

Next-Generation Teacher Evaluation Reform: Taking on Teacher Quality from the Statehouse

Betheny Gross and Ashley Jochim
Center on Reinventing Public Education

November 2013

© 2013 Edvance Research, Inc.

The productivity of public education systems depends importantly on teacher quality, given that teachers account for a high share of education dollars¹ and are the single most important school-based driver of student learning.² Stanford economist Eric Hanushek estimates that replacing 6 to 10 percent of the worst teachers with just average teachers would measurably improve student achievement and help close the achievement gap for essentially no new dollars.³

States have turned toward evaluation as one of the primary levers for advancing teacher quality from the statehouse. As reported by the New Teacher Project, traditional evaluation systems suffer from a variety of design flaws, including infrequent, unfocused, and undifferentiated evaluations.⁴ Reforms have been developed to improve the quality of teacher evaluation measures and to better integrate evaluation data into decisions about compensation, tenure, and promotion. Since 2009, 37 states have adopted legislation that requires districts to incorporate estimates of teacher impact on student performance—often computed as value-added measures—into teacher evaluations, and to adopt more frequent, rigorous, and differentiated observation systems.⁵

While states have made great progress in these regards, they have also faced a variety of setbacks and new challenges. Some of the challenges are technical, such as ensuring quality data sources and training leaders and teachers to new norms for evaluation. Other challenges are more philosophical, such as how to incentivize and support what is essentially a schoolhouse activity—performance evaluation—from the statehouse. And there are still more challenges to come that states would do well to anticipate.

1. National Center on Education Statistics, *Revenues and Expenditures for Public Elementary and Secondary Education: School Year 2008–09 (Fiscal Year 2009)* (Washington, DC: U.S. Department of Education, 2011).

2. See, for example, Emily A. Hassel and Bryan C. Hassel, *3X for all: Extending the reach of education's best* (Chapel Hill, NC: Public Impact, 2009); Jonah E. Rockoff, "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data," *American Economic Review* 94, no. 2 (2004): 247–252.

3. Eric A. Hanushek, "The Economic Value of Higher Teacher Quality," *Economics of Education Review* 30 (2011): 466–479.

4. Daniel Weisberg, Susan Sexton, Jennifer Mulhern, and David Keeling, *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness*, 2nd ed. (New York: The New Teacher Project, 2009).

5. National Council on Teacher Quality, *State of the States 2012: Teacher Effectiveness Policies* (Washington, DC: National Council on Teacher Quality, 2012).

How this work moves forward depends on the answers to two questions: What are the big lessons from the design and implementation of the current crop of teacher evaluation systems? And what should a next-generation teacher evaluation system look like?

Interviews with several experts with diverse experiences in teacher evaluation offer a variety of lessons to those working from the statehouse to drive improvements in teacher quality. We interviewed the following experts:

- Dan Goldhaber, a university-based labor economist who directs the Center on Education Data and Research.
- Luke Kohlmoos, director of the Tennessee Department of Education's teacher evaluation program.
- Brad Jupp, a senior program advisor on teacher initiatives to U.S. Secretary of Education Arne Duncan.
- Christopher Thorn, director of the Carnegie Foundation's Advancing Teaching and Improving Learning program.
- Janice Poda, the strategic initiative director for the education workforce in the Council of Chief State School Officers.

Five Organizations Doing Work on Teacher Evaluation

- [The U.S. Department of Education](#) has been working with states to transform teacher evaluation systems since at least 2009, when the Obama administration's [Race to the Top](#) competitive grant program included [teacher evaluation](#) as a key area of reform. Later, in 2011, the department required applicants to the [Elementary and Secondary Education Act Flexibility](#) program to change state teacher evaluation systems to include the use of student growth measures. In 2013, the department allowed states to delay implementation of these required changes in response to challenges associated with putting the Common Core State Standards into place.
- [The Council on Chief State School Officers \(CCSSO\)](#) is an organization representing leaders of state education agencies (SEAs). Along with the National Governors Association, CCSSO was the chief architect of the Common Core State Standards Initiative. Since 2011, it has organized the [State Consortium on Educator Effectiveness](#), a collaborative of 28 states working to improve teacher quality. In 2011, it initiated the [Interstate Teacher Assessment and Support Consortium](#), which developed a new vision for teaching and strategies for improving teacher practice. In 2012, CCSSO convened a task force, [Transforming Educator Preparation and Entry into the Profession](#), that issued a call to action to chief state school officers and identified action steps that states could take to improve the quality of the education profession.

