breaking-in-big-shift-regents-vote-to-exclude-state-tests-from-teacher-evals-until-2019/#.VyDtJvkrLIV>.
- 24 Stephen Sawchuk, "Unions, Teachers, and Opting Out: A Complex Calculus," *Education Week*, last modified October 7, 2014, [http://blogs.edweek.org/edweek/teacherbeat/2014/10/unions\\_teachers\\_opting\\_out.html](http://blogs.edweek.org/edweek/teacherbeat/2014/10/unions_teachers_opting_out.html).
- 25 Sarah Darville, "92 Percent of City Teachers Earn High Marks in Newest Round of Evaluations," *Chalkbeat*, last modified December 14, 2015, <http://www.chalkbeat.org/posts/ny/2015/12/14/92-percent-of-city-teachers-earn-high-marks-in-newest-round-of-evaluations/#.V17aLsrLIU>; John Mooney, "Overwhelming 97 Percent of State's Teachers Rated 'Effective' in New Evaluations," *NJ Spotlight*, last modified June 2, 2015, <http://www.njspotlight.com/stories/15/06/01/overwhelming-97-of-nj-teachers-earn-satisfactory-scores-on-new-evaluation-system/>; Stephen Sawchuk, "Teachers Highly Rated under Hawaii, Delaware Systems," *Education Week*, last modified August 22, 2014, [http://blogs.edweek.org/edweek/teacherbeat/2014/08/the\\_two\\_latest\\_states\\_to.html](http://blogs.edweek.org/edweek/teacherbeat/2014/08/the_two_latest_states_to.html).
- 26 Caitlin Beatson, "Staff Performance Evaluation Results 2014-15," *Indiana Department of Education*, accessed October 8, 2016, <http://www.in.gov/sboe/files/er-data-presentation-sboe-14-15.pdf>.
- 27 Lauren Sartain, Sara Ray Stoelinga, and Eric R. Brown, "Rethinking Teacher Evaluation in Chicago: Lessons Learned from Classroom Observations, Principal-Teacher Conferences, and District Implementation," *Consortium on Chicago School Research*, November 2011, <https://consortium.uchicago.edu/sites/default/files/publications/Teacher%20Eval%20Report%20FINAL.pdf>.
- 28 "Supporting Effective Teaching in Tennessee: Listening and Gathering Feedback on Tennessee's Teacher Evaluations," *State Collaborative on Reforming Education*, accessed October 8, 2016, <https://www.joomag.com/magazine/score-evaluation-full-report/0258001001389028027?short>.
- 29 Chad Aldeman and Carolyn Chuong, "Teacher Evaluations in an Era of Rapid Change: From 'Unsatisfactory' to 'Needs Improvement,'" *Bellwether Education Partners*, August 2014, [http://bellwethereducation.org/sites/default/files/Bellwether\\_TeacherEval\\_Final\\_Web.pdf](http://bellwethereducation.org/sites/default/files/Bellwether_TeacherEval_Final_Web.pdf).
- 30 Lauren Sartain and Matthew P. Steinberg, "Teachers' Labor Market Responses to Performance Evaluation Reform: Experimental Evidence from Chicago Public Schools," *The Journal of Human Resources* 51, no. 3 (2016): <http://jhr.uwpress.org/content/51/3/615.abstract>
- 31 Thomas Dee and James Wyckoff, "Incentives, Selection, and Teacher Performance: Evidence from IMPACT," *Journal of Policy Analysis and Management* 34, no. 2 (2015): <http://cepa.stanford.edu/content/incentives-selection-and-teacher-performance-evidence-impact>.
- 32 *Ibid.*, [http://bellwethereducation.org/sites/default/files/Bellwether\\_TeacherEval\\_Final\\_Web.pdf](http://bellwethereducation.org/sites/default/files/Bellwether_TeacherEval_Final_Web.pdf).
- 33 John Mooney, "Teacher-evaluation Ratings Released for Schools around the State," *NJ Spotlight*, last modified July 16, 2015, <http://www.njspotlight.com/stories/15/07/15/teacher-evaluation-ratings-released-for-schools-around-the-state/>.
- 34 Matthew A. Kraft and Allison F. Gilmour, "Revisiting the Widget Effect: Teacher Evaluation Reforms and the Distribution of Teacher Effectiveness," *scholar.harvard.edu*, February 2016, [http://scholar.harvard.edu/files/mkraft/files/kraft\\_gilmour\\_2016\\_revisiting\\_the\\_widget\\_effect\\_wp.pdf?m=1456772152](http://scholar.harvard.edu/files/mkraft/files/kraft_gilmour_2016_revisiting_the_widget_effect_wp.pdf?m=1456772152).

