The State of Teacher Evaluation Reform

State Education Agency Capacity and the Implementation of New Teacher-Evaluation Systems

Patrick McGuinn   November 2012

Center for American Progress

The State of Teacher Evaluation Reform

State Education Agency Capacity and the Implementation of New Teacher-Evaluation Systems

Patrick McGuinn   November 2012

Contents

Introduction and summary

The Obama administration's Race to the Top competitive grant program initiated an unprecedented wave of state teacher-evaluation reform across the country.[1] To date, most of the scholarly analysis of this activity has focused on the design of the evaluation instruments[2] or the implementation of the new evaluations by districts and schools.[3] But little research has explored how *states* are managing and supporting the implementation of these reforms. As U.S. Department of Education Secretary Arne Duncan has remarked: "…because teacher evaluation systems are still a work in progress, it is vital that school leaders and administrators continue to solicit feedback, learn from their mistakes, and make improvements."[4] It has become increasingly clear that the role of state education agencies will be critical as school districts enter what for most will be uncharted territory. As Edward Crowe argued in his recent Center for American Progress report on teacher preparation, "The capacity and commitment of states to implement these Race to the Top activities will determine success or failure."[5] And as highlighted in recent news reports, many states are struggling to implement their new teacher-evaluation systems and most of the Race to the Top winners have asked to extend their timetables for completing this work.[6]

This paper offers an assessment of how early adopter states' departments of education have undertaken the preparation and implementation of new evaluation systems. It also identifies challenges and lessons that can be used to guide future reform efforts in this area. Developing new teacher-evaluation systems has been identified by scholars and policymakers alike as a crucial part of improving teacher quality and raising student academic performance across the country.[7] It is imperative that we learn more about the most effective way for state education agencies to support districts in this difficult work.

This assessment of the activities of state departments of education is based on comparative case studies of six states: Colorado, Delaware, New Jersey, Pennsylvania, Rhode Island, and Tennessee. These particular states were selected because they are "early adopters" in the area of teacher-evaluation reform and

because their states and/or education agencies have undertaken different approaches to implementing the reforms. Two of the states—Tennessee and Delaware—were initial Race to the Top winners, while the other states won smaller grants in later rounds. Research consisted of a review of the scholarly and think tank research on state education agency capacity and teacher-evaluation systems; analysis of reports and data from the state education departments' websites and from organizations such as the Council of Chief State School Officers; a study of media coverage of the reform efforts in the six states; and 15 interviews with national experts on teacher-evaluation reforms and state education agency and local education agency staff in each state.

The central questions probed and answered in this report include:

- How are state education departments adjusting to their new, more ambitious roles and responsibilities in the wake of Race to the Top?
- What steps are state education agencies taking to restructure themselves for these new responsibilities?
- What kinds of capacity—financial, personnel, technical—have state education agencies added to support the implementation of new teacher-evaluation systems, and what kinds of capacity are still lacking?
- To what extent and in what ways are state education agencies relying on external capacity by contracting outside consultants to provide technical assistance with this work?
- What is the role of philanthropic organizations in supporting state education agencies in this work?
- How rapidly and effectively are states implementing their new teacher-evaluation systems?
- How are states approaching this work differently from one another—do some approaches appear to be more or less effective than others?
- What challenges are emerging and how are states addressing them?
- What lessons can be learned from these early-adopter states that can inform teacher-evaluation reform in the rest of the country?

It is clear that state education agencies are working hard to realign their organizations with the many new responsibilities that have been thrust upon them in the wake of the federal No Child Left Behind Act and Race to the Top programs.[8] State efforts to implement new teacher-evaluation reforms offer excellent examples of the ways that state education agencies are adapting to their new role as well as the ways in which ongoing capacity gaps continue to impede their work.

Improving teacher quality has become the centerpiece of the Obama administration's education agenda and of the contemporary school-reform movement. The many challenges that have already emerged, however, also highlight how difficult this work is and how it is complicated by short timelines and limited state education agency staffing and funding. A number of key challenges to implementing new teacher-evaluation systems have emerged from the work of the early-adopter states. Some of these challenges, which can inform the efforts of other states going forward, include:

- **The philosophical/statutory/constitutional debate over the proper role of state education agencies.** It is important to recognize that all state education agencies are not the same—each agency has a unique history and operates in a different fiscal, political, statutory, and constitutional context. In particular, states vary significantly in their attachment to local control of schools and the proper role of the state in education. This has a major impact on how state education agencies approach teacher-evaluation reform. A related issue revolves around the traditional focus of state education agencies on compliance and accountability activities, which has made local education agencies wary of being candid about whether and how they might be struggling to implement reform and made them reluctant to seek out assistance.