- **The [Carnegie Foundation](#)** is an independent policy and research center focused on improving teaching and learning and narrowing the divide between research and practice. Two projects shape the foundation’s work on teacher evaluation: The [Building a Teaching Effectiveness Network](#) brings together leaders in education practice, policy, and research to focus on developing and retaining effective teachers in the nation’s schools. The [Advancing Teaching–Improving Learning \(ATIL\)](#) program seeks to help those working in assessment and evaluation to learn from emerging practices in order to build more effective information systems to advance teacher quality. In 2013, ATIL released [A Human Capital Framework for a Stronger Teacher Workforce](#).
- **The [Center for Education Data and Research \(CEDR\)](#)** is a research center affiliated with the University of Washington, Bothell. CEDR addresses the disjunction between research, policy, and practice by conducting high-quality research and disseminating it to policymakers. The center’s research is concentrated in the areas of school and teacher effectiveness, educational accountability and governance, and teacher labor markets. Projects include an examination of [Denver Public Schools’ Professional Compensation System for Teachers](#), an evaluation of [Washington State’s new assessment of effective teaching](#), and an assessment of the [effectiveness of teacher preparation and recertification policies](#) in Washington State.
- **The [Tennessee Department of Education](#)** makes Tennessee a vanguard state in evaluation reform. Led by Education Commissioner Kevin Huffman, the state won the first round of Race to the Top funds, which it used to implement the state’s [First to the Top](#) plan. As part of this package of reforms, the state launched the [Tennessee Educator Acceleration Model](#), which relies on a combination of frequent observation, student growth data, and professional development to drive improvements in teacher quality.

WHAT’S NEXT FOR TEACHER EVALUATION?

The experts interviewed all agree that teacher evaluation is essential to any effort to improve teacher quality. Conversations reflected four shared themes on what state and other reform leaders should work toward next:

- Acknowledging that evaluation is just one part of a broader talent management system that improves teacher quality throughout the pipeline.
- Making teacher evaluation relevant to instructional practice.
- Managing implementation with an eye toward continuous improvement.
- Understanding that changes to standards, curricula, and instructional models could shake the foundations of evaluation systems.

Evaluation is Only One Piece of the Puzzle

Teacher evaluation generates data. It does not in itself improve teacher quality. As Christopher Thorn puts it, “Evaluation is the gateway drug to improvement, but it doesn’t deliver it.” The extent to which teacher evaluation is useful depends on how evaluation data are integrated into other aspects of the talent management system. Ultimately, as Dan Goldhaber explains, evaluation must fundamentally alter the incentives and behavior of teachers in the labor market.

Teacher evaluation data can be used to improve teacher quality at several points: upon entry, through its impact on teacher training and hiring; at midcareer, in shaping professional development and compensation; and at exit, through dismissal and tenure denial. Few states are currently leveraging teacher evaluation to approach talent management at all three of these points.⁶

The extent to which teacher evaluation is useful depends on how evaluation data are integrated into other aspects of the talent management system.

In most states, teacher evaluation debates have centered on using the data for decisions about tenure and dismissal. But, as Thorn explains, relying on exit pathways to improve teacher quality is very costly to the system and to students. Some economists estimate that it costs upward of \$15,000 to fire and replace a teacher.⁷

Moving forward, states will want to consider a comprehensive talent initiative that addresses the entire talent pipeline, from making the profession one that talented people want to enter to developing their skills to rewarding excellent performance. Tennessee is one state attempting to systemically improve teacher quality throughout the pipeline. As Luke Kohlmoos explains, the Tennessee Department of Education views evaluation reform as just one component of a broader effort to improve teacher quality. Anchored by the [First to the Top initiative](#),⁸ Tennessee is complementing evaluation reform

6. National Council on Teacher Quality, *State of the States 2012: Teacher Effectiveness Policies* (Washington, DC: National Council on Teacher Quality, 2012).