- 35 "Roadmap for a Teacher-student Data Link: Key Focus Areas to Ensure Quality Implementation," *Data Quality Campaign*, accessed October 8, 2016, <http://2pido73em67o3eytaq1cp8au.wengine.netdna-cdn.com/wp-content/uploads/2016/03/DQC-roadmap-TSDL.pdf>.
- 36 "Using Data to Improve Teacher Effectiveness: A Primer for State Policymakers," *Data Quality Campaign*, January 2012, <http://2pido73em67o3eytaq1cp8au.wengine.netdna-cdn.com/wp-content/uploads/2016/05/Using-Data-to-Improve-Teacher-Effectiveness.pdf>.
- 37 Matt Barnum, "The War over Evaluating Teachers—Where It Went Right and How It Went Wrong," *The 74*, last modified November 3, 2015, <https://www.the74million.org/article/the-war-over-evaluating-teachers-where-it-went-right-and-how-it-went-wrong>.
- 38 "The National Network of State Teachers of the Year Releases 2016 Results of First Federal Policy Survey," *NNSTOY.org*, April 5, 2016, <http://www.nnstoy.org/download/Press-Release-for-Federal-Policy-Survey-final-1.pdf>; Linda Lyons, "Teachers Favor Common Core Standards, Not the Testing," *Gallup*, last modified October 29, 2014, <http://www.gallup.com/poll/178997/teachers-favor-common-core-standards-not-testing.aspx>.
- 39 Sara Mead, Andrew Rotherham, and Rachael Brown, "The Hangover: Thinking about the Unintended Consequences of the Nation's Teacher Evaluation Binge," *Bellwether Education Partners*, September 2012, [http://bellwethereducation.org/sites/default/files/TeacherQuality\\_TheHangover\\_0.pdf](http://bellwethereducation.org/sites/default/files/TeacherQuality_TheHangover_0.pdf).
- 40 "Race to the Top District Competition Draft," *US Department of Education*, accessed October 8, 2016, <http://www.ed.gov/race-top/district-competition>.
- 41 "Leading Change: The Emerging Human Capital Landscape in Personalized Learning Initiatives," District Reform Support Network, *US Department of Education*, September 2016, <https://rttd.grads360.org/services/PDCService.svc/GetPDCDocumentFile?fileId=23534>.
- 42 Kaylan Connally and Melissa Tooley, "Beyond Ratings: Re-envisioning State Teacher Evaluation Systems as Tools for Professional Growth," *New America*, March 2016, [https://static.newamerica.org/attachments/12744-beyond-ratings-3/NA\\_BeyondRatingsPaper.deba47a82ff04af2833cebdbeed0c3ab.pdf](https://static.newamerica.org/attachments/12744-beyond-ratings-3/NA_BeyondRatingsPaper.deba47a82ff04af2833cebdbeed0c3ab.pdf).
- 43 "Teacher Evaluation and Support Systems: A Roadmap for Improvement," *Aspen Institute*, last modified March 2016, <http://www.aspendr.org/portal/browse/DocumentDetail?documentId=2956&download&admin=2956%7C2710081212>.
- 44 "Principles for Teacher Support and Evaluation Systems," *Council of Chief State School Officers*, accessed October 8, 2016, <http://www.ccsso.org/Documents/2016/Principles%20for%20Teacher%20Support%20and%20Evaluation%20Systems.pdf>.
- 45 "The Mirage: Confronting the Hard Truth about Our Quest for Teacher Development," *TNTP*, last modified August 4, 2015, <http://tntp.org/publications/view/the-mirage-confronting-the-truth-about-our-quest-for-teacher-development>.
- 46 Ty Tagami, "Teachers and Testing Foes Get a Win under the Gold Dome," *myajc.com*, last modified March 15, 2016, <http://www.myajc.com/news/news/local-education/teachers-and-testing-foes-get-a-win-under-the-gold/nqk8K/>; Weston Morrow, "Alaska Repeals Teacher Evaluation Plan," *newsminer.com*, last modified June 26, 2016, [http://www.newsminer.com/news/education/state-repeals-teacher-evaluation-plan/article\\_123b3d6a-3b74-11e6-9997-837770d45b7c.html](http://www.newsminer.com/news/education/state-repeals-teacher-evaluation-plan/article_123b3d6a-3b74-11e6-9997-837770d45b7c.html).
- 47 Stephen Sawchuk, "Teacher Evaluation Overhauled in Los Angeles Pact," *Education Week*, last modified June 15, 2016, [http://blogs.edweek.org/edweek/teacherbeat/2016/06/teacher\\_evaluation\\_overhauled\\_los\\_angeles\\_pact.html](http://blogs.edweek.org/edweek/teacherbeat/2016/06/teacher_evaluation_overhauled_los_angeles_pact.html).
- 48 Amber M. Northern, PhD, and Daniela Fairchild, "Measures of Effective Teaching: Final Reports," *Thomas B. Fordham Institute*, last modified January 9, 2013, <https://edexcellence.net/commentary/education-gadfly-weekly/2013/january-10/measures-of-effective-teaching-final-reports.html>; "(MET) Measures of Effective Teaching Project," *Bill & Melinda Gates Foundation*, accessed October 9, 2016, <http://k12education.gatesfoundation.org/teacher-supports/teacher-development/measuring-effective-teaching/>.
- 49 "Toward Trustworthy and Transformative Classroom Observations: Progress, Challenges, and Lessons in SREB States," *Southern Regional Education Board*, February 2015, [http://www.sreb.org/sites/main/files/file-attachments/sreb\\_coreportonline.pdf](http://www.sreb.org/sites/main/files/file-attachments/sreb_coreportonline.pdf).