- **The amount of flexibility in state evaluation systems varies greatly.** States vary widely in the amount of centralization and standardization they have mandated—through statute or regulation—in the new teacher-evaluation systems. This variance has a major impact on the state education agency's approach to supporting implementation. A clear tension is emerging between a state's desire to give districts flexibility to select or adapt evaluation instruments that are best suited to their particular circumstances, and the state education agency's limited capacity to provide implementation support for a wide array of instruments.

- **State education agency restructuring and the human capital demands.** State education agencies in many states are undergoing a radical restructuring and re-staffing as they embrace a shift from being compliance monitors to service delivery/school-improvement organizations. This restructuring is difficult and time-consuming work and, while necessary to carry out new responsibilities over the long term, creates a number of short-term challenges. It will take some time for this organizational shake out to be completed and for new structures and staff to acclimate to their new roles. Many state education agencies have created new teacher-effectiveness units, but the degree to which these units have been well-integrated with other units appears to vary and longstanding concerns about agency siloing persist.

- **Internal versus external capacity.** In the short term, state education agencies are dealing with their internal capacity gaps by relying on two different kinds of external capacity: outside consultants and foundations. There is some concern, however, that reliance on outside grants and consultants may preclude or delay the development of the fiscal self-sufficiency and internal capacity that can support these systems over the long term.

- **Funding streams and the "fiscal cliff."** There is a great deal of concern about state education agencies' lack of capacity to implement these reforms, particularly for states that did not win a Race to the Top grant or secure foundation support (which is the majority of states). Given the current tight fiscal climate, most states have been unable or unwilling to allocate new money to support the implementation of these reforms. State education agencies appear to vary widely in the way that they have spent external funds, the degree to which they are dependent on them, and the extent to which they have begun to bring these expenses on budget. As a result the eventual end of federal and foundation grants—part of the upcoming "fiscal cliff"—is likely to affect states in different ways.

- **Evaluating the evaluators.** One of the primary activities of state education agencies in supporting their local education agencies with teacher-evaluation reform has been providing training to the administrators that will be conducting the new observations. States vary widely in their approach here, however, for both philosophical and capacity reasons with some state education agencies (such as Tennessee) directly training all evaluators, some (such as Colorado and Pennsylvania) adopting a train-the-trainer model, and others (such as New Jersey) leaving the training entirely up to districts.

- **Implementation timetables and sequencing.** Most state reform statutes have established rapid timetables for the installation of new teacher-evaluation systems. While all states are struggling to meet these timetables, it is becoming clear that some states are struggling more than others due to the fact that states vary in terms of their experience with statewide evaluation systems. A related challenge centers on the extent to which evaluation reforms are—or are not—being connected to the implementation of other reforms such as new principal evaluations and new common core standards and assessments.

- **Value-added/growth scores for teachers in nontested subjects.** Perhaps the single biggest challenge in implementing new evaluation systems that has emerged from the field is the fact that the majority of teachers do not teach in tested subjects

or grades and as a consequence standardized student achievement data is not available to be used in their ratings. Districts are working independently to develop their own student-learning objectives, but the quality of the results appears to be mixed and messy both within and across states. This is an enormous problem and it is clear that many state education agencies are struggling to address it.

- **Networks, policy learning, and politics.** Policy learning and continuous improvement requires that local education agencies, state education agencies, and the U.S. Department of Education be transparent and forthcoming about what is working and what is not and that lessons learned be regularly shared within and between states. But on the ground the reality appears to be that not enough communication and sharing of information about effective measures is happening yet. Balancing their support and compliance monitoring functions will continue to require a delicate balancing act for state education agencies and the Department of Education, but getting the balance—and the communication—right will be crucial to the evaluation reform effort going forward.

The lessons derived from these challenges form the basis for the following recommendations:

- Individual states need to think carefully about the work that needs to be done to implement a new teacher-evaluation system, assess the existing capacity at the local and state education agency levels, and define an appropriate role for the state education agency that is commensurate with state constitutional and statutory provisions.

- Given their limited resources, state education agency leaders have to think carefully about how best to reallocate existing staff and budgets to focus on new responsibilities, build capacity, and eventually bring work that is funded by external grants on budget. Federal regulations and state budgeting and civil service requirements that constrain the ability of state education agencies to do so should be revised with an eye toward permitting greater managerial flexibility.

- State education agencies need to think about comparative advantage and economies of scale—where the state can provide something districts cannot. Providing technical assistance and policy interpretation, creating communication networks for information sharing, expanding assessment portfolios, and establishing online training modules are several areas where state education agencies could add real value.

- State legislatures and state education agencies should tailor their implementation timelines to the unique needs and resources of their particular state. They should also determine how the evaluation work ought to be sequenced with and connected to the roll out of other related education reforms, particularly those reforms around teacher preparation, professional development, principal evaluation, and common standards and assessments.