7. Anthony T. Milanowski and Allan R. Odden, “A New Approach to the Cost of Teacher Turnover,” *School Finance Redesign Working Paper 13* (Seattle, WA: Center on Reinventing Public Education, 2007).

8. “First to the Top,” Tennessee Department of Education, accessed October 21, 2013: www.tn.gov/firsttothetop/programs.html

by using evaluation data to change licensure rules, differentiate professional development, use high-performing teachers to conduct job-embedded professional development for their peers, incentivize reforms to teacher compensation practices in districts, and rate teacher preparation programs. Much of this work is still being formed, but the broader view—of talent reform as a systemic endeavor—is in place.

Make Evaluation Relevant to Instructional Practice

In early debates over implementation of evaluation reform, technical challenges related to measuring performance in untested subjects and accounting for nonclassroom inputs crowded out concerns over teaching and learning. The central implementation challenge confronting districts and states is how to make teacher evaluation matter for instructional practice and, ultimately, for children’s experiences in the classroom.

Two of the interviewed experts, Thorn and Jupp, emphasize that the success of new teacher evaluation systems will depend on their ability to act as a tool to improve teachers’ instructional practice, in ways summative evaluations alone cannot. Thorn explains that “improvement is problem-focused . . . [and] evaluation isn’t.”

Evaluation will affect teaching and learning only if the evaluation tools—most notably, observation of instructional practice—result in meaningful feedback to teachers that naturally guides their professional development. To Jupp, this means states must go beyond their initial investments in constructing sophisticated evaluation tools and toward building the internal capacity to supervise the work of schools and teachers that implement evaluation systems. Jupp is optimistic that with state support, instructional leaders will become skilled at wielding observation tools to provide differentiated supports to struggling teachers.

Thorn, however, cautions that the early evidence from vanguard states isn’t promising. He notes that in Tennessee, principals were overwhelmed by the number of observations they had to perform for each teacher. Moreover, the observation data ultimately revealed very limited variation in scores, calling into question whether the tools could be counted on to differentiate the quality of teachers or target job-embedded professional development.

Continuously Improve Teacher Evaluation Based on Feedback

Lessons emerging from Tennessee’s implementation suggest that teacher evaluation systems need further refinement in response to educator feedback and technical implementation challenges. Based on a statewide “listening campaign” launched by the State Collaborative on Reforming Education, a Tennessee nonprofit advocacy group, the state differentiated its observation system so that higher performing teachers received fewer observations. The state also changed the evaluation formula in nontested subjects, and is soliciting proposals for alternative evaluation metrics. This persistent

solicitation and incorporation of feedback into teacher evaluation policy has mediated some of the conflict over evaluation that has dominated in other states and districts—and that did so in Tennessee at first.

In addition to the political push and pull over teacher evaluation, significant technical challenges remain and suggest the need for further refinement. In Tennessee’s case, for example, misalignment between teachers’ value-added assessments and observation data led the state to institute stronger supports for observers.

Being receptive to feedback does not mean abandoning program goals. In Tennessee, the state continued to roll out the new evaluation system despite calls to delay the implementation of consequences for poor performance.

Future Reforms May Disrupt Evaluation Work

Today’s teacher evaluation systems are based on a particular model of what a school looks like, including content of instruction, progression of students, and composition of instructional staff. Poda, Jupp, and Goldhaber all warn that this vision of education will change, probably dramatically. The longevity of evaluation reforms will depend on whether the evaluation systems can be adapted to accommodate both the curriculum changes that are already taking hold across the country and the deep technological changes that are soon to come.

In the immediate future, while curriculum standards are changing across the country, Poda and Jupp emphasize that conversations about academics cannot be separate from conversations about evaluation reform. The connection between curriculum and evaluation, Jupp explains, was an aha moment for officials in the U.S. Department of Education, who realized that rolling out curriculum reform in one office and evaluation reform in another creates “massive confusion” for educators.