- 50 Jay Chambers, Iliana Brodziak de los Reyes, and Caitlin O'Neil, "How Much Are Districts Spending to Implement Teacher Evaluation Systems?: Case Studies of Hillsborough County Public Schools, Memphis City Schools, and Pittsburgh Public Schools," *American Institutes for Research*, May 2013, [http://www.air.org/sites/default/files/downloads/report/What\\_Districts\\_Are\\_Spending\\_to\\_Implement\\_Teacher\\_Evaluation\\_Systems\\_Final\\_0.pdf](http://www.air.org/sites/default/files/downloads/report/What_Districts_Are_Spending_to_Implement_Teacher_Evaluation_Systems_Final_0.pdf).
- 51 Dougherty and Jacobs, "State of the States 2015," <http://www.nctq.org/dmsView/StateofStates2015>.
- 52 Douglas N. Harris, "Creating a Valid Process for Using Teacher Value-Added Measures," *Albert Shanker Institute*, last modified November 28, 2012, <http://www.shankerinstitute.org/blog/creating-valid-process-using-teacher-value-added-measures>.
- 53 Thomas Toch, "Grading the Graders: A Report on Teacher Evaluation Reform in Public Education," *educationpost.org*, May 2016, <http://educationpost.org/wp-content/uploads/2016/05/Thomas-Toch-Grading-the-Graders.pdf>.
- 54 Deborah Loewenberg Ball, David K. Cohen, and Brian Rowan, "Study of Instructional Improvement (SII) (ICPSR 26282)," *ICPSR*, May 20, 2010, <http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/26282#>.
- 55 "The Three Essentials: Improving Schools Requires District Vision, District and State Support, and Principal Leadership," *SREB*, August 2010, <http://www.wallacefoundation.org/knowledge-center/Documents/Three-Essentials-to-Improving-Schools.pdf>; Mead, Rotherham, and Brown, "The Hangover," [http://bellwethereducation.org/sites/default/files/TeacherQuality\\_TheHangover\\_0.pdf](http://bellwethereducation.org/sites/default/files/TeacherQuality_TheHangover_0.pdf).
- 56 Frederick M. Hess and Linda Darling-Hammond, "How to Rescue Education Reform," *New York Times*, last modified December 5, 2011, [http://www.nytimes.com/2011/12/06/opinion/how-to-rescue-education-reform.html?\\_r=4&src=tp&mtrref=wordpress.redirectingat.com&gwh=6799430CF7CC1E67DAC4DCC81BF0F674&gwt=pay&assetType=opinion](http://www.nytimes.com/2011/12/06/opinion/how-to-rescue-education-reform.html?_r=4&src=tp&mtrref=wordpress.redirectingat.com&gwh=6799430CF7CC1E67DAC4DCC81BF0F674&gwt=pay&assetType=opinion).
- 57 Dougherty and Jacobs, "State of the States 2015," <http://www.nctq.org/dmsView/StateofStates2015>.
- 58 Shaina Cavazos, "Report: Indiana Teachers Are Being Evaluated with 'Erratic' Mix of 200 Systems," *Chalkbeat*, last modified July 5, 2016, <http://www.chalkbeat.org/posts/in/2016/07/05/report-finds-major-inconsistencies-in-teacher-evaluation-plans-across-indiana/#.V5FfCvkrLIX>.
- 59 Dougherty and Jacobs, "State of the States 2015," <http://www.nctq.org/dmsView/StateofStates2015>.
- 60 Chad Aldeman and Ashley LiBetti Mitchel, "No Guarantees: Is It Possible to Ensure Teachers Are Ready on Day One?" *Bellwether Education Partners*, February 2016, pages 12–15, [http://bellwethereducation.org/sites/default/files/Bellwether\\_NoGuarantees\\_Final.pdf](http://bellwethereducation.org/sites/default/files/Bellwether_NoGuarantees_Final.pdf).
- 61 "The MetLife Survey of the American Teacher: Challenges for School Leadership," *metlife.com*, February 2013, <https://www.metlife.com/assets/cao/foundation/MetLife-Teacher-Survey-2012.pdf>.
- 62 Julie O'Connor, "Flunking Out Bad Teachers," *nj.com*, last modified December 19, 2010, [http://www.nj.com/njvoices/index.ssf/2010/12/flunking\\_out\\_bad\\_teachers.html](http://www.nj.com/njvoices/index.ssf/2010/12/flunking_out_bad_teachers.html).
- 63 Ben Backes, James Cowan, Dan Goldhaber, Cory Koedel, Luke Miller, and Zeyu Xu, "The Common Core Conundrum: To What Extent Should We Worry that Changes to Assessments and Standards Will Affect Test-based Measures of Teacher Performance?" *National Center for Analysis of Longitudinal Data in Education Research*, February 2016, <http://www.caldercenter.org/sites/default/files/Working%20Paper%20152.pdf>.
- 64 Joy Resmovits, "California's Parents Don't Know They Have a Voice in School Spending Decisions," *Los Angeles Times*, last modified September 8, 2016, <http://www.latimes.com/local/education/la-me-usc-education-poll-20160907-snap-story.html>.

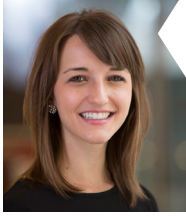


## Acknowledgments

Bellwether Education Partners would like to express gratitude for the generous support from the Bill and Melinda Gates Foundation for this project.

We would also like to thank the following individuals who provided advice and thoughtful feedback on this project: Bryan Hassel, Public Impact; Sandi Jacobs, EducationCounsel; Brad Jupp, U.S. Department of Education; Jason Kamras, DC Public Schools; Dan Weisberg, TNTP; and Analiza Wolf, Achievement First. The views and opinions expressed in this paper, however, are the responsibility of the authors alone and should not be interpreted to reflect the views of others.

## About the Authors



### Kaitlin Pennington

Kaitlin Pennington is a senior analyst at Bellwether Education Partners. She can be reached at [kaitlin.pennington@bellwethereducation.org](mailto:kaitlin.pennington@bellwethereducation.org).



### Sara Mead

Sara Mead is a partner at Bellwether Education Partners. She can be reached at [sara.mead@bellwethereducation.org](mailto:sara.mead@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 so, we provide a unique combination of exceptional thinking, talent, and hands-on strategic support.

© 2016 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](http://www.creativecommons.org). If you have any questions about citing or reusing Bellwether Education Partners content, please contact us.