- States need to think long term about how to produce a large and stable supply of administrators—state education agency staff as well as school principals and district superintendents—with the training, technical expertise, and field experience to address their current human-capital challenges around teacher-evaluation reform. Partnering with a state's higher education system or with management consultants to devise new training and certification programs that reflect the different work and skill set required is crucial.

- The learning curve for local education agencies, state education agencies, and the U.S. Department of Education during the implementation of new teacher-evaluation systems will be steep and mistakes will inevitably be made, but it is crucial that the work be transparent and that information about effective methods be shared up and down the education delivery chain. State education agencies and the Department of Education need to create a safe space where practitioners within and across states can be candid about the mistakes they are making and the support they need without fear of triggering punitive oversight or interventions by a higher authority.

The remainder of the paper will provide a review of previous research on state education agency capacity and teacher-evaluation reform, analyze state education agency implementation efforts in the six case study states, and elaborate on the lessons and challenges that have emerged from the early-adopter states.

Implementation and state capacity gaps

Race to the Top inaugurated an unprecedented wave of state teacher-evaluation reforms.[9] The National Council on Teacher Quality reports that 36 states and the District of Columbia have changed their teacher-evaluation policies since 2009. There has been a large increase in the number of states that require annual teacher evaluations (currently 43 states), and those incorporating student achievement (32 states), differentiated levels of performance (26 states), annual classroom observations (39 states), multiple observations each year (22 states), and performance-based tenure decisions (9 states).[10] The National Council of Teacher Quality found that "the landscape is quickly and dramatically changing when it comes to rethinking and rebuilding teacher evaluations in school systems in the United States. There is a great deal of promise and potential in these policy trends. At the same time, however, it is clear that policy is only part of what is necessary… Even the best evaluation system can be implemented poorly or undermined."[11]

It is extraordinarily difficult to drive change from the state capitol all the way down to the classroom level. For teacher-evaluation reform to succeed, state policy changes must result in changes in district practice. In turn, changes in district practice must change the behavior of principals and teachers at the school level, and changes at the school level must deliver improved student performance. As a result the vigor and effectiveness of state and district implementation efforts will be critical. But commitment alone may not be sufficient, for, as Harvard's Richard Elmore argues, states suffer from a "capacity gap" that undermines their ability to monitor and enforce mandates and provide technical guidance.[12] A 2011 Center for Education Policy survey, for example, found that operating budgets for a majority of state education agencies have declined by 10 percent or more since 2007, and that "only a handful of states believe they have all three elements in place—adequate expertise, staffing levels, and funding—to carry out key American Recovery and Reinvestment Act related reform activities."[13]

Paul Pastorek, former Louisiana state superintendent of education, has expressed concern that the U.S. Department of Education and many states have been insuf-

ficiently attuned to these capacity deficits. "I think some [states] may be underes-timating the resources and energy that these kinds of initiatives require . . . state departments of education are not designed to implement these programs," says Pastorek.[14] Furthermore, a recent study by the Data Quality Campaign found that state data systems are woefully inadequate, pointing out that only 11 states nationwide (and only 4 of the 12 Race to the Top winners) have all the compo-nents they deem essential.[15] And many states and districts have little experience implementing some of the reform approaches contained in their Race to the Top applications. A 2011 study by the U.S. Government Accountability Office con-cluded that states are struggling to implement the reforms in their Race to the Top applications and that their "overly optimistic" timelines are unlikely to be met.[16] The dozen winners from the competition have formally amended their Race to the Top plans more than 25 times, usually to scale back proposed reforms or push back timetables.[17] District efforts to circumvent compliance with state mandates are a further challenge, as are the debates over memoranda of understanding and implementing the Obama administration's new school-restructuring approaches.[18]

Adam Tucker, senior program officer at the Bill & Melinda Gates Foundation, observes, "The work is moving to another phase. Now that the challenges of adopt-ing the evaluation reforms have been done, the next phase of work is design and implementation. This is very challenging work to implement at the school level where the rubber meets the road." As noted by TNTP (originally The New Teacher Project), "Now comes the hard part. As states across the country have already learned, strong implementation will determine whether a new evaluation system lives up to its potential. Even the most elegantly designed evaluation system won't succeed unless schools implement it consistently and accurately."[19] TNTP identified five sets of investments that states should make to support schools with implementa-tion—tools and systems (including a teacher value added model, student learning measures, assessment rubrics, and data systems); training for evaluators and district staff; communications with stakeholders; monitoring and support (including support teams, metrics of success, and evaluator accountability); and sustainabil-ity (through feedback and improvement and the reallocation of state and district resources to support implementation over the long haul).