Poda sees the result of this disconnect already playing out in states adopting the Common Core. She notes that there is considerable dissonance between what teacher evaluators typically look for in classrooms and what is needed for a classroom to succeed under the Common Core. States will need to look closely at their data to identify what matters for students’ success given the new standards, and improve their evaluation rubrics and professional development accordingly.

Looking further down the line, Goldhaber sees technology profoundly changing the work of teachers and the organization of schools. “The only way teaching is going to be radically more productive,” he says, “is if schools leverage technology and use big data to determine the match between a person and their best job—their talent and job fit. For example, maybe there is someone who is not a good teacher face-to-face but is phenomenal at facilitating online discussion or debate.”

Already, blended-learning schools show what is possible. In the [Carpe Diem Schools](#) and [Rocketship Education](#) charter networks, teams of teachers oversee the independent work of several classrooms' worth of students, pulling students into small-group instruction to target specific remediation or acceleration needs. Other blended schools maintain the familiar classroom structure and size, but basic content delivery (what was traditionally provided through lectures) and routine practice exercises are shifted online, freeing the classroom teacher to spend more time diagnosing and addressing the unique needs of students. Other models for leveraging technology in schools will continue to emerge.

“Of course, changing the nature of teachers’ work like this raises a whole host of issues about evaluation, because it would have to be very different,” Goldhaber says. Evaluation systems will need to acknowledge that different people are successful in different contexts, and account for a variety of career paths for teachers. And, importantly, evaluation systems will need to account for a range of instructional contexts. In a school where a given student learns math from a variety of teachers, paraprofessionals, and online programs, measuring student success as a value-added measure attached to one teacher, as is typically done in today’s new evaluation systems, would not be plausible. Likewise, observation rubrics that were built around the idea of one teacher instructing a group of 25 or 30 students may no longer make sense in a more fluid instructional environment. “An effective evaluation could not be standardized in that world,” Goldhaber says. “You would need nuanced judgment.”

LOOKING FORWARD: WHAT STATES MUST DO

Discussion with these thought leaders reveals several recommendations for states looking toward the next generation of teacher evaluation reform.

First, states must learn to view evaluation as one component of a broader talent system, and must work to fully utilize and integrate all their points of leverage to improve teacher quality. This includes taking the following steps:

- Integrate evaluation data into other parts of the talent pipeline—for example, when rating teacher preparation programs, providing professional development, and establishing compensation schemes.
- Use multiple levers to improve teacher quality at each point in the pipeline (entry, development, and exit).
- Rethink how the SEA staffs its professional development and teacher licensure offices so that they work seamlessly with those responsible for teacher evaluation.
- Consider using carrots and sticks to change weakly rated educator preparation programs.
- Streamline the number and type of licenses and certification pathways to reduce fragmentation in teacher preparation programs.

Second, it is clear that evaluation tools will need to do more to impact classroom teaching. There are a few steps states can take to help move this process along:

- Provide school instructional leaders with training and support to perform high-quality evaluation and leverage evaluations for talent improvement.
- Ensure agency staff understand the work of teacher supervision; states should not rely on measurement models alone.
- Ensure district and school-level personnel have some flexibility to differentiate observation demands based on teacher need; a high-performing teacher may not need to be observed as often as a struggling one.
- Use high-performing teachers to provide additional observations and conduct job-embedded professional development.
- Support districts to build additional formative assessment tools that can be used to more frequently assess student progress and support teacher development.

Third, emerging implementation challenges and the likelihood of new ones on the horizon suggest a need to formalize a process for continuous improvement. This might include the following:

- A process to seek feedback from educators, advocacy groups, and others.
- Careful monitoring of data to identify and target assistance to schools and districts showing large disparities between observation and value-added data.

Finally, to enable evaluation systems to adapt to the inevitable fundamental shifts in instruction and the work of teachers, states should work to do the following:

- Integrate teacher evaluation into the broader discussion of school reform.
- Break down organizational barriers between teacher evaluation and other dimensions of teaching and learning.
- Provide districts and school leaders with the flexibility to use evaluation in different school contexts.

These recommendations will ensure that states leverage teacher evaluation as a tool for driving teacher quality in both the schools of today and those of tomorrow.