

STATE OF THE STATES 2022

# Teacher and Principal Evaluation Policies

NOVEMBER 2022

---

## INTRODUCTION

### **Supporting teachers and principals by recognizing strong performance and helping them grow is more urgent than ever.**

Results of the 2022 National Assessment of Educational Progress reveal alarming results: Since 2019, scores declined substantially for all students, while disparities widened for students already most affected by opportunity gaps.<sup>1</sup> As districts and states help students recover in the wake of a global pandemic, supporting teachers and principals by recognizing strong performance, and helping them grow and improve where necessary, is more urgent than ever.

Strong teacher and principal evaluation systems have the potential to help teachers and principals improve their practice, to exit teachers who are perennially ineffective, to retain teachers who are effective and learn from them, and to increase the overall quality of a district's teacher workforce.<sup>2</sup>

As states respond to widespread concerns (both real and perceived) about teacher shortages and resignations, evaluation systems also have a role to play: Schools need access to fair, valid evaluation systems to help identify and retain highly effective teachers and principals, as well as to support those who are struggling.

As is true of all policies, implementation matters. A recent working paper from the Annenberg Institute generated disappointment and reflection after researchers found that, on the whole, states' changes to evaluation systems have not yielded the student outcomes they had hoped for.<sup>3</sup> These findings stood in stark contrast to the well-documented success that systems like those in **Dallas, Denver, the District of Columbia, Chicago**, and the state of **Tennessee** have had building strong evaluation systems that directly contributed to improved student learning and higher teacher quality.<sup>4</sup>

But a closer look at Annenberg’s research gives reason for optimism. Researchers found bright spots across the country (obscured by the larger trend) where exemplary evaluation systems made a significant impact on student achievement. These exemplary evaluation systems had a number of evidence-based practices in common, including the use of multiple measures to evaluate teachers’ effectiveness (particularly student growth and student surveys), meaningful differentiation between teachers, regular and sustained opportunities for observation and feedback, guaranteed written feedback, and alignment with professional learning.<sup>5</sup> These findings add to the evidence that evaluation, *done well*, can make a difference for students and educators.<sup>6</sup>

Given the importance of state policies to set the conditions for successful evaluation systems, NCTQ has regularly collected data, starting in 2011, to chart states’ progress in adopting evidence-based evaluation practices. In this report, we analyze statewide policies for teacher and principal evaluations in the 50 states and the District of Columbia, using data collected in fall of 2021 and verified by states in early 2022, in order to answer the following questions:

- **What role does the state play in teacher and principal evaluation design?**
- **What components are included in a teacher or principal’s evaluation?**
- **When, where, how, and by whom are evaluations conducted?**
- **Are evaluations used for support and improvement?**

---

## FINDINGS

### **States have largely retreated or stalled in adopting evidence-based teacher and principal evaluation policies.**

Since our last analysis in 2019, states have largely retreated or stalled in adopting evidence-based teacher and principal evaluation policies that support student learning. While state evaluation systems did experience disruptions throughout the pandemic, this pattern follows a trend that began as early as 2016. Since then, states have continued to move away from including measures of student academic growth as part of evaluations, and several have dropped the use of student surveys as well.

While several states have made progress in adopting effective practices like annual observations, most still have significant room to improve when it comes to requiring practices that support teachers’ growth and development, such as requiring annual feedback for all teachers. Even as some states have lowered standards for entry into the profession, too many still do not require the basic support structures necessary to help new teachers improve, such as additional observation and feedback that begin early in the year.

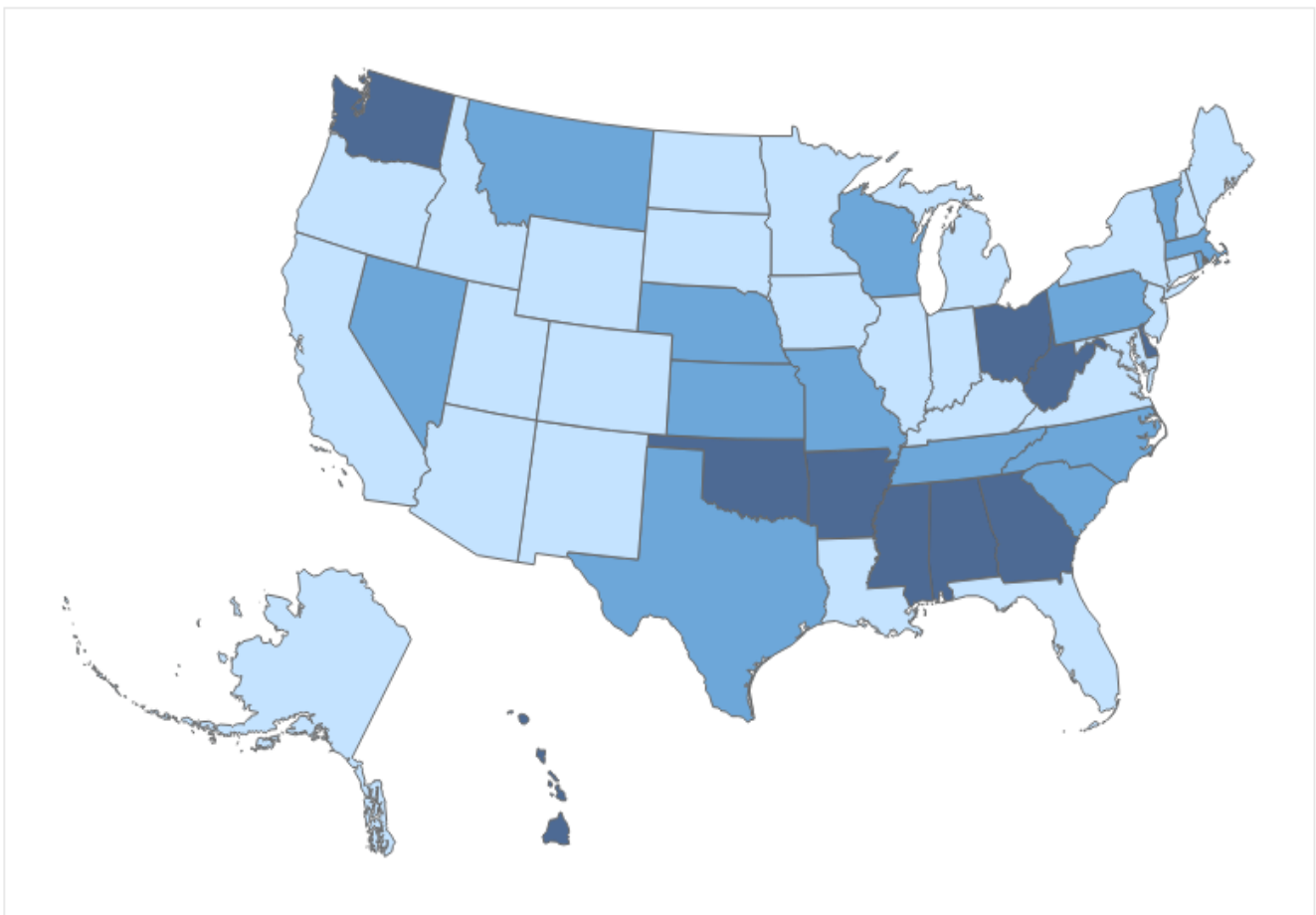
States have also lost ground or failed to make progress in measuring meaningful outcomes for principals, continuing a trend away from factoring student academic growth and survey results into principal evaluations.

## Teacher evaluation

Only 10 states require a teacher evaluation system that is the same statewide. Fourteen states allow districts to opt in or out of their statewide teacher evaluation system, and in 27 states the district designs their own evaluation systems based on criteria laid out by the state. While there are a number of reasons that a state might allow an evaluation system that is not uniform across all districts, without at least some common elements, consistency and comparability are limited.

*Figure 1.*

### What role does the state play in teacher evaluation design?



- Districts design their own evaluation systems based on criteria from the state
- Districts may either use the state's evaluation system or develop their own
- Districts must use the state's evaluation system

## What components are included in a teacher's evaluation?

Research has shown that it takes multiple sources of information to provide a fair and accurate understanding of a teacher's performance, and that evaluations based on multiple measures are more likely to be reliable and predictive.<sup>7</sup> In 2018, NCTQ studied four large school districts and two states where evaluation had led to meaningful improvements in teacher quality. All six evaluation systems measured different facets of teachers' effectiveness from varied perspectives (e.g., student growth as measured by state assessments, observations by school administrators, and student surveys).<sup>8</sup>

Common elements of an evaluation using multiple measures might include formal observations; measures of students' academic growth, including on state assessments; and student survey data. Of these elements, we find that states are generally far more reliant on observations, and have significantly decreased the use of any other sources of evidence, particularly those tied to quantitative measures of student learning.

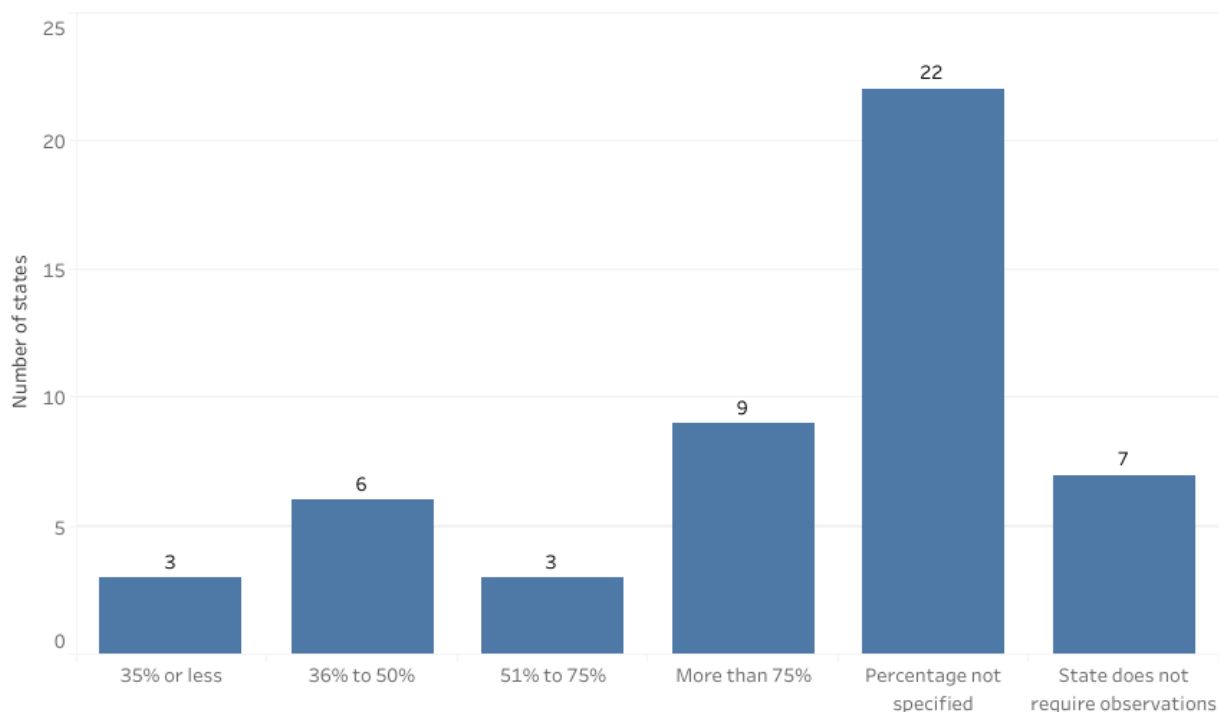
### Observations

Observations (particularly when they are based on a clearly defined rubric) provide a rich source of information about multiple aspects of a teacher's skills and impact on students, and are a useful starting point for providing actionable, specific, and relevant feedback.<sup>9</sup> Recent survey data suggests that most teachers find observations helpful in improving their instructional practice.<sup>10</sup> Seven states (the **District of Columbia, Missouri, Montana, Nebraska, New Hampshire, North Dakota, and Vermont**) currently do not require observations as part of a teacher's evaluation. Twenty-two states require observations, but do not specify the percentage of a teacher's evaluation made up by observations. Of the 22 states that do specify a percentage, just under one half specify that observations make up the bulk (75% or more) of an educator's evaluation score.

While observations are a critical factor in any feedback cycle, there are well-documented limitations to their usefulness and reliability in understanding teacher performance, including patterns of bias<sup>11</sup>—all the more reason for states to include multiple measures, carefully weighing a range of evidence to provide feedback and evaluate performance.

Figure 2.

## What percentage do observations account for in a teacher's overall evaluation score?



*Note: Although Delaware does require observations, the state is currently transitioning key evaluation system policies and has not announced the percentage that they will account for in a teacher's evaluation score.*

### Measures of student growth

As part of an effective evaluation system, observations should be considered together with measures of student academic growth, which might include measures like student learning objectives (SLOs), district assessments, statewide assessments, or other shared measures.