While we are beginning to understand the kinds of capacities that state educa-tion agencies will need to implement teacher evaluation reforms effectively, we have much less understanding about whether state education agencies possess these capacities and how they are actually going about this work in the field. In a recent survey of state education agencies, Cynthia Brown of the Center for

American Progress and her colleagues note that a wave of recent reforms has "put immense stress on agencies that were originally conceived as tiny departments primarily designed to funnel money to local school districts. Yet it is not at all clear that state education agencies are prepared for this demanding new role."[20] Sara Mead, Andrew Rotherham, and Rachael Brown of Bellwether Education Partners recently cautioned that policymakers must be careful to avoid a "policy hangover" and that "if advocates of 2.0 teacher evaluation rush too quickly to create new systems or do so without appropriate humility about what we do and do not know… the nation's teacher evaluation spree could turn into a big headache."[21] Let's turn now to the state case studies and analysis, which will hopefully provide some insights—some ideational aspirin, if you will—that can help prevent the onset of the potential teacher-evaluation "hangover" or at the very least, mitigate its effects.

Six state case studies

Quick facts

Number of teachers: 66,558
Number of schools: 1,803
Number of districts: 137

How Tennessee's teacher evaluations are structured:
Tennessee has a statewide teacher evaluation model that local districts are required by law to adopt. The system uses three components to arrive at a teacher's level of effectiveness: observation data (50 percent); student growth (35 percent); and student-achievement data selected by the educator and his/her supervisor from a list of state approved options (15 percent). Teachers with three or less years in the classroom are observed six times per year, while more experienced teachers are observed four times per year. The evaluation system was implemented statewide in the 2011-2012 school year.

Tennessee was a first-round Race to the Top winner—it received a $500 million grant—and had a pre-existing statewide teacher-evaluation system in place prior to Race to the Top. It was one of the earliest states in the country to initiate teacher-evaluation reform and its value-added student-growth system, the Tennessee Value-Added Assessment System enacted in 1992, informed much of the subsequent debate across the country. Sara Heyburn, assistant commissioner for teachers and leaders, points out, however, that the old evaluation contained infrequent and subjective observations, and that the fidelity of implementation was completely on the district. As a consequence, results were not collected by the state and there was limited accountability. Tennessee's recent reforms established a statewide teacher-

evaluation model that local districts are required by law (First to the Top Act of January 2010) to adopt. A waiver option is available for the observation instrument but 126 of 137 districts are currently using the state model. As Heyburn notes, "The state has a lot of centralized decision-making—there is a significant state role in terms of design, development and implementation around evaluation."

Tennessee's evaluation system has three components: observation data (50 percent), student growth (35 percent), and student-achievement data selected by the educator and his/her supervisor from a list of state board approved options (15 percent). Apprentice teachers (those with three years or less in the classroom) are required to be observed six times per year, while professional teachers are observed four times per year. During the 2010-11 school year the teacher-evaluation advisory committee piloted four different evaluation methods in approximately 30 districts and recommended the TAP (system for teacher and student achievement) teaching standards[22] as the model for the observation component of the evaluation. Most districts have adopted that model, which is the only one that Tennessee's department of education is providing implementation support for. The evaluation system was implemented statewide in 2011-2012. *Note: The Memphis school district, which is part of the Gates Foundation's Intensive Partnerships for Effective Teaching grant, and a few other districts have devised their own models.*

The state's department of education worked with TNTP, formerly known as The New Teacher Project, to survey school districts and identify capacity needs around teacher evaluation. Unlike many other states that are relying on a "train-the-trainer" model, the Tennessee department of education trained all 5,000 plus observers and evaluators itself, through a four-day summer training session in groups of 40 to 50 (a total of 102 cohorts.) The department contracted with the National Institute for Excellence in Teaching to deliver the training and also provide ongoing and on-demand support to schools during the academic year. Timothy Gaddis, the assistant superintendent for teaching, learning, and assessment for the Williamson County Schools (and the former director of evaluation for the Tennessee department of education), calls the institute's training and support "really, really strong." The state created an online training portal with videos and other instructional resources, along with an online certification test (that measures inter-rater reliability) that all observers must complete. Ninety-seven percent of principals passed the test, though many apparently had to take it multiple times before passing—a potential red flag.[23]

Assessing the new evaluation system, Gaddis remarks that he wishes he had "one more year or so to roll out" the new system and notes that in order to meet Race to the Top timelines "we had to roll it out with a lot of bugs in the system." Gaddis recalls that he had to say "'I don't know' a huge amount of the time and this ate away at our credibility."

Assistant Commissioner Heyburn terms the first year of the state's new evaluation system a success, saying, "The big win from our standpoint is that we have implemented a multi-measure evaluation system in all districts and all schools across the state that provides regular feedback to teachers and focuses them on instructional practice and student results in a very real way. The conversation change that we have seen at the school and district level over the past year has been really astounding." Heyburn admits that there were many challenges associated with initial implementation, but that ultimately Tennessee was able to fully implement the new system in the field in the 2011-2012 school year.

The implementation of Tennessee's new system encountered some political turbulence the result of media coverage that highlighted serious concerns raised by state legislators early on in the process as well as an effort to roll back some elements of the new evaluation system.[24] This underscores how state education agencies must be cognizant of both the policy and political challenges involved in implementing and sustaining difficult and controversial reforms.

After winning its Race to the Top grant, Tennessee contracted with the U.S. Education Delivery Institute to conduct a "capacity review" of the state's department of education. Their review concluded that "the organization and the work wasn't organized in a way that supported implementation … [and] reinforced that intentional change had to happen in order to improve capacity, regardless of how that would affect components, departments, and people in the agency."[25]

In April 2012 Kevin Huffman came on board as Tennessee's new education commissioner and reorganized the state education agency. A new division of teachers and leaders was created within the state's department of education to bring together all of the different elements of the human-capital continuum—from educator preparation, licensure and certification to recruitment, staffing, and compensation to evaluation and professional development—which in the past had all been functions of different offices that were not connected and didn't necessarily sync with one another. Commenting on the new division, Heyburn says she believes that "it

Unlike many other states that are relying on a "train-the-trainer" model, the Tennessee department of education trained all 5,000 plus observers and evaluators itself.

has really helped us to have the state level human capital management work under the same division to ensure that we have the same vision for teaching and leading across the spectrum of an educator's career. To get this right, we have to work in tandem across these offices." Tennessee—like other states—has struggled to find the appropriate staff to fill some positions in this new division. Heyburn went on to note, "These new evaluation models require unique skills and expertise but there are not a lot of people yet with experience in this work."

Heyburn reports that while she has had a good bit of freedom to recruit and hire the people she needs, several changes of hand around the evaluation work during the first phase of implementation have been a challenge.

In addition to getting the department capacity right, Tennessee has used Race to the Top money to hire consultants to support some time-limited initial work, including year-one training and the state data system evaluation. Gaddis expresses concern, however, about staff turnover at the state education agency during such a critical time. "It is tough to not have that consistency at a time of great change and uncertainty—the new folks have a lot of catching up to do before they can be effective." He also says it is "critical that the folks who are driving and are the face of big change really have credibility at the district level" but notes that the hiring of "nontraditional" state education agency staff and out of state consultants has created a "credibility problem." Heyburn has a slightly different assessment, saying she believes that hiring new staff and drawing on select external support, while not perfect, has enabled the state department of education to build the internal capacity needed to support evaluation reform while avoiding funding new positions with Race to the Top money in a way that might not be sustainable.

Gov. Bill Haslam (R) asked a nonprofit organization, the State Collaborative on Reforming Education, or SCORE, to provide an independent assessment of the first-year rollout of the new system. The nonprofit led a five-month listening and feedback process that included nine regional roundtables and an online teacher questionnaire. Their report concluded that overall, "Tennessee's evaluation system is improving both the quality of instruction in the classroom as well as the establishment of accountability for student results."[26] They found, however, that educators questioned whether principals had the time and ability to effectively assess teachers and believed that there was "inconsistent interpretation and implementation of the rubric." The report found that two-thirds of teachers do not have individual value-added student growth data for their grades and subjects, while also encouraging the state to develop better metrics for teachers in nontested subjects. The report

also criticized the lack of high-quality professional learning opportunities tied to teachers' performance feedback and the lack of attention to linking the new evaluation system to the pending implementation of the common core state standards. Assistant Superintendent Gaddis cautions that there have been "serious growing pains" with the implementation of the new system along with some "fundamental unfairness" related to the requirement that every teacher's evaluation be 35 percent based upon a growth score. He notes that "since 60 percent of our teachers have no standardized test, the state had to put measures in place that really aren't a good direct indicator of growth in the area of the teacher's responsibility."

In response to a request from the state legislature, the Tennessee department of education also conducted its own internal analysis of the first-year implementation, which it released in July 2012.[27] It conducted more than 120 stakeholder meetings across the state, met with all of the state's 136 directors of schools, and compared the data from the teacher observations and value-added scores to student test scores. The review was conducted by the Tennessee Consortium on Research, Evaluation, and Development, which is funded by the state's Race to the Top grant. Overall, the report concluded that evaluators successfully identified high-performing teachers but "systematically failed to identify the lowest-performing teachers, leaving these teachers without access to meaningful professional development and their students and parents without a reasonable expectation of improved instruction in the future."[28] Assistant Commissioner Heyburn notes that the biggest challenges were at the school level and not the district level and points out that 72 schools—across 50-plus districts—have been targeted for additional support in year two of the program.

According to the teacher surveys "it was clear that there were communication challenges" and that teachers and school leaders "struggled to find the support and guidance needed to navigate the early stages of implementation." The state education department pledged to develop a "more centralized communication strategy" to address this issue going forward.[29] Heyburn acknowledged that "communication is key" but it is something that state education agencies as a rule don't do well, particularly in terms of working with districts. As she explained, "We had to reset our [communications] strategy early on in terms of our ability to respond to districts so that they could get answers to their questions about implementation in real time." Heyburn says staffing an online question and answer hotline and holding regional technical assistance meetings helped accomplish this strategy. In addition, the state education department announced plans to hire five full- and part-time evaluation coaches, in collaboration with the National Institute for Excellence in Teaching, who are being deployed regionally to support

district implementation efforts during the 2012-13 school year. These coaches will devote special attention to those schools where year-one scores were most out of alignment. The department also intends to expand the online resources available through the TAP portal to include more model lessons and other resources that tie directly to rubric indicators. Recognizing the crucial but time-consuming role that principals play as evaluators in this new system, the state department of education is also working with districts to help them reallocate managerial tasks in order to free up more time for them to work as instructional leaders.
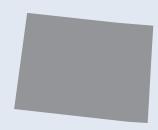
While the various reports generated a great deal of concern and criticism of Tennessee's evaluation reforms, they also demonstrated the importance of transparency in the implementation of teacher-evaluation reforms. Heyburn notes the "need to reflect on year one and make some changes and continue the cycle of improvement of this system and our support at the state level."

As in other states, the state education agency in Tennessee is struggling to define an appropriate and constitutional role in implementing teacher evaluation reform and to balance being supportive and flexible around district efforts to ensure compliance. After the SCORE report was released, newspaper coverage reported "the state is looking for clarity regarding when officials can intervene in districts with a wide gap between value-added and observation scores."[30] Given Tennessee's long experience with value-added data and the fact that its effort was relatively well-funded and supported, the challenges that have emerged there should serve to underscore the difficulty of this work. Finally, Heyburn emphasizes that it is "important to frame this from the beginning as continuous improvement of the evaluation system, just as we are talking about continuous improvement of teaching practice. We are in the early stages of learning about how to build effective evaluation systems and learning more and more all the time."

Colorado

<div>

Quick facts

Number of teachers: 48,543
Number of schools: 1,835
Number of districts: 178

How Colorado's teacher evaluations are structured:
Senate Bill 10-191 directs school districts to adopt new teacher-evaluation systems that are based 50 percent on student academic growth and 50 percent on observations and/or other methods that measure professional practice. The system incorporates four performance level ratings for educators. Local school districts can adopt the state model wholly or in part, but district evaluations must meet or exceed the state criteria and be subject to state review. The new evaluation system is to go into effect statewide in the 2013-2014 school year following a two-year pilot.

</div>

In 2010 Colorado enacted one of the most sweeping educator evaluation reform statutes in the country (Senate Bill 10-191), directing school districts to adopt new teacher-evaluation systems based on two major components: 50 percent on student academic growth and 50 percent on observations and/or other methods that measure professional practice. Beginning in 2011 and continuing through to the start of the 2013 school year, Colorado is piloting elements of its new evaluation system that will go into effect statewide in 2013-14. Katy Anthes, the executive director of educator effectiveness in the Colorado Department of Education, emphasizes the importance of the two-year pilot program versus the single-year pilots being used elsewhere, pointing out that it allows them more time to learn and refine from mistakes. And unlike many states, Colorado's legislature has been actively involved in the rollout of the new evaluation system, including being required by statute to approve plans developed by the state's department of education related to its teacher-evaluation system. Observers also report that State Sen. Michael Johnston—the sponsor of the reform bill—has been actively involved in the implementation of the pilot.

In contrast to Tennessee's more centralized approach, Colorado's emphasis on local control has led to an optional state model for teacher evaluation that districts can choose to adopt wholly or in part, or they can develop their own evaluation

systems. District-created evaluation systems must meet or exceed the state criteria and must be submitted to the state's department of education as part of the annual assurances, and are subject to state review as needed. The Colorado department of education, however, only offers training and support for the state model, incentivizing districts to adopt it. Going forward it will be interesting to see how much variation—if any—exists in the various evaluation systems. Another area to watch will be how frequently and in what way the state uses its audit authority.

Another unique feature of Colorado's approach to this work is that it decided to implement a new principal-evaluation system, also as a pilot, before embarking on the pilot phase of the new teacher-evaluation system, a move that Anthes believes has really helped the latter. Both new evaluation systems will be implemented statewide at the same time.

Prior to the passage of Senate Bill 191, the Colorado Department of Education did not have an educator effectiveness unit and though it has expanded a bit, it remains small. Due to staff and resource constraints, the state's department of education has adopted a "train-the-trainer model." Anthes notes, "We know we can't be everywhere and do this work with only eight trainers so we are partnering with our regional service delivery agencies and district personnel to go out and train people at the school level." She also says that the state's department of education plans to build long-term capacity by creating criteria and an approval process that will allow programs and districts to become "state approved training programs," where staff from professional associations, boards of cooperative educational services, school districts, and institutions of higher education can assist with the work. The training—one and a half days long—provides an orientation to the law and a "deep dive" into the rubric—how it should be used and scored—followed by simulated practice evaluations and feedback sessions. The Colorado Department of Education has also devoted one staffer to communication and outreach along with creating online modules that walk districts through the implementation process. In order to ensure compliance, it requires that districts report every July (the assurances noted above) detailing the type of evaluation system that each district is using and how districts are meeting state standards. One way the state's department of education will monitor district progress is by looking at the evaluation data to determine if there is a meaningful correlation of educator ratings and student achievement. Anthes admitted to "capacity issues" at both the state education agency and district level, saying doing this work can be "exhausting."

Nina Lopez, vice president at the Colorado Legacy Foundation, a private nonprofit organization supporting education reform, emphasizes that local context

matters for implementing these kinds of reforms. An interesting dynamic in Colorado, for example, has been figuring out an acceptable—and constitutional—role for the state education agency. Colorado is a local control state and the state constitution explicitly declares school curriculum to be the sole province of local districts and places limits on what the state education agency may require. Lopez observes "there were decades where districts just told the state to 'leave us alone.'" But after some initial resistance to the state's evaluation reforms, Lopez says districts are now requesting support from the state education agency, a situation she calls "unprecedented." Many districts have come to recognize that they do not have the capacity to do this work on their own and are concerned about the potential legal challenges that would emerge if personnel decisions are made by an evaluation system that is not valid and reliable. Colorado contains a large number of small, rural, and geographically isolated school districts that according to Lopez "have limited resources available to create a valid and reliable evaluation instrument." These small districts' emphasis on personal relationships between principals and teachers and their difficulty in recruiting qualified new teachers and principals further complicates attempts to create rigorous evaluation systems.

Lopez notes that there are "huge challenges in doing this work" and that "no one could have fully contemplated all of the details, all of the steps involved and what would be required because no one had ever really put all of these pieces together before at the state level." In particular, she says that the "importance of data and analytics and assessments" can't be overstated: "The magnitude and primacy of doing solid analysis of the data both for the evaluations and the feedback for teachers is really, really going to need to work well for this thing to realize its fullest potential and the capacity across states and our state is highly variable at best." There is a wide variety of labor-intensive and time-consuming tasks that are necessary to facilitate the roll out of these new systems, including the creation of common course codes, teacher-student data links, roster verification, data analysis templates, electronic version of observation rubric, and professional development resources connected to the rubric. The state education department's initial estimate was that only three staff would be necessary to implement the new teacher-evaluation reforms but Lopez says it has now become the "core focus" of the agency. She reports that the American Institutes for Research was hired to provide technical expertise to the department.

One of the most unique—and successful—aspects of Colorado's approach to educator-evaluation reform has been the partnership between the state's department of education and the Colorado Legacy Foundation, a private, nonprofit foundation created in 2008 with philanthropic support from a variety of sources,

Another unique feature of Colorado's approach to this work is that it decided to implement a new principal-evaluation system, also as a pilot, before embarking on the pilot phase of the new teacher-evaluation system.

including the Gates and Ford foundations (it does not receive any state or federal funding). The foundation does not engage in advocacy (unlike groups elsewhere such as Tennessee SCORE) but rather sees its role as supporting education reform in the state and being a "critical friend" to the Colorado department of education. The foundation, which currently has 40 staffers, has played a crucial role in providing supplemental funding and staffing for the department of education and the State Council on Educator Effectiveness. The foundation has also helped convene state stakeholders for conversations and hires consultants to provide expertise and guidance that can assist the state's department of education with strategic planning around the launching of new initiatives. Lopez says the "state education agency often does not have the resources or agility to be able to quickly turn on a dime and focus on new things but we are well-positioned to give them momentum quickly." She also notes that the foundation can work with districts in a way that the Colorado department of education cannot because the foundation does not have a compliance role. In addition, the foundation is able to create "a safe place for districts to fail and learn" and is in a position to be candid about districts' efforts and challenges. There is regular communication and coordination between the foundation, the department of education, and the State Council on Educator Effectiveness, facilitated by weekly meetings between the foundation and the education commissioner, not to mention the fact that Lopez formerly worked for the Colorado department of education and served on the council on educator effectiveness.

The foundation is also working with 13 "integration districts" on a project funded by the Gates foundation that is trying to provide focused support and analysis for how educator evaluation reforms are integrated with other reforms such as the implementation of new principal evaluations and new instructional tools aligned with common core standards and assessments. Lopez notes that the foundation is meeting with the state monthly to share information gleaned from the integration districts, which in turn can be used to inform their own work. An example of the kind of work being done by the foundation is the adaptation of the Harvard "tripod" student survey for use in the collection of student feedback as part of teacher evaluations, which the foundation funded. It will launch a pilot program using the "tripod" evaluation tool and will conduct a validation study as well as develop protocols for its use. All of the information and resources that are developed through the pilot process will be shared with the Colorado department of education for use statewide.

New Jersey

Quick facts

Number of teachers: 110,202
Number of schools: 2,634
Number of districts: 613

How New Jersey's teacher evaluations are structured:
New Jersey's reform legislation created requirements for four categories for teacher ratings based on multiple measures of student learning and growth, and multiple observations. It does not address how much student achievement should be incorporated into teacher evaluations. Districts can choose from one of the state-approved models or develop their own evaluation model, with evidence of validity. Statewide implementation is scheduled for the 2013-2014 school year.

The state's reform statute, TEACHNJ, created multiple ratings categories for teachers—highly effective, effective, partially effective, and ineffective—that are based on multiple objective measures of student learning and growth and multiple observations. The statute also called for the creation of an Educator Effectiveness Task Force during the 2010-2011 school year, which recommended a teacher-evaluation implementation pilot in 10 school districts in the 2011-2012 school year. Another 20 districts joined the original 10 during the pilot for 2012-2013, and other districts were instructed to build capacity for the statewide implementation scheduled for the 2013-2014 school year. The pilot guidance purposely was ambiguous on how and how much student achievement should be incorporated into teacher evaluations, noting only that standardized assessments must be used, but must not be the predominant factor. As of October 12, 2012 the New Jersey department of education had approved 14 different teaching-practice-evaluation instruments, which met the state's "technical requirements."[31] Districts can choose from one of the state-approved models or develop their own evaluation model (within one year) and present data to show its validity.[32]

One unique aspect of New Jersey's implementation of teacher evaluation reform was the use of a competitive application process for districts interested in participating in the pilot program. For the second year of the pilot, the state department

of education received 49 applications and 10 districts were selected to receive grants totaling $1 million to support implementation. Another interesting feature was the requirement that all districts in the state engage in a year of focused capacity building in advance of the roll out of the new evaluation system. It is important to note that as complicated and labor intensive as teacher evaluation reform is, the reform is hardly the only major initiative being implemented by the New Jersey Department of Education. The agency's capacity is being further strained by the simultaneous roll out of a new principal-evaluation system as well as a wide array of other reforms, including charter school expansion, school turnarounds, and the continued state department of education control of New Jersey's three largest districts: Newark, Jersey City, and Paterson.

New Jersey's Department of Education created a State Evaluation Advisory Committee and District Evaluation Advisory Committees to solicit feedback on the pilot program, with the district committees meeting monthly to discuss implementation challenges. The education department also contracted with the Rutgers University Graduate School of Education to conduct an independent evaluation of the year one pilot that includes site visits and administrator surveys.[33] The Rutgers evaluation has two components—a review of the process focusing on the implementation of the pilot and stakeholders' response to it; and a review of outcomes, which looks at the distribution of evaluation ratings and at student growth data. Another interesting aspect of New Jersey's work in this area is that the state began piloting a new teacher-evaluation system before the state legislature enacted the new evaluation and tenure law. This enabled the lessons learned from that pilot to be incorporated into the statute.

The first year of the pilot program revealed a number of challenges. The majority of teachers do not have standardized assessments for their students, and even for those who do, the state's data system is only now starting to link student scores to individual teachers. Based on the experience in the first year of the pilot, the state made a number of changes for the second year. It increased the number of observations for core teachers, introduced the use of double-scoring (two independent observations), shifted to unannounced observations, and increased flexibility in weighting for tested and nontested grades and subjects. In addition it required schools to use external evaluators to provide a second set of observations to deal with the issue of principal bias. The external evaluators must come from another school, from the district office, or are retired but still retaining certification; and every evaluator has to be trained and calibrated on the rubric being used by the district. Evaluator training is not done by the state department of education but is left to the districts, many of

which are outsourcing the training, in some cases through new multidistrict consortia.[34] Peter Shulman, chief talent officer/assistant commissioner of teacher and leader effectiveness for the New Jersey Department of Education[35] observes, "We are being less prescriptive