Helping students to grow academically is core to a teacher's role<sup>12</sup> and should be a component of any evaluation. Evaluations are also more likely to be valid measures of a teacher's performance when quantitative measures of student learning are combined with qualitative measures like observations.<sup>13</sup>

States have continued to lose ground on including measures of student growth in evaluations. Between 2019 and 2022, four states—**Indiana**, **Mississippi**,<sup>14</sup> **North Dakota**, and **Oregon**—dropped requirements for including objective measures of student growth in teachers' evaluations. Of the 30 states that use measures of student growth, 19 specify the percentage of a teacher's evaluation that growth should comprise, ranging from 10% to 50%.

Figure 3.

**Are measures of student growth required as part of a teacher's evaluation score?**

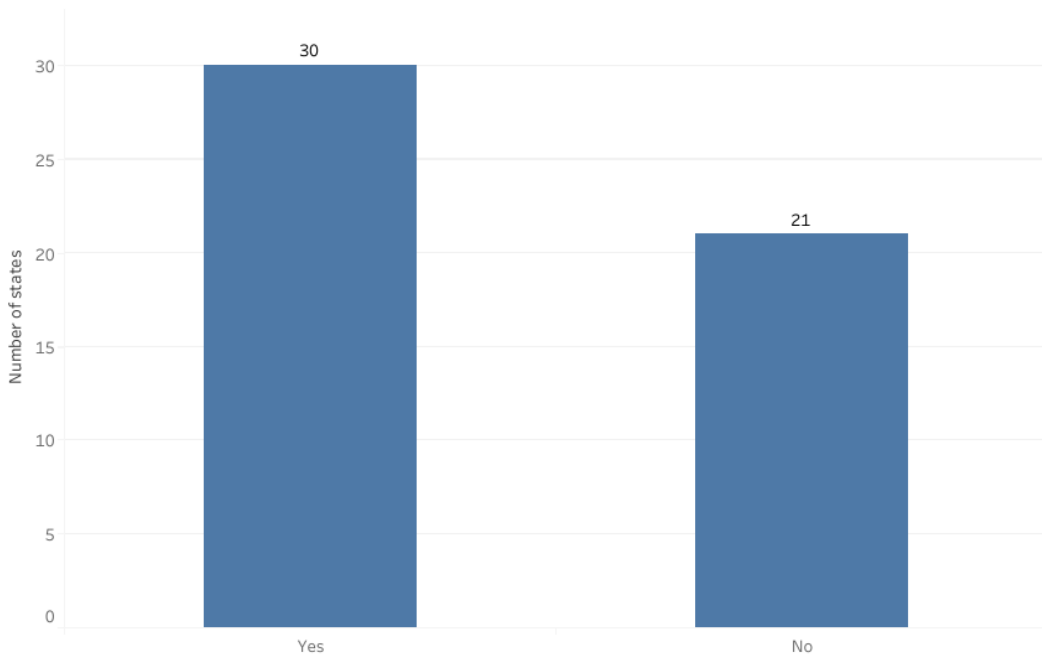
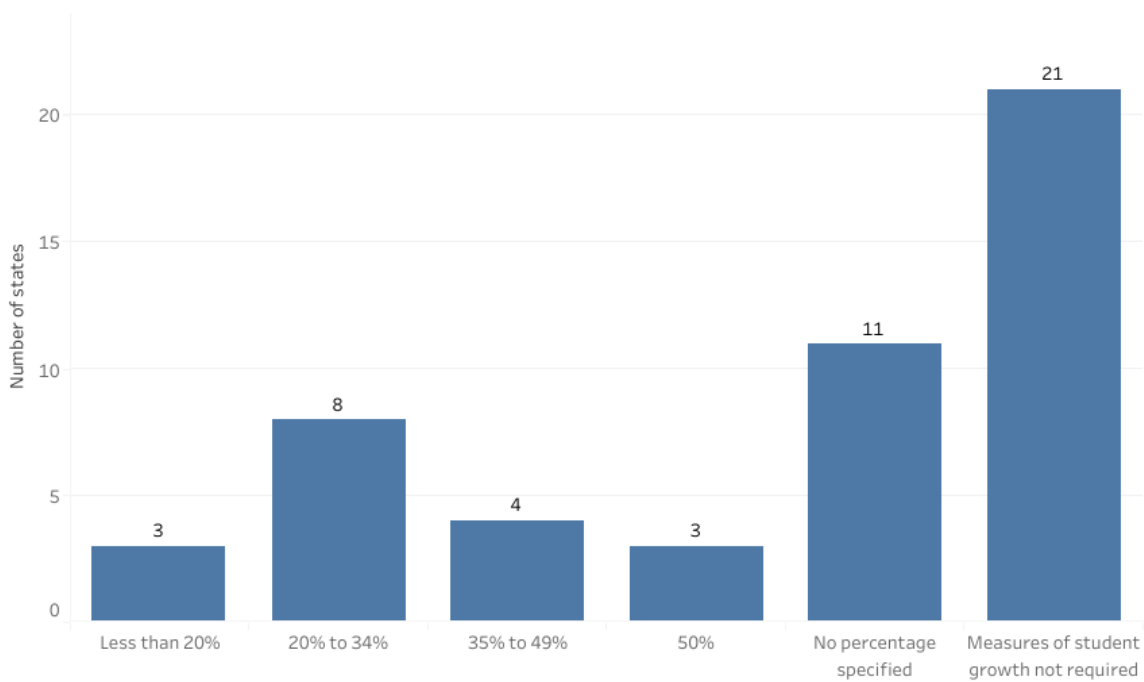


Figure 4.

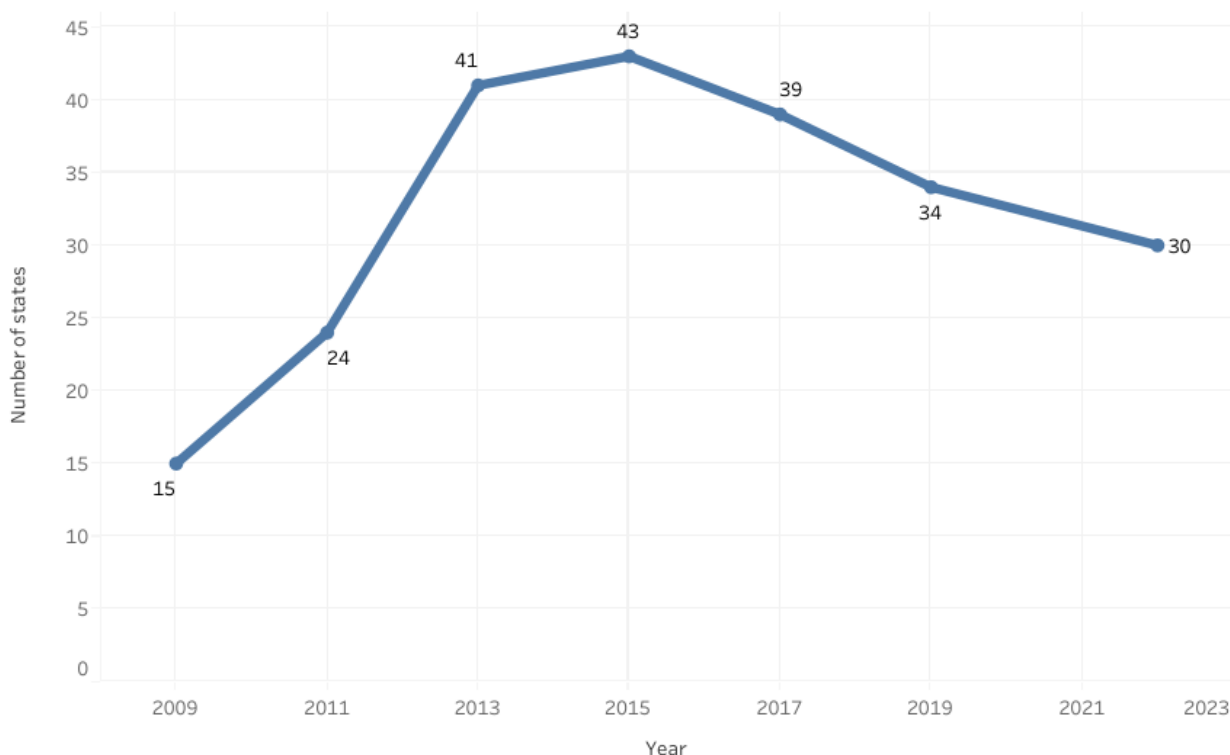
**How much of a teacher's evaluation score comes from measures of student growth?**



*Note: Delaware is not currently included in this figure, as the state is currently transitioning key evaluation system policies.*

Figure 5.

## How many states' teacher evaluation systems require measures of student growth?



*Note: These figures include states with explicit requirements in policy for student growth, regardless of the status of implementation.*

### State assessments to measure student learning

Between 2019 and 2022, the number of states where statewide assessments are required or explicitly allowed in evaluations decreased from 27 to 23. **Alabama, New York, and Virginia** added state tests as required or explicitly allowed measures, while **Arizona, Delaware, Indiana, North Dakota, Oregon, South Carolina, and West Virginia** dropped state assessments as a required or explicit measure of student growth. While pandemic disruptions may have prompted or accelerated a move away from use of state assessment data in at least a few states, many had already announced these changes by the early winter of 2020.<sup>15</sup>

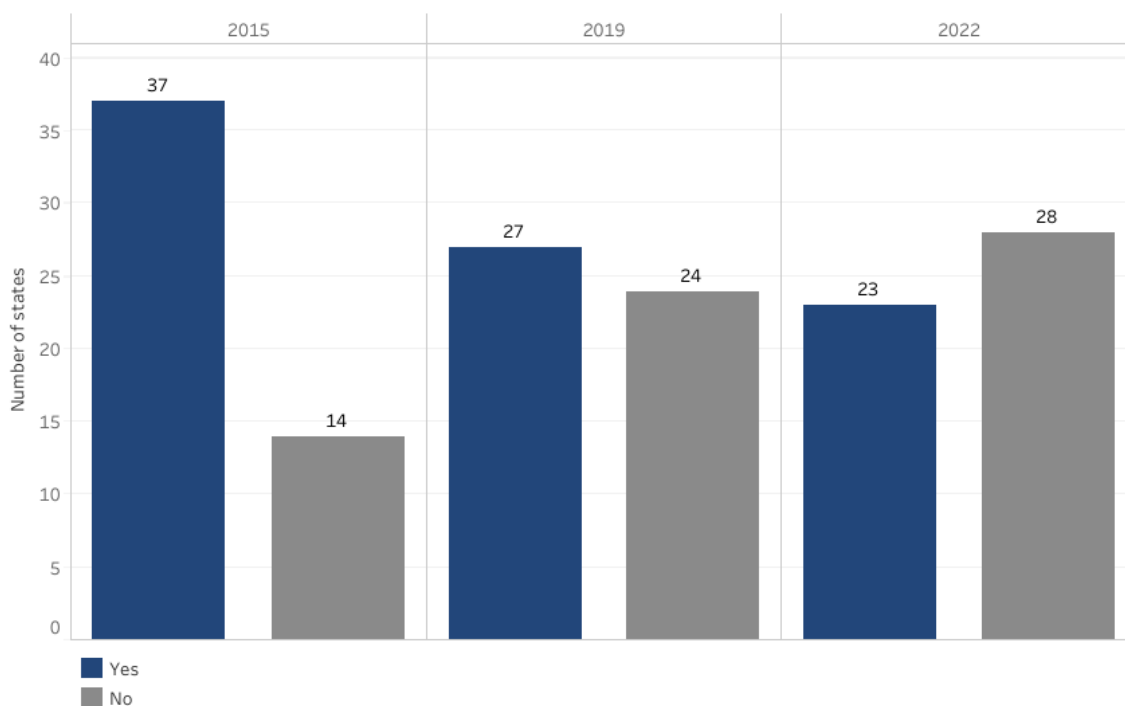
### Pandemic Disruptions to State Assessments

All 50 states and the District of Columbia were granted waivers from the U.S. Department of Education to forgo statewide assessments in the spring of 2020 in response to the onset of the pandemic. This had a significant impact on state teacher evaluation systems that include statewide assessments of student learning. As a result, many districts and states paused evaluations altogether or excluded state tests from teachers' evaluations.<sup>16</sup> Those changes continued into the 2020-2021 school year, with disruptions still affecting statewide assessments in spring of 2021.<sup>17</sup>

Beyond these fluctuations, a longer-term trend is also clear: States continue to back away from using measures of student growth and using valid and reliable assessments of student learning in evaluation. Between 2015 and 2022, 14 states dropped requirements or allowances for the use of statewide assessments in evaluation. Without shared quantitative measures, it is more challenging for states to accurately assess the equitable distribution of effective teachers across districts and student populations. Moreover, a lack of shared measures also means that educators statewide are not held to the same expectations for student learning.

Figure 6.

### Do states explicitly allow or require data from state standardized tests in teacher evaluations?



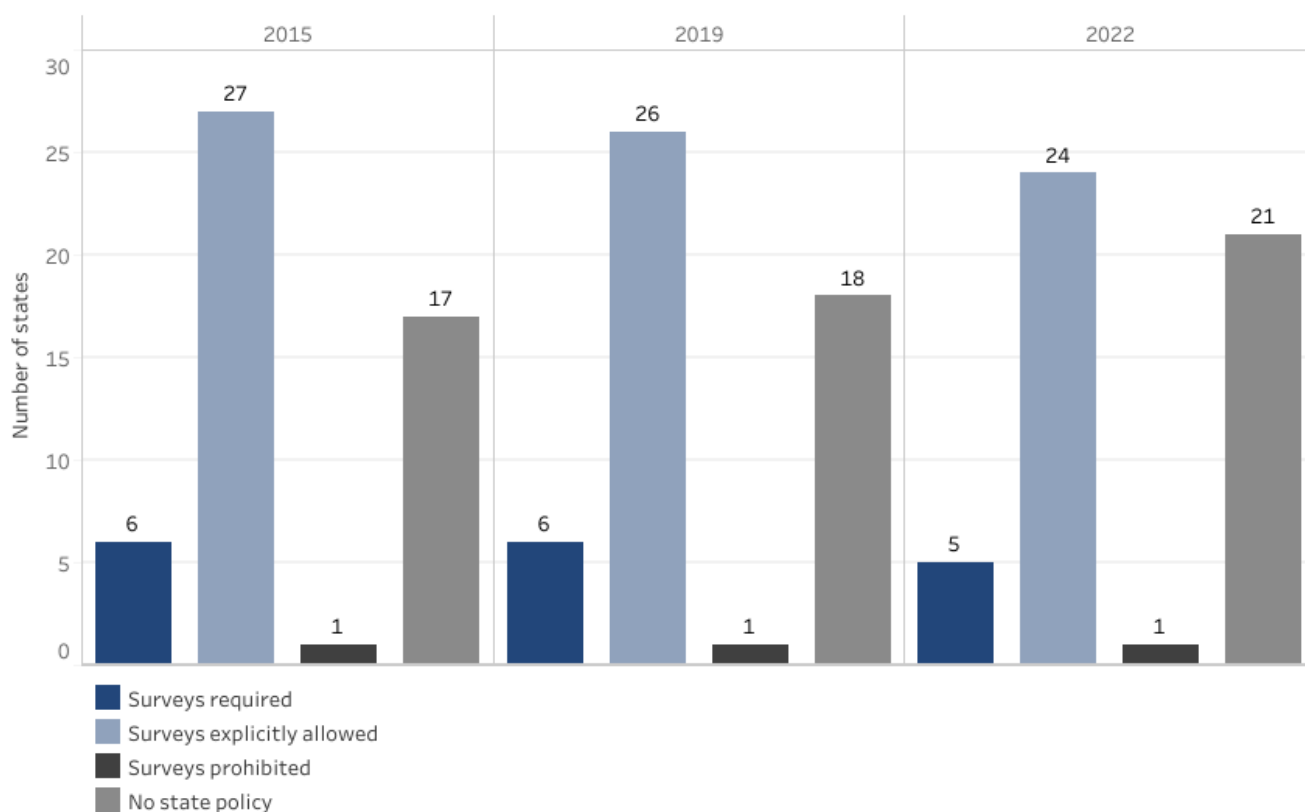
### Student surveys

Another common component of effective teacher evaluation systems are student surveys, which give students a chance to give feedback on their teachers' classroom climate and instructional skills. Research shows that student survey ratings are positively correlated with learning gains, and that they are an accurate and consistent measure of teacher quality.<sup>18</sup> Despite this, states have lost some ground in the use of student surveys: Three fewer states require or explicitly allow the use of student surveys than did in 2019.



Figure 7.

## What is the role of student surveys in teacher evaluation?



### Evaluation rating categories

An evaluation rating system with three categories or more is important to meaningfully distinguish performance, and evidence shows that binary systems favor nearly all teachers being rated satisfactory.<sup>19</sup> While 37 states use a system that includes three or more rating categories in order to differentiate performance (with the majority, 31, using a four category system), 14 use either a binary system or do not specify rating categories. Three states still use a five category system: **North Carolina, Oklahoma, and Tennessee.**

### When, where, how, and by whom are evaluations conducted?

Regular feedback is a critical element in helping teachers grow their skills and promote positive student outcomes. Annual evaluations for all teachers is a key feature of successful evaluation systems that improve teacher effectiveness and increase student learning.<sup>20</sup>

## Evaluation frequency

Twenty-two states require districts to evaluate all teachers every year, and the majority of states (37) only require that probationary<sup>21</sup> teachers receive an evaluation once a year. Only eight states (**Alabama, Hawaii, Illinois, Maryland, Massachusetts, Ohio, Rhode Island** and **Washington**) require that teachers with low performance ratings receive additional evaluations. Two additional states, **Texas** and **Wyoming**, explicitly make additional evaluations for low-performing teachers an option in their state policies.

Figure 8.

### How many states require all teachers to be evaluated annually?

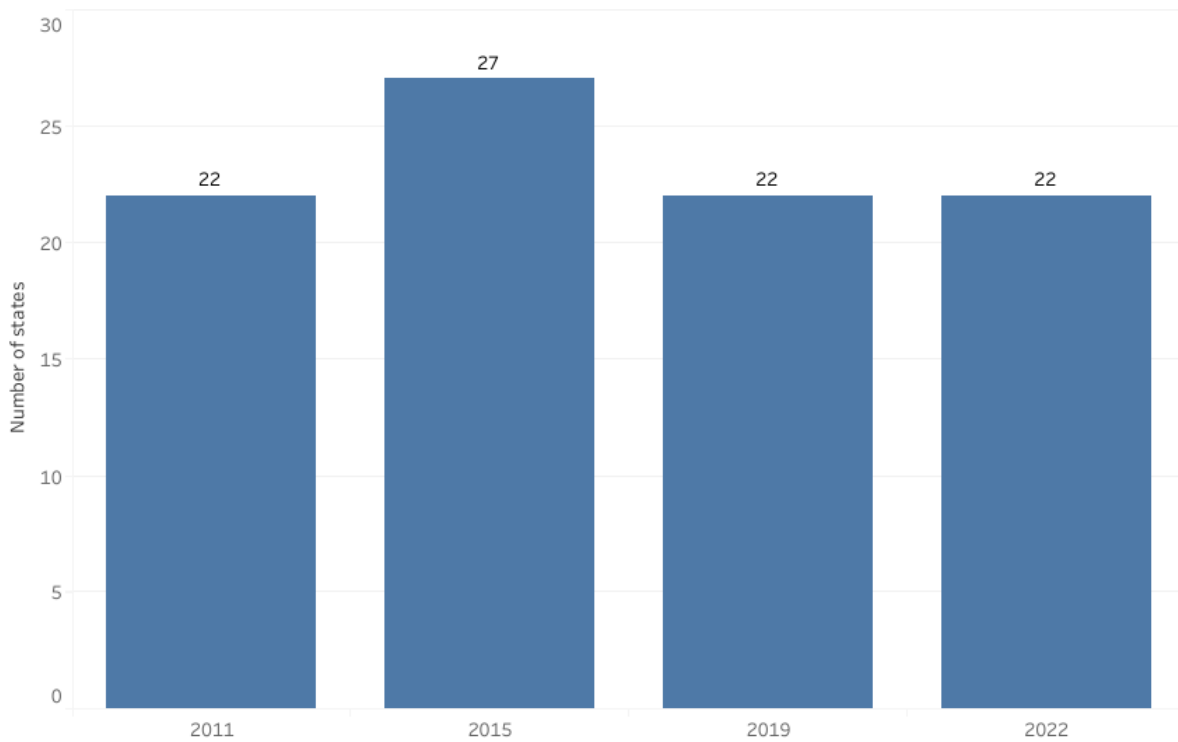


Figure 9.

### Are all non-probationary teachers evaluated annually?

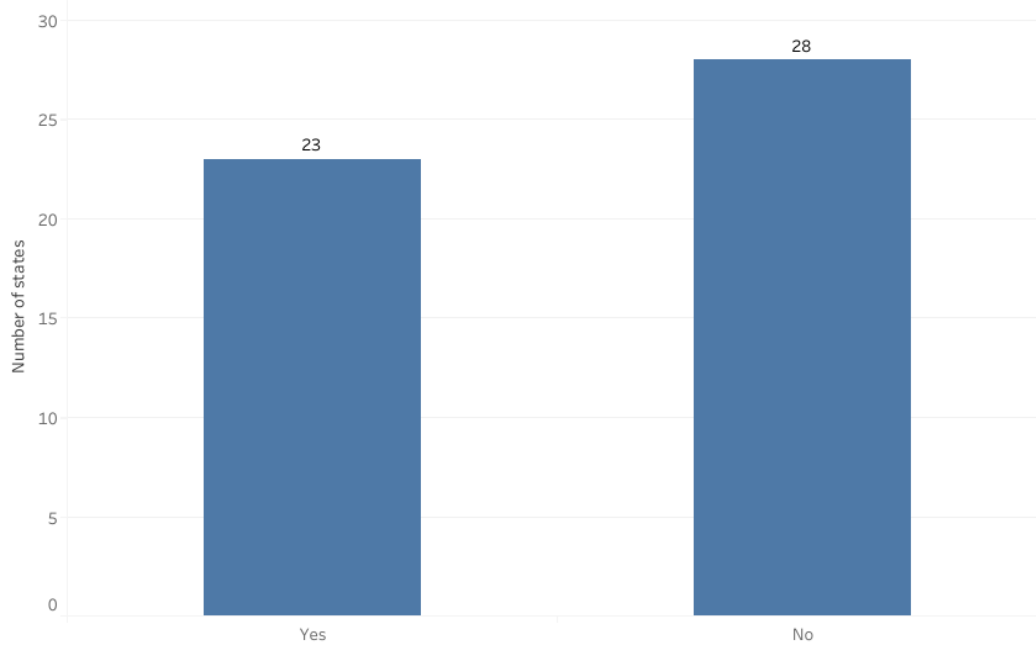
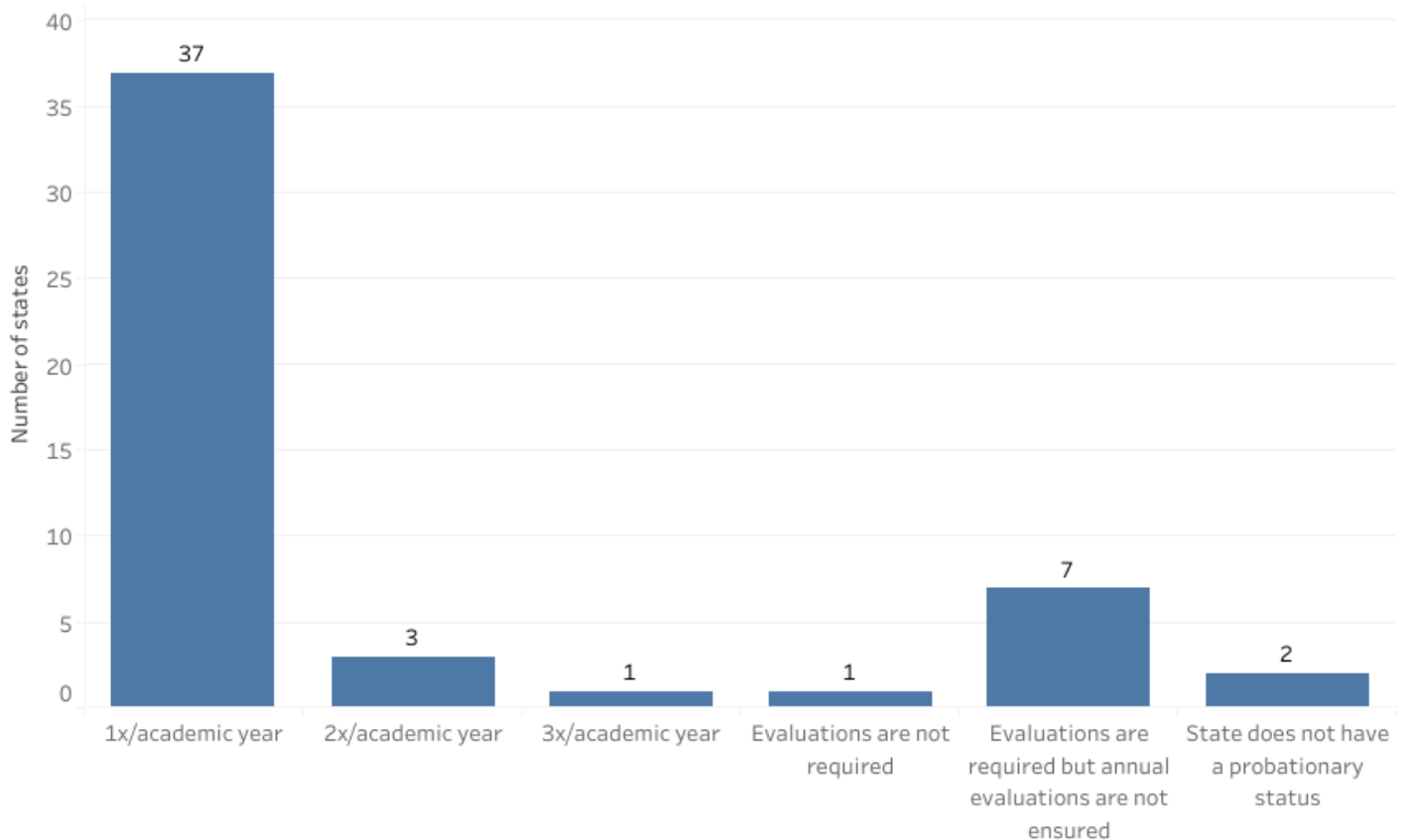


Figure 10.

### How frequently are probationary teachers required to receive an evaluation?



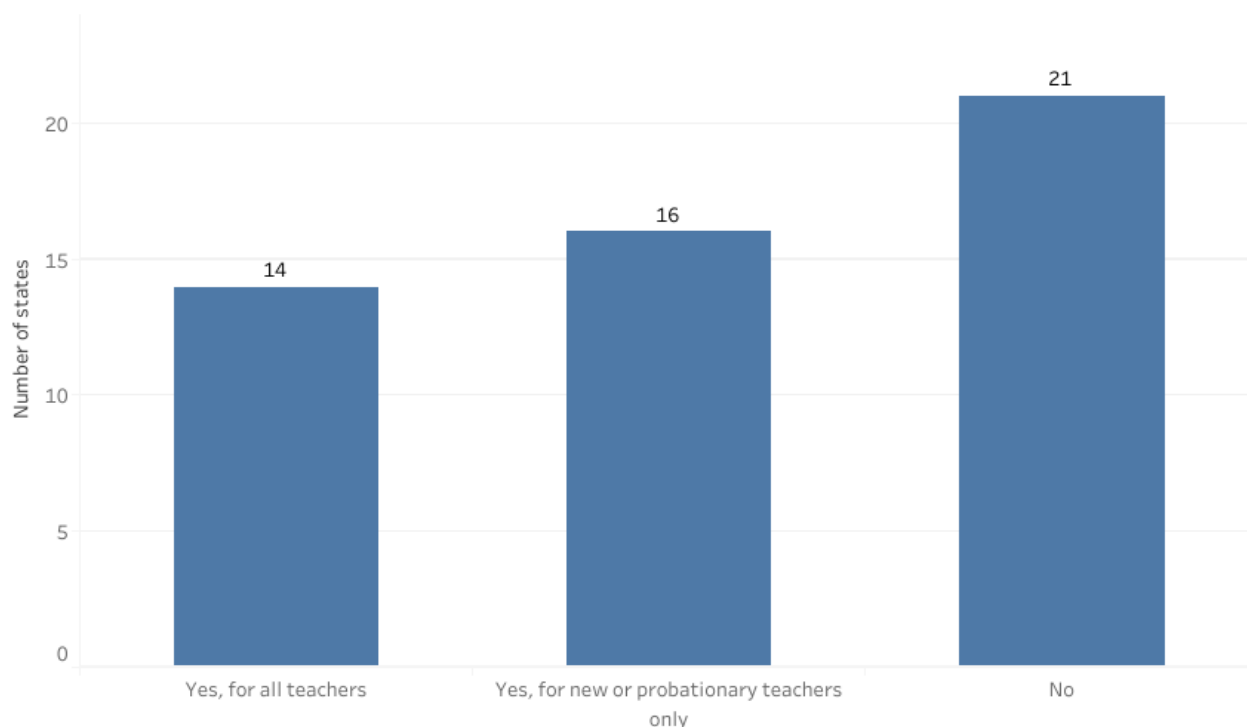
## Observation frequency

It is widely accepted that opportunities for expert observation, feedback, and practice are important for all teachers, but particularly new teachers. Research suggests that more than one observation is necessary to accurately assess teacher performance as part of an evaluation.<sup>22</sup> At least one recent study found that teachers who are observed four or more times per year report a more positive view of their evaluation system, compared to those who are observed less often.<sup>23,24</sup>

Observations are also more likely to yield reliable information about a teacher's performance when teachers receive more of them, particularly when they are conducted by more than one observer.<sup>25</sup> Yet only 14 states require all teachers be observed multiple times each year; an additional 16 states require multiple observations for early career/probationary teachers only. When it comes to ensuring that new teachers are set up for success from the beginning, only 17 states require that new teachers receive observation and feedback early in the school year. Five states (**Iowa, Maryland, New Jersey, New York, and South Carolina**) require use of multiple observers, while an additional 15 states allow but do not require their use.

Figure 11.

### Do states require teachers to be observed multiple times per year?



## Observer qualifications

Only 19 states articulate specific certification requirements for observers, while 38 require some training for evaluators. Statewide policies like these are a lever to set a standard that all teachers are observed by a knowledgeable observer and are well calibrated to the observation protocol or rubric—two elements particularly critical to effective evaluations.<sup>26</sup>

## Video and recorded observations

Prior to the onset of the pandemic and subsequent shift to remote instruction, four states (**Massachusetts, Michigan, New Jersey, and New York**) allowed some form of virtual observation for evaluations.<sup>27</sup> Since April 2020, that number has more than tripled, as 10 additional states made changes to allow virtual observations in response to the use of remote instruction<sup>28</sup> during the pandemic. Of the states with policies to accommodate virtual observation (our analysis included recorded observations, or live observations of teachers in a virtual/hybrid learning environment), some (**Oklahoma**, for example) specifically articulate that virtual observation is only to be used in a virtual learning environment. Others, like **Massachusetts, Michigan, New Jersey, and New York** (all of which enacted this flexibility pre-pandemic), allow for self-recording. **New Jersey**, building on the success of a pilot program for highly effective teachers, provides flexibility for tenured teachers who have received a “highly effective” rating on their most recent summative evaluation to replace one traditional, announced observation with a number of alternative activities, including videotaping a lesson and providing a reflection on that lesson.<sup>29</sup>

The flexibility of using technology to gather evidence, intended to support continued feedback and growth during an exceptional circumstance, could have long-term benefits, should districts and states choose to expand the use of self-recording to in-person teaching. In a recent study of 400 teachers, researchers found that when teachers videotaped themselves delivering a lesson and then watched the footage and discussed with an observer later, they reported more positive feelings about the observation and feedback process, and had a higher retention rate than peers not selected for videotaped observations.<sup>30</sup> The flexibility to permit video observations may allow districts to adopt a practice that has better buy-in from educators, and has the potential added benefit of helping manage observers’ time, which is often a challenge with in-person observations.

## Are evaluations used for support and improvement?

Evaluations should be connected to timely, specific, and actionable feedback, and give teachers opportunities for growth and chances to demonstrate improvement. As schools take on the urgent work of helping students recover from the pandemic, this is especially important. Yet far too many states still do not require that teachers receive any feedback after an observation, or that evaluations will be used to provide targeted growth and

support. States also miss critical opportunities to use evaluation data at the state level to drive system-level improvement in how teachers deemed effective are distributed across the state.

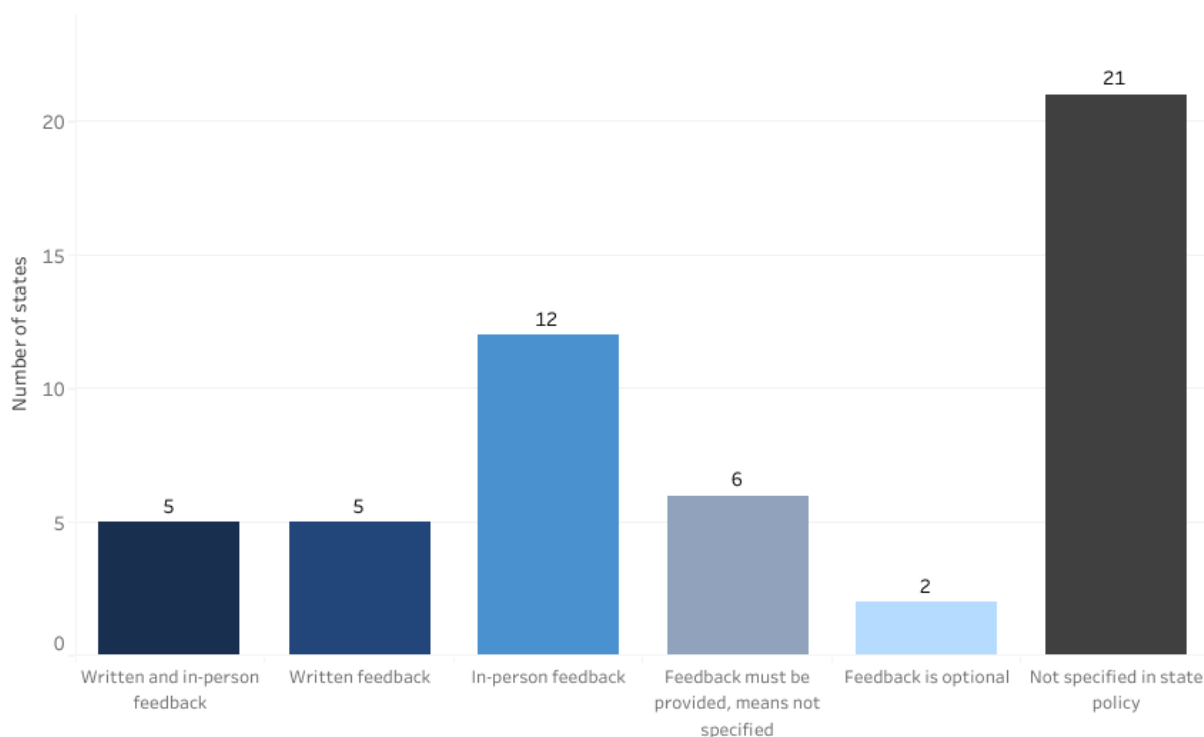
### Observation and evaluation feedback

Too many states still do not explicitly require feedback to be provided to teachers after an observation: 19 states do not have a statewide policy that requires feedback to teachers in any form (whether written, in-person or otherwise), while two states specifically designate observation feedback as optional.

Further, some states still do not explicitly require feedback to be provided to a teacher as part of their evaluation overall: eight states (including **Alabama, Alaska, the District of Columbia, Iowa, Minnesota, Montana, New Hampshire, and Vermont**) do not require teachers to receive feedback either written or in-person, after an evaluation.

Figure 12.

### What feedback do states require after observations?



### Connection to professional development opportunities and improvement plans

Research suggests that observations are more likely to positively impact teachers' effectiveness when they are connected directly to professional development opportunities.<sup>31</sup> Yet 20 states do not explicitly connect evaluation results to professional development, missing a critical opportunity to require aligned support to help teachers to



## Principal evaluation

The research is clear: Strong school leaders create strong schools.<sup>34</sup> As states continue to look for ways to combat teacher turnover and help students recover academically, principals are key leaders of this work in their schools, and their evaluations should reflect that.

Principals have an important role to play in school quality, particularly in their support for and management of teachers. Evidence has shown a relationship between principal effectiveness and student academic outcomes:<sup>35</sup> A recent meta-analysis estimated the impact of having an effective principal on student learning was nearly as large as having an effective teacher.<sup>36</sup>

Principals also play a role in teacher recruitment and retention,<sup>37</sup> retaining effective teachers and exiting consistently low-performing teachers,<sup>38</sup> and shaping teachers' experiences of school climate.<sup>39</sup> They also influence students' perception of school climate, student attendance, in-school discipline, and parents' perceptions of the school.<sup>40</sup> Given their critical role to both students and teachers, principals must receive meaningful feedback and opportunities for support through comprehensive evaluations. As with teachers, these systems can also serve to identify exemplary principals from whom others can learn, to support those who are struggling, and to ultimately exit leaders who do not improve with time and support.

### **What role does the state play in designing principal evaluations?**

Fourteen states set all criteria for principal evaluations, while 21 states set minimum criteria for what is included, and 16 states play no role in designing principal evaluations. As is the case for teacher evaluations, some flexibility for districts in designing evaluations may be useful; however, setting no shared standards or measures defining the central elements of a principal's job risks differing expectations, inconsistent attention to the core responsibilities of the job, and highly varied evaluation implementation in different communities across the state.

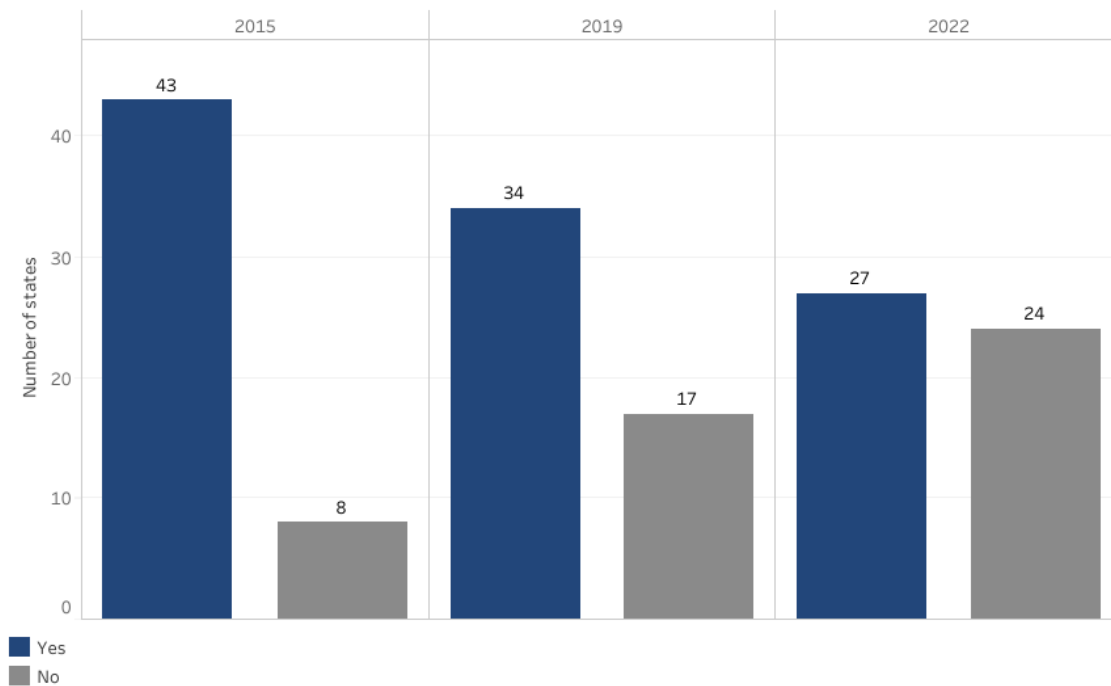




**Oregon**, and **South Dakota** removed requirements to include measures of student growth in principal evaluation. Interestingly, fewer states require measures of student growth to be included in principals' evaluations (27) compared to teachers' (30).

*Figure 15.*

### Do states require measures of student growth in principal evaluations?



### State assessments to measure student learning

Research suggests that principals have a major impact, both direct and indirect, on student achievement.<sup>43</sup> Only 10 states factor state assessments into a principal's evaluation score, compared to 12 that require these tests to be reflected in teachers' evaluation scores.

### Surveys

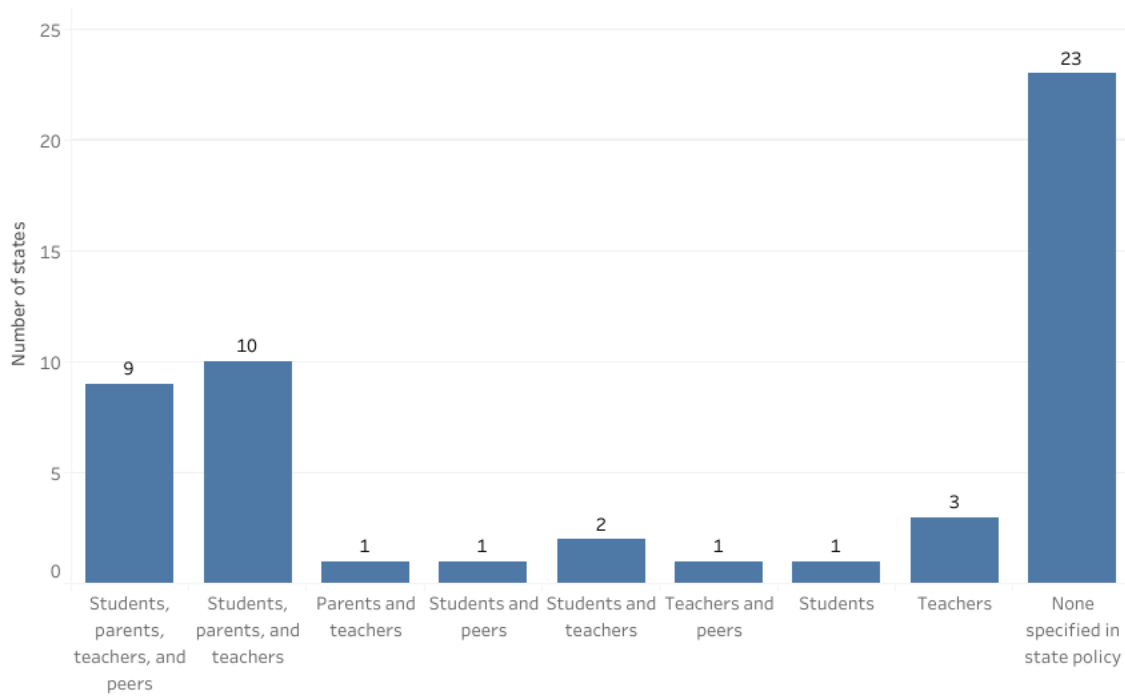
Principals play a key role in influencing the overall climate of a school, and at least one study has concluded that a principal's biggest influence on student learning is mediated through their ability to create a positive school climate.<sup>44</sup> Other studies have validated the importance of principals' leadership and influence on school climate to retaining teachers, too.<sup>45</sup> Fostering healthy school climates that re-engage and support students in the wake of several years of widespread trauma and disruption is critical; so too is fostering a school climate that prevents teacher burnout and motivates teachers to stay.

Given this, survey data from students, teachers, and the wider school community can be a valuable tool in helping provide feedback to principals and to measure a principal's success. Twenty-eight states explicitly allow or require surveys to be included in principal

evaluations in some form. (For a breakdown of what kind of surveys are permitted, see Figure 16.) The state of **Michigan**, for instance, requires a mix of feedback from students, teachers, and parents all be included in a principal’s evaluation score. Since 2019, at least one state that had previously required surveys for principal evaluations, **Georgia**, dropped this requirement.

Figure 16.

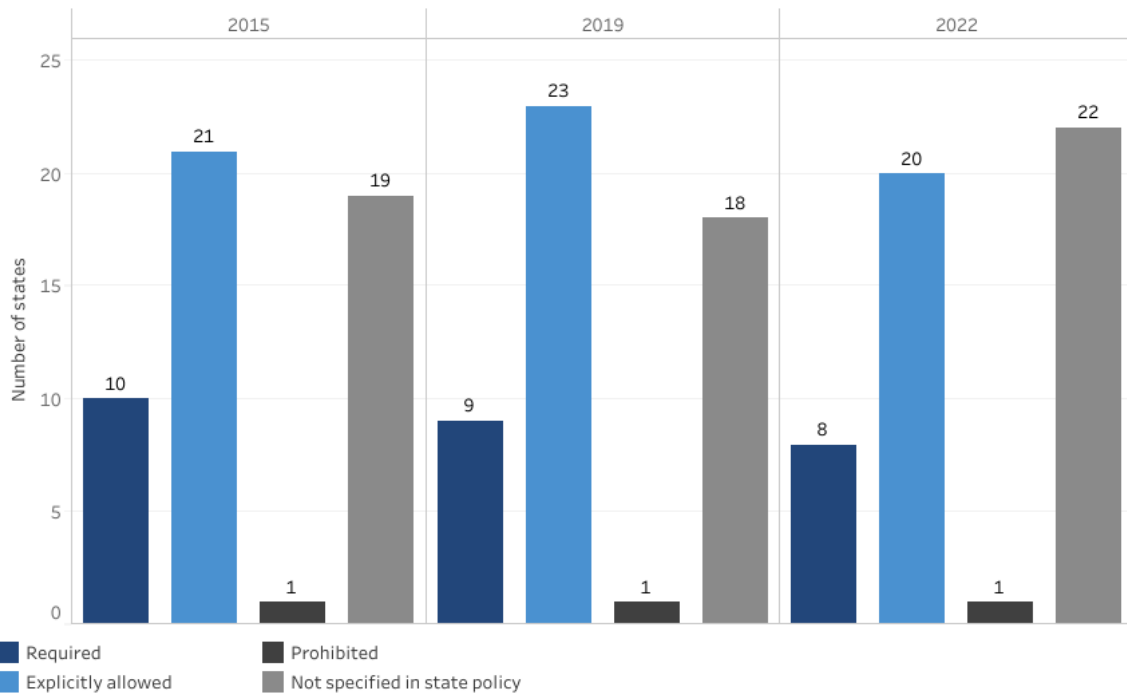
**What types of surveys are required or explicitly allowed as part of a principal's evaluation score?**



*Note: Credit is given to states that require input from students, parents, teachers, and peers, and this feedback may be in the form of a survey.*

Figure 17.

## What is the role of surveys in principal evaluations?



*Note: Surveys may include student, parent, teacher, and/or peer surveys.*

### Link to instructional leadership

Much of the conversation on principal quality in recent years has centered around enhancing the role of principals as instructional leaders, setting a standard for strong instruction across the school, and helping teachers meet that standard. The urgency of instructional leadership has only heightened in the wake of the pandemic, but evidence suggests that far too few principals feel that they are able to fulfill that aspect of their role, given the many competing demands on their time.<sup>46</sup> Despite the importance of clarifying a principal's role in instructional leadership, many states have failed to use evaluation to signal that it is a priority: 18 states still do not explicitly link principal evaluations to their role as instructional leaders by including specific criteria related to this role in their evaluations.

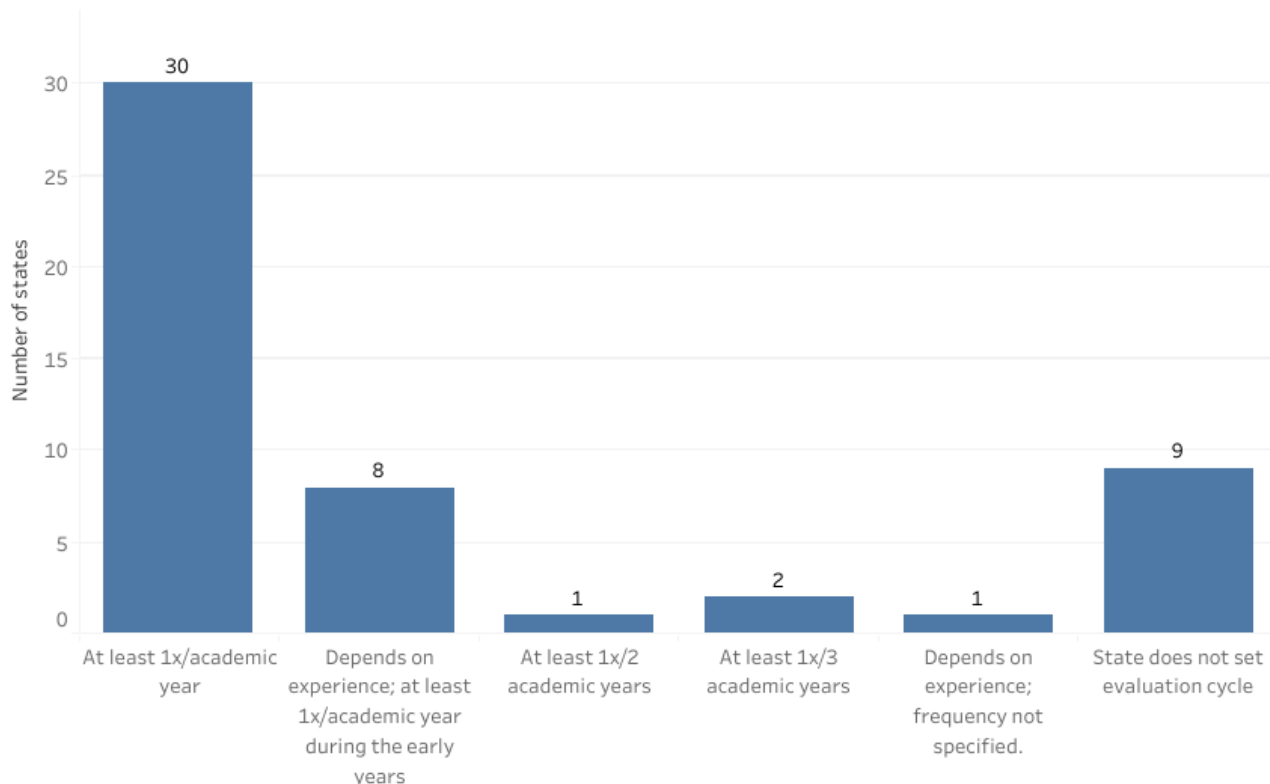
### How often are principals evaluated?

Like teachers, principals need to receive regular, actionable feedback and formal evaluations of their performance. Thirty states require that principals are evaluated each year, while eight set the frequency of evaluation based on principals' years of experience, with states more likely to require evaluation in the early years of their career. One implementation challenge that some states may face is that principal employment contract cycles and evaluation cycles do not line up. For example, in some districts,

principals may be on a three-year employment cycle but a two-year evaluation schedule, meaning that feedback cycles and employment decisions are not aligned.

Figure 18.

### How frequently are principal evaluations required?

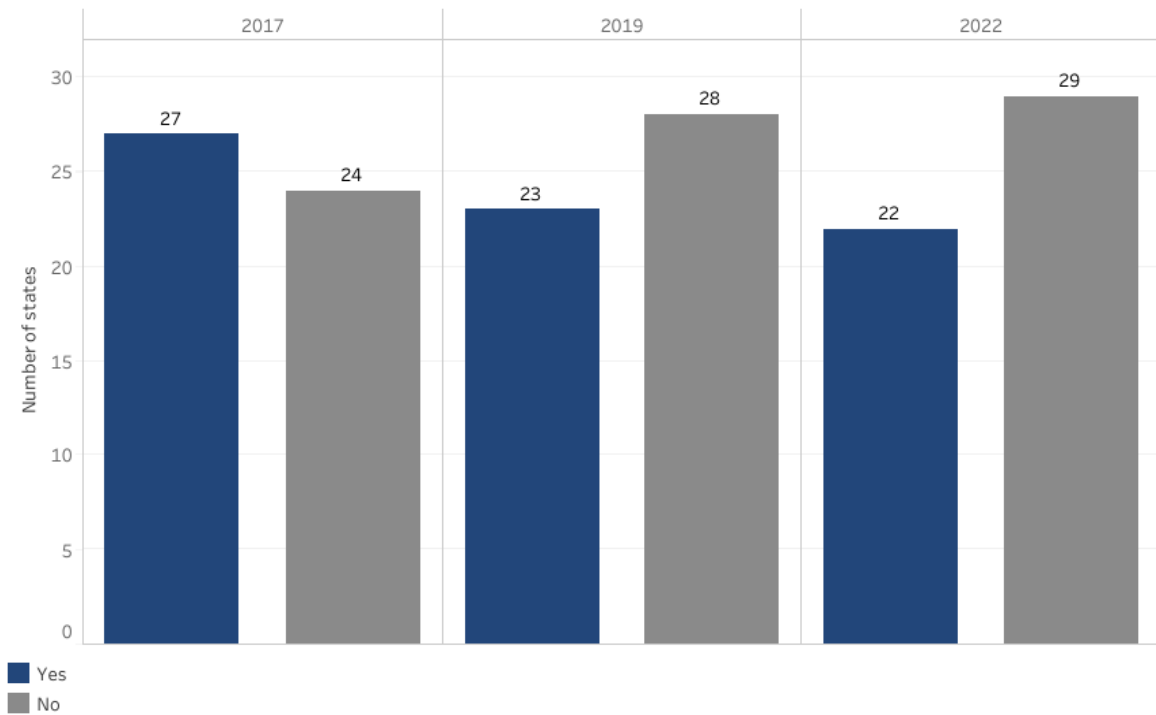


### Are principal evaluations used for support and improvement?

If evaluation systems are designed to help principals hone their practice and improve student learning, then they must be linked to improvement systems. Yet too few states require that principals with less-than-effective ratings are placed on improvement plans: 22 states require improvement plans, while 29 either do not require improvement plans as remediation for low ratings or do not have a system of improvement plans at all. Since 2019, **Georgia, Mississippi, Ohio, and Virginia** have added new requirements, while **Nevada, New Mexico, South Carolina, and Utah** removed requirements for improvement plans for less-than-effective principals.<sup>47</sup> These policy shifts have resulted in one fewer state overall requiring improvement plans for principals deemed ineffective.

Figure 19.

## Do states require improvement plans for principals with less-than-effective ratings?



Note: Four states (**Massachusetts, Michigan, Nevada, and Washington**) require that high-performing principals are evaluated less frequently.

## RECOMMENDATIONS

Evidence supports key policies and practices in evaluation that improve teacher and principal skills and ultimately student outcomes: use of multiple measures (including student surveys and academic growth measures), regular opportunities for feedback, and more. (For a comprehensive list of policy conditions that standout systems have in common, see Figure 20.) States have a central role to play in both setting policy conditions and supporting effective implementation. We recommend high-leverage state policies and practices below.

Figure 20.

## Components of a strong evaluation system

- ✓ Multiple measures
- ✓ Student surveys
- ✓ Objective measures of student growth
- ✓ At least three rating categories
- ✓ Annual observations and evaluations for all teachers
- ✓ Professional development tied to evaluation
- ✓ Written feedback after each observation

Source: Putman, H., Ross, E., & Walsh, K. (2018).

---

## Policy recommendations

### Focus on student growth

Coming out of pandemic disruptions, states should begin with a renewed commitment to accelerating academic growth, and reflect the importance of this goal when designing teacher and principal evaluations. Student growth should be included as part of a range of evidence-based multiple measures, like surveys and observations. In response to concerns over gaps in available student data, states may consider temporary adjustments to their student growth model, such as expanding the years within the model, rather than eliminate it.

### Require multiple observations, regular feedback, and annual evaluations

Multiple observations, regular feedback, and required annual evaluations for all teachers are important elements of effective evaluation systems that contribute to increased student learning.<sup>48</sup> High-quality evaluation can be part of a comprehensive effort to ensure that all new teachers receive the regular feedback they need. Research suggests that frequent observations that are followed by timely, specific feedback have a discernible impact on teachers' improved practice.<sup>49</sup> States play an important role in setting requirements for the timing and content of evaluations.

### Support new teachers

At a time when retaining and supporting effective teachers is taking on even more urgency, all teachers (but especially new teachers) deserve supportive, actionable feedback and opportunities for growth and development on a regular basis. Data consistently shows that early career teachers have the highest attrition rate, yet they are too often left alone in their classrooms with little support.<sup>50</sup> States can require that novice teachers receive more opportunities for observation and feedback, as they do in

**Delaware, New Mexico, Ohio, West Virginia, and Wisconsin**, all of which require four observations for novice teachers each year.

Consistent evaluations with multiple opportunities to see teachers' practice and provide feedback contribute to teachers' growth and development. This is particularly important during a time when many states have lowered standards for entry into the teaching profession, allowing some new teachers to take on responsibility for classrooms without demonstrating they have mastered the knowledge and skills necessary to be successful.<sup>51</sup> If states choose to pursue policies that allow less-prepared candidates to enter, then they must simultaneously commit to policies that support and evaluate these new teachers early and regularly.

### **Collect and publish statewide data**

Collecting and publishing effectiveness data is critical to understanding which students have access to impactful teachers (and which students do not). There is widespread and long-standing evidence that effective teachers are distributed inequitably,<sup>52</sup> and evaluation data is critical to identifying and ameliorating these gaps.<sup>53</sup> States can begin to right these inequities by using evaluation data to see where they exist. Following the example set by states like **Colorado, Arkansas, and Kentucky**, states can develop systems that support improvement. Further, when states collect evaluation data, they hold districts accountable for providing teachers with feedback and evaluation—a core responsibility of state agencies.

### **Measure what matters for principals**

Principals' impact on both school climate and educator satisfaction has a demonstrated relationship to improved student outcomes, teacher success, and retention,<sup>54</sup> and they should be measured accordingly. To gain a comprehensive picture of a principal's impact, states should consider multiple measures of principal effectiveness, including surveys and measures of student learning. States can also explore new methods to measure student learning in principal evaluation, such as including multiple years of data in a principal's evaluation,<sup>55</sup> or temporarily adjusting current growth models to account for any gaps in student data.

### **Design systems with consequences**

Evaluation systems should recognize strong performance through incentives like sizable bonuses<sup>56</sup> and provide real opportunities for teachers and principals to improve. For teachers and principals who do not improve over time, state policies should provide a clear process to exit.

---



## How can states support quality implementation?

Beyond setting overarching policy conditions, states may also be searching for additional levers to ensure that their evaluation systems are implemented well. Though much of the work of making evaluation meaningful and effective happens in schools and districts, states have a range of tools available to them to support quality implementation. Of the systems that have had meaningful success, what most had in common were sustained investment that lasted beyond one system leader; a commitment to getting all stakeholders invested in the system; and a commitment and follow-through to iterating and evaluating the evaluation system, and improving over time.<sup>57</sup> Below, we offer additional steps for states working to improve how their evaluation system operate in practice:

### **Analyze and act on statewide data**

In order to understand how their evaluation systems are working, states should collect, analyze, and report on evaluation ratings. States can use this data to answer (and ultimately, act on) key questions about the relationship between evaluation data and student growth and achievement; identify inequities for teachers (see below); and target the inequitable distribution of teacher talent. States like **Colorado** have made progress in collecting and analyzing evaluation data from across the state in order to better understand the distribution and assignment of teachers.

### **Address disproportionate impact**

State and local policymakers must take issues of potential racial bias in evaluation systems seriously—an effort vital to fundamental fairness and equity, and to ensuring that states support and retain the diverse teacher workforce that their students deserve. Recent evidence suggests that gaps exist in evaluation scores for teachers of color—in some cases, researchers have traced these disparities to observer bias, finding that white observers systematically assign Black teachers lower ratings, or that observers were more likely to assign higher scores to teachers of the same race in general,<sup>58</sup> while others have found that racial gaps in observation scores could be traced directly back to the student populations that teachers of color were more likely to teach in the first place.<sup>59</sup>

- Researchers have emphasized that their findings are not cause to discontinue evaluation, but to continuously improve systems, increasing fairness, equity, and trust.<sup>60</sup> While there is not yet strong evidence on interventions that work to ameliorate systemic bias in evaluation systems as a whole, researchers have suggested that using an evaluation system with multiple measures could partially mitigate this risk.<sup>61</sup> To address observer bias, states could explore efforts to diversify the general pool of observers, increase the overall validity and reliability of observations by using multiple observers over the course of multiple observations,<sup>62</sup> and calibrate observations (see recommendation below).
- Before taking action, states need to understand the existing data. They should begin by requiring that districts submit teacher evaluation data annually, disaggregated by teacher subgroups, and analyze that data at the state level to

understand impact and identify any disproportionate effects related to teacher demographic characteristics.<sup>63</sup> In service of transparency, systems can follow the lead of **District of Columbia Public Schools**, which has published highly detailed data on the trends in equity and mitigating implicit bias in its evaluation system.<sup>64</sup>

### **Collect user feedback**

States should aim to understand what both principals and teachers think about the usefulness of the evaluation systems they use, and get their view of how evaluation systems are implemented. **Idaho**, for instance, has conducted annual surveys of K-12 administrators to gauge how well they implement evaluation requirements like timing and number of observations. While the state's most recent survey found troubling gaps in practice, it has been able to document some improvement over several years.<sup>65</sup>

### **Focus on continuous improvement**

Using the feedback they collect, states can adjust evaluation systems and make improvements over time to their evaluation policies and practices. **Tennessee** stands out as a notable example of a state that made marked improvement in building trust as it worked to refine its evaluation system. As detailed in NCTQ's 2018 Report [Making a Difference](#), Tennessee continually refined its evaluation system in response to educator feedback, meeting with over 7,500 educators to incorporate their input. Ultimately, the state made impressive gains in teachers' beliefs about their evaluation systems: In 2012, only 38% of Tennessee teachers surveyed said that their school's teacher evaluation process led to improvements in their teaching, a number that rose to 72% by 2018.<sup>66</sup>

### **Sponsor statewide evaluator training**

In order to influence the quality of evaluator preparation, state agencies can also provide access to statewide training to all evaluators, though this would depend on having statewide common expectations or common elements within the evaluation system. **Delaware**, for instance, provides annual statewide training for all evaluators.<sup>67</sup>

### **Certify and calibrate observer skills**

In order to promote access to effective evaluators, states can require that observers demonstrate their knowledge and skill as an observer through a state certification process, and provide resources to evaluators in order to calibrate ratings effectively. **Texas**, as part of the state Teacher Incentive Allotment, partners with [Texas Tech](#) as a third-party reviewer of evaluation data to validate its accuracy. **Massachusetts** offers an optional resource known as the [Online Platform for Teaching and Informed Calibration \(OPTIC\)](#) to support educators and evaluators to build a shared understanding of high-quality instruction and improve the feedback that teachers receive. OPTIC uses video and interactive displays as part of a dynamic calibration training experience for both evaluators and teachers aligned to the standards for effective teacher practice and the standards for student learning.

### **Link to teacher preparation**

To increase alignment between expectations for pre-service teachers and in-service teachers, states can require that all teacher preparation programs in the state use

measures of performance for teacher candidates that are aligned to the state's professional teaching standards and evaluation standards. **Massachusetts**, for instance, uses the Massachusetts Candidate Assessment of Performance (CAP), a practice-based assessment that aligns the evaluation of pre-service candidates to the teacher evaluation of in-service teachers. In a 2019 study, the CAP was found to be predictive of future scores on teachers' in-service evaluation scores.<sup>68</sup>

---

## DATA

### [Download full dataset](#)

Download the full teacher and principal evaluation policy data collected by NCTQ and used in this analysis.

---

## ACKNOWLEDGEMENTS

### **Authors**

Abigail Swisher and Dr. Patricia Saenz-Armstrong

### **Data collection and analysis**

Kelli Lakis and Lisa Staresina

### **Project leadership**

Dr. Heather Peske, NCTQ President  
Shannon Holston, Chief of Policy and Programs  
Hannah Putman, Managing Director of Research

### **Communications and advocacy**

Nicole Gerber, Ashley Kincaid, Andrea Browne Taylor, and Shayna Levitan

### **Reviewers**

*Special thanks to the following individuals for providing review and feedback on this project. Inclusion does not imply endorsement.*

Dr. Matthew A. Kraft  
Associate Professor of Education and Economics  
Brown University

Ron Noble, Jr.  
Assistant Superintendent  
Methuen Public Schools

## Project funders

*This report is based on research funded by the following foundations. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the project funders.*

Daniels Fund  
The Joyce Foundation

**Suggested citation:** Swisher, A. & Saenz-Armstrong, P. (2022). *State of the States 2022: Teacher and Principal Evaluation Policies*. Washington, D.C.: National Council on Teacher Quality.

---

## ENDNOTES

<sup>1</sup> National Center for Education Statistics (2022). *NAEP long-term trend assessment results: Reading and Mathematics*. The Nation's Report Card. Retrieved August 29, 2022, from <https://www.nationsreportcard.gov/highlights/ltt/2022/>.

<sup>2</sup> Successful evaluation systems have not always fulfilled all of these functions at once; however, there are solid proof points that evaluation systems can fulfill these functions. See Dee, T. S., James, J., & Wyckoff, J. (2021) Is Effective Teacher Evaluation Sustainable? Evidence from District of Columbia Public Schools. *Education Finance and Policy*, 16(2): 313–46; Adnot, M., Dee, T. S., Katz, V., & Wyckoff, J. (2016). Teacher Turnover, Teacher Quality, and Student Achievement in DCPS. *Educational Evaluation and Policy Analysis*, 0162373716663646; Dee, T. S. & Wyckoff, J. (2015). Incentives, Selection, and Teacher Performance: Evidence from IMPACT. *Journal of Policy Analysis and Management*, 34(2): 267–97; Dotter, D., Chaplin, D. D., & Bartlett, M. (2021). *Impacts of School Reforms in Washington, DC on Student Achievement*. Mathematica Policy Research; Steinberg, M. P. & Sartain, L. (2015). Does teacher evaluation improve school performance? Experimental evidence from Chicago's Excellence in Teaching project. *Education Finance and Policy*, 10(4), 535-572; Sartain, L. & Steinberg, M. P. (2021). *Can Personnel Policy Improve Teacher Quality? The Role of Evaluation and the Impact of Exiting Low-Performing Teachers*. (EdWorkingPaper: 21-486). Retrieved from Annenberg Institute at Brown University: <https://doi.org/10.26300/d201-7y89>.

<sup>3</sup> Bleiberg, J., Brunner, E., Harbatkin, E., Kraft, M. A., & Springer, M. (2021). *The Effect of Teacher Evaluation on Achievement and Attainment: Evidence from Statewide Reforms*. (EdWorkingPaper: 21-496). Retrieved from Annenberg Institute at Brown University: <https://doi.org/10.26300/b1ak-r251>.

<sup>4</sup> Dee, T. S., James, J., & Wyckoff, J. (2021); Adnot, M., Dee, T. S., Katz, V., & Wyckoff, J. (2016); Dee, T. S. & Wyckoff, J. (2015); Dotter, D., Chaplin, D. D., & Bartlett, M. (2021); Steinberg, M. P. & Sartain, L. (2015); Sartain, L. & Steinberg, M. P. (2021); Commit! (2017). *Overview of Dallas ISD STAAR Achievement at "Meets" Post-Secondary Standard Across Various Demographics and Subjects 2012-2017*. Dallas, TX: Commit; Denver Public Schools. (2018). *DPS Students Again Outpace State in Academic Growth on 2018 CMAS*. Retrieved from: <https://www.dpsk12.org/dps-students-again-outpace-state-in-academic-growth-on-2018-cmas/>; Tennessee Department of Education cited in Putman, H., Ross, E., & Walsh, K. (2018). *Making a Difference: Six Places Where Teacher Evaluation Systems Are Getting Results*. Washington, D.C.: National Council on Teacher Quality. <https://www.nctq.org/publications/Making-a-Difference>.

<sup>5</sup>Putman, H., Ross, E., & Walsh, K. (2018), see Appendix: Key Components of an Evaluation System.

<sup>6</sup> These six systems and states validated by Annenberg were identified and profiled by Putman, H., Ross, E., & Walsh, K. (2018). Notably, implementation of any evaluation system will be mediated by the contextual factors of the ecosystem in which the policy is implemented, such as leadership quality.

<sup>7</sup> Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Identifying effective classroom practices using student achievement data. *Journal of Human Resources*, 46(3), 587-613; Taylor, E. S., & Tyler, J. H. (2012); Cantrell, S. & Kane, T. J. (2013). *Ensuring Fair and Reliable Measures of Effective Teaching: Culminating Findings from the MET Project's Three-Year Study*. Seattle, WA: Bill & Melinda Gates Foundation, Policy and Practice Brief, Measures of Effective Teaching project.

<sup>8</sup> Putman, H., Ross, E., & Walsh, K. (2018).

<sup>9</sup> Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011); Marsh, J. A., Bush-Mecenas, S., Strunk, K. O., Lincove, J. A. & Huguet, A. (2017). Evaluating Teachers in the Big Easy: How Organizational Context Shapes Policy Responses in New Orleans. *Educational Evaluation and Policy Analysis*, 39(4), 539-570; Stecher, B. M., Garet, M. S., Hamilton, L. S., Steiner, E. D., Robyn A., Poirier, J., Holtzman, D. J., Fulbeck, E. S., Chambers, J., & Brodzia de los Reyes, I. (2016). *Improving Teaching Effectiveness: Implementation: The Intensive Partnerships for Effective Teaching Through 2013-2014*. RAND Corporation. Retrieved from: [https://www.rand.org/pubs/research\\_reports/RR1295.html](https://www.rand.org/pubs/research_reports/RR1295.html); Strunk, K. O., Weinstein, T. L., & Makkonen, R. (2014). Sorting Out the Signal: Do Multiple Measures of Teachers' Effectiveness Provide Consistent Information to Teachers and Principals? *Education Policy Analysis Archives*, 22(100), As of May 11, 2018: <http://www.redalyc.org/html/2750/275031898100>; Taylor, E. S. & Tyler, J. H. (2012).

<sup>10</sup> Tuma, A. P., Hamilton, L. S., & Tsai, T. (2018). *How Do Teachers Perceive Feedback and Evaluation Systems?: Findings from the American Teacher Panel*. RAND Corporation. Retrieved from: [https://www.rand.org/pubs/research\\_briefs/RB10023.html](https://www.rand.org/pubs/research_briefs/RB10023.html).

<sup>11</sup> Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011); Taylor, E. S. & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *The American Economic Review*, 102(7), 3628-3651; Ho, A. D. & Kane, T. J. (2013). *The Reliability of Classroom Observations by School Personnel*. Research Paper. Seattle, WA: Bill & Melinda Gates Foundation, Measures of Effective Teaching project; Steinberg, M. P. & Sartain, L. (2021). What Explains the Race Gap in Teacher Performance Ratings? Evidence From Chicago Public Schools. *Educational Evaluation and Policy Analysis*, 43(1), 60-82, <https://doi.org/10.3102/0162373720970204>; Chi, O. L., (2021) A Classroom Observer Like Me: The Effects of Race-congruence and Gender-congruence Between Teachers and Raters on Observation Scores. *Education Finance and Policy*, [https://doi.org/10.1162/edfp\\_a\\_00367](https://doi.org/10.1162/edfp_a_00367); Grissom, J. A., Bartanen, B., & Jones, A. A. (2019). *Retaining Teachers of Color in an Era of High-Stakes Teacher Evaluation: Investigating Racial Differences in Teacher Evaluation Ratings and Teacher Turnover*. Working Paper, Vanderbilt University. Retrieved from: [https://cdn.vanderbilt.edu/vu-my/wp-content/uploads/sites/2824/2019/04/14200330/retentionoftoc\\_grissom\\_bartanen\\_jones.pdf](https://cdn.vanderbilt.edu/vu-my/wp-content/uploads/sites/2824/2019/04/14200330/retentionoftoc_grissom_bartanen_jones.pdf).

<sup>12</sup> Teachers who influence students' academic achievement make an impact on students that extends beyond their short-term achievement, influencing, for example, students' likelihood of pursuing postsecondary education and later earnings. See, for instance: Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9), 2633-79; Jackson, C. K. (2012). *Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in North Carolina* (Working Paper No. 18624). Cambridge, MA: National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w18624>.

<sup>13</sup> Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011); Taylor, E. S. & Tyler, J. H. (2012).

<sup>14</sup> NCTQ's analysis has always included states with explicit requirements in policy for student growth, including policies adopted but not yet implemented. During the time that it was written in state policy, Mississippi's requirement to include objective measures of student growth was never implemented.

<sup>15</sup> Our analysis of state policy documents could not give a precise timeline for every state that dropped assessments between 2019 and 2022, as some state policy changes were undated.

<sup>16</sup> Holston, S. (2020, November 9). Evaluating teachers during the pandemic. National Council on Teacher Quality. <https://www.nctq.org/blog/Evaluating-teachers-during-the-pandemic>; Nittler, K. & Saenz-Armstrong, P. (2020, May 1). Teacher evaluations and support during COVID-19 closures. National Council on Teacher Quality. <https://www.nctq.org/blog/Teacher-evaluations-and-support-during-COVID--19-closures>.

<sup>17</sup> Gewertz, C. (2021, November 16). State Test Results Are In. Are They Useless? *Education Week*. Retrieved October 21, 2022, from <https://www.edweek.org/teaching-learning/state-test-results-are-in-are-they-useless/2021/10>.

<sup>18</sup> Wallace, T., Kelcey, B., & Ruzek, E. (2016). What can student perception surveys tell us about teaching? Empirically testing the underlying structure of the Tripod student perception survey. *American Educational Research Journal*, 53(6), 1834-1868, <https://files.eric.ed.gov/fulltext/ED540960.pdf>; Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment*. Research Paper. Seattle, WA: Bill & Melinda Gates Foundation, Measures of Effective Teaching project.

<sup>19</sup> Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. TNTP. Retrieved from: <https://tntp.org/publications/view/the-widget-effect-failure-to-act-on-differences-in-teacher-effectiveness>.

<sup>20</sup> Bleiberg, J., Brunner, E., Harbatkin, E., Kraft, M. A., & Springer, M. (2021); Putman, H., Ross, E., & Walsh, K. (2018).

<sup>21</sup>For the purposes of this data collection, teachers' probationary status was defined by state definitions of probationary, which may vary, but often refer to teachers early in their career, not yet fully certified or in the first tier of licensure.

<sup>22</sup> Cantrell, S. & Kane, T. J. (2013); Kane, T. J., & Staiger, D. O. (2012). *Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains*. Research Paper. Seattle, WA: Bill & Melinda Gates Foundation, Measures of Effective Teaching project.

<sup>23</sup> Tuma, A. P., Hamilton, L. S., & Tsai, T. (2018)

<sup>24</sup> Of note, only five states (Delaware, New Mexico, Ohio, West Virginia, and Wisconsin) actually require four or more observations yearly for new teachers.

<sup>25</sup> Cantrell, S. & Kane, T. J. (2013)

<sup>26</sup> Steinberg, M. P., & Sartain, L. (2015)

<sup>27</sup> Our analysis classified virtual observations broadly, to include policies from fully virtual, to live video, or recordings. Some of these policies specify that virtual observations specifically apply to virtual and hybrid classrooms, while others are more general.

<sup>28</sup> For our purposes here, remote and virtual are used interchangeably. States use a variety of terms in their observation policies.

<sup>29</sup>Of note, the teacher must still receive one unannounced observation: <https://nj.gov/education/AchieveNJ/teacher/iqt/execution/reflective.pdf>.

<sup>30</sup> Kane, T., Blazar, D., Gehlbach, H., Greenberg, M., Quinn, D., & Thal, D. (2020). Can Video Technology Improve Teacher Evaluations? An Experimental Study. *Education Finance and Policy*, 15(3): 397-427. [https://doi.org/10.1162/edfp\\_a\\_00289](https://doi.org/10.1162/edfp_a_00289).

<sup>31</sup> Shaha, S. H., Glassett, K. F., & Copas, A. (2015). The impact of teacher observations with coordinated professional development on student performance: A 27-state program evaluation. *Journal of College Teaching & Learning*, 12(1), 55; Hunter, S. B. (2022). High-leverage teacher evaluation practices for instructional improvement. *Educational Management Administration & Leadership*. doi:10.1177/17411432221112995.

<sup>32</sup> Sass, T. R., Hannaway, J., Xu, Z., Figlio, D. N., & Feng, L. (2012). Value added of teachers in high-poverty schools and lower poverty schools. *Journal of Urban Economics*, 72(2-3), 104-122; Steele, J. L., Pepper, M. J., Springer, M. G., & Lockwood, J. R. (2015). The distribution and mobility of effective teachers: Evidence from a large, urban school district. *Economics of Education Review*, 48(1), 86-101; Goldhaber, D., Quince, V., & Theobald, R. (2018). Has it always been this way? Tracing the evolution of teacher quality gaps in US public schools. *American Educational Research Journal*, 55(1), 171-20; Goldhaber, D., Lavery, L., & Theobald, R. (2015). Uneven Playing Field? Assessing the Teacher Quality Gap Between Advantaged and Disadvantaged Students. *Educational Researcher*, 44(5), 293-307; Glazerman, S. & Max, J. (2011). *Do low income students have equal access to the highest performing teachers?* Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance; Cooc, N. & Yang, M. (2016). Diversity and equity in the distribution of teachers with special education credentials: Trends from California. *AERA Open*, 2(4), 2332858416679374, <https://caldercenter.org/sites/default/files/CALDER%20WP%20259-0122.pdf>.

In rare instances, research has found little evidence of inequitable access to effective teachers, although that may in part be due to the way equity is measured, such as looking for differences within rather than across districts. See: Isenberg, E., Max, J., Gleason, P., & Deutsch, J. (2022). Do Low-Income Students Have Equal Access to Effective Teachers? *Educational Evaluation and Policy Analysis*, 44(2), 234-256. <https://doi.org/10.3102/01623737211040511>.

<sup>33</sup> Saenz-Armstrong, P. (2021). *State of the States 2021: State Reporting of Teacher Supply and Demand Data*. Washington, D.C.: National Council on Teacher Quality.

<sup>34</sup> Grissom, J. A., Egalite, A. J., & Lindsay, C. A. (2021). *A Systematic Synthesis of Two Decades of Research*. New York: The Wallace Foundation. Retrieved from <http://www.wallacefoundation.org/principalsynthesis>.

<sup>35</sup> Wu, H. & Shen, J. (2021). The association between principal leadership and student achievement: A multivariate meta-meta-analysis. *Educational Research Review*, 100423; Clifford, M., Hansen, U. J., & Wraight, S. (2014). *Practical guide to designing comprehensive principal evaluation systems: A tool to assist in the development of principal evaluation systems*. Center on Great Teachers and Leaders; Rice, J. K. (2010). *Principal effectiveness and leadership in an era of accountability* (Brief 8). National Center for Analysis of Longitudinal Data in Education Research; Glasman, N. S. & Heck, R. H. (1992). The changing leadership role of the principal: Implications for principal assessment. *Peabody Journal of Education*, 68(1), 5-24; Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004). *How leadership influences student learning: A review of research for the Learning from Leadership Project*. New York: The Wallace Foundation.

<sup>36</sup> Grissom, J. A., Egalite, A. J. & Lindsay, C. A. (2021)

<sup>37</sup> Boyd, D., Grossman, P., Ing, M., Lankford, H., Loeb, S., & Wyckoff, J. (2011). The influence of school administrators on teacher retention decisions. *American Education Research Journal*, 48(2), 303-333; Kimball, S. (2011). Strategic talent management for principals. *Strategic management of human capital in education: Improving instructional practice and student learning in schools* (pp. 133-152). New York, NY: Routledge Publishing; Rice, J. K. (2010); Clark, D., Martorell, P., & Rockoff, J. (2009). *School principals and school performance* (No. w17803). National Bureau of Economic Research; Ingersoll, R. M. (2001). *A different approach to solving the teacher shortage problem*. Center for the Study of Teaching and Policy, University of Washington; Ladd, H. (2011). Teachers' perceptions of their working conditions: How predictive of planned and actual teacher movement? *Educational Evaluation and Policy Analysis*, 33(2), 235-261; Luekens, M. T., Lyter, D. M., Fox, E. E., & Chandler, K. (2004). *Teacher attrition and mobility: Results from the teacher follow-up survey, 2000-01*. National Center for Education Statistics, <https://link.springer.com/article/10.1007/s10984-015-9198-x>.

- <sup>38</sup> Beteille, T., Kalogrides, D., & Loeb, S. (2009). *Effective schools: Managing the recruitment, development, and retention of high-quality teachers* (Working Paper 37). National Center for Analysis of Longitudinal Data in Education Research.
- <sup>39</sup> Kraft, M. A., Marinell, W. H., & Shen-Wei Yee, D. (2016). School Organizational Contexts, Teacher Turnover, and Student Achievement: Evidence From Panel Data. *American Educational Research Journal*, 53(5), 1411–1449. <https://doi.org/10.3102/0002831216667478>.
- <sup>40</sup> Allensworth, E. & Hart, H. (2018, March). *How Do Principals Influence Student Achievement?* University of Chicago Consortium on School Research. Retrieved from: <https://consortium.uchicago.edu/publications/how-do-principals-influence-student-achievement>; Bartanen, B. (2020). Principal Quality and Student Attendance. *Educational Researcher*, 49(2), 101–113. <https://doi.org/10.3102/0013189X19898702>; Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2012). *Estimating the effect of leaders on public sector productivity: The case of school principals* (No. w17803). National Bureau of Economic Research; Louis, K. S., Leithwood, K., Wahlstrom, K. L. Anderson, S. E., Michlin, M., & Mascall, B. (2010). *Learning from leadership: Investigating the links to improved student learning*. Center for Applied Research and Educational Improvement/University of Minnesota and Ontario Institute for Studies in Education/University of Toronto; Clark, D., Martorell, P., & Rockoff, J. (2009); Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004).
- <sup>41</sup> Rice, J. K. (2010); Wu, H., Shen, J. (2022).
- <sup>42</sup> Bartanen, B., Husain, A. N., & Liebowitz, D. D. (2022). *Rethinking Principal Effects on Student Outcomes*. (EdWorkingPaper: 22-621). Retrieved from Annenberg Institute at Brown University: <https://doi.org/10.26300/r5sf-3918>.
- <sup>43</sup> Grissom, J. A., Egalite, A. J. & Lindsay, C. A. (2021); Wu, H. & Shen, J. (2021); Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004).
- <sup>44</sup> Allensworth, E. & Hart, H. (2018, March); Sebastian, J., Huang, H., & Allensworth, E. (2016). The role of teacher leadership in how principals influence classroom instruction and student learning. *American Journal of Education*, 123(1), 69-108.
- <sup>45</sup> Kraft, M. A., Marinell, W. H., & Shen-Wei Yee, D. (2016); Aldridge, J. M., & Fraser, B. J. (2016). Teachers' views of their school climate and its relationship with teacher self-efficacy and job satisfaction. *Learning Environments Research*, 19(2), 291-307; Kraft, M. A. & Papay, J. P. (2014). Can professional environments in schools promote teacher development? Explaining heterogeneity in returns to teaching experience. *Educational Evaluation and Policy Analysis*, 36(4), 476-500; Aldridge, J. M. & Fraser, B. J. (2016). Teachers' views of their school climate and its relationship with teacher self-efficacy and job satisfaction. *Learning Environments Research*, 19(2), 291-307.
- <sup>46</sup> Tooley, M. (2017). *From Frenzied to Focused: How School Staffing Models Can Support Principals as Instructional Leaders*. New America. Retrieved from: <https://www.newamerica.org/education-policy/policy-papers/frenzied-focused>.
- <sup>47</sup> One consideration is that principal supervisors may more readily use the termination clause in a principal's contract more readily than using the evaluation process to influence principal performance.
- <sup>48</sup> Putman, H., Ross, E., & Walsh, K. (2018). See Appendix: Key Components of an Evaluation System.
- <sup>49</sup> Scheeler, M. C., Ruhl, K. L., & McAfee, M. K. (2004). Providing performance feedback to teachers: A review. *Teacher Education and Special Education*, 27(4), 396–407. <https://doi.org/10.1177/088840640402700407>; Thurlings, M., Vermeulen, M., Bastiaens, T., & Stijnen, S. (2013). Understanding feedback: A learning theory perspective. *Educational Research Review*, 9, 1–15. <https://doi.org/10.1016/j.edurev.2012.11.004>; Kraft, M. A. & Christian, A. (2022). Can Teacher Evaluation Systems Produce High-Quality Feedback? An Administrator Training Field Experiment. *American Educational Research Journal*, 59(3), 500–537. <https://doi.org/10.3102/00028312211024603>.



- <sup>50</sup>Ingersol, R. M., Merrill, E., Stuckey, D., & Collins, G. (2018). *Seven Trends: The Transformation of the Teaching Force*. Updated October 2018. CPRE Research Report# RR 2018-2. Consortium for Policy Research in Education; Weisberg, D., et al. (2009).
- <sup>51</sup> Swisher, A. (2022, July 28). Setting sights lower: States back away from elementary teacher licensure tests. National Council on Teacher Quality. <https://www.nctq.org/blog/Setting-sights-lower:-States-back-away-from-elementary-teacher-licensure-tests>; Peske, H. (2022, July 28). We wouldn't lower standards for pilot licenses—so why teachers? National Council on Teacher Quality. <https://www.nctq.org/blog/We-wouldnt-lower-standards-for-pilot-licenses-so-why-teachers>.
- <sup>52</sup> Sass, T. R., Hannaway, J., Xu, Z., Figlio, D. N., & Feng, L. (2012). Value added of teachers in high-poverty schools and lower poverty schools. *Journal of Urban Economics*, 72(2-3), 104-122; Steele, J. L., Pepper, M. J., Springer, M. G., & Lockwood, J. R. (2015); Goldhaber, D., Quince, V., & Theobald, R. (2018); Goldhaber, D., Lavery, L., & Theobald, R. (2015); Glazerman, S. & Max, J. (2011).
- <sup>53</sup>Levitan, S., Holston, S. & Walsh, K. (2022). *Ensuring Students' Equitable Access to Qualified and Effective Teachers*. Washington, D.C.: National Council on Teacher Quality. Retrieved from <https://www.nctq.org/publications/Ensuring-Students-Equitable-Access-to-Qualified-and-Effective-Teachers>.
- <sup>54</sup>Allensworth, E. & Hart, H. (2018, March); Kraft, M. A., Marinell, W. H., & Shen-Wei Yee, D. (2016).
- <sup>55</sup>While this evidence is not specific to principals, at least one study of teacher value added models found that pooling multiple years of data to estimate a teacher's value-added greatly reduced random measurement error: The Strategic Data Project. (2011). *Value-added measures: How and why the strategic data project uses them to study teacher effectiveness*. Center for Education Policy Research at Harvard University. Retrieved [https://hwpi.harvard.edu/files/sdp/files/sdp-va-memo\\_0.pdf](https://hwpi.harvard.edu/files/sdp/files/sdp-va-memo_0.pdf).
- <sup>56</sup> Pham, L. D., Nguyen, T. D., & Springer, M. G. (2021). Teacher Merit Pay: A Meta-Analysis. *American Educational Research Journal*, 58(3), 527–566. <https://doi.org/10.3102/0002831220905580>.
- <sup>57</sup> Putman, H., Ross, E., & Walsh, K. (2018).
- <sup>58</sup> Chi, O. L., (2021); Grissom, J. A., Bartanen, B., & Jones, A. A. (2019); Of note, the findings on racial matching are mixed, with at least one study finding racial bias in observations that was not significantly different when Black teachers were evaluated by a principal who shared their race and gender. See: Campbell, S. L. (2020). Ratings in black and white: a quantcrit examination of race and gender in teacher evaluation reform. *Race Ethnicity and Education*, 1-19.
- <sup>59</sup> Steinberg, M. P. & Sartain, L. (2021). The study determined that these differences were traced solely back to the student populations these teachers were more likely to teach: students from low-income families, students with higher reported incidences of behavioral issues, and students with lower literacy.
- <sup>60</sup> Campbell, S. L. & Ronfeldt, M. (2018). Observational evaluation of teachers: Measuring more than we bargained for?. *American Educational Research Journal*, 55(6), 1233-1267; Chi, O. L., (2021).
- <sup>61</sup> Gerber, N. (2021, April 22). Bias in teacher observations: No easy solutions. National Council on Teacher Quality. <https://www.nctq.org/blog/Bias-in-teacher-observations:-No-easy-solutions>.
- <sup>62</sup>White, T. (2014). *Adding eyes: The rise, rewards, and risks of multi-rater teacher observation systems*. Carnegie Foundation for the Advancement of Teaching; Cantrell, S., & ð, T. J. (2013); Whitehurst, G., Chingos, M., & Lindquist, K. (2015). Getting classroom observations right. *Education Next*, 15(1), pp. 63-68.
- <sup>63</sup> Campbell, S. L. (2020).

<sup>64</sup> District of Columbia Public Schools (2021, August). *IMPACT Data Trends: Equity and Mitigating Implicit Bias*. District of Columbia Public Schools. Retrieved from: [https://dcps.dc.gov/sites/default/files/dc/sites/dcps/page\\_content/attachments/EquityDisparate-Outcomes-memo\\_IMPACT-Review\\_August-2021.pdf](https://dcps.dc.gov/sites/default/files/dc/sites/dcps/page_content/attachments/EquityDisparate-Outcomes-memo_IMPACT-Review_August-2021.pdf).

<sup>65</sup> Dean, N. (2021). *2020-2021 Educator Evaluation Review: FY2022 Report to the Idaho State Board of Education*. Idaho State Board of Education. Retrieved from <https://boardofed.idaho.gov/resources/educator-evaluation-review-fy2022-report/>.

<sup>66</sup> Putman, H., Ross, E., & Walsh, K. (2018)

<sup>67</sup> Delaware Administrative Code, Regulation 108A

<sup>68</sup> Chen, B., Cowan, J., Goldhaber, D., & Theobald, R. (2019). *From the clinical experience to the classroom: Assessing the predictive validity of the Massachusetts candidate assessment of performance*. National Center for Analysis of Longitudinal Data in Education Research. American Institutes for Research. <https://caldercenter.org/publications/clinical-experience-classroom-assessing-predictive-validitymassachusetts-candidate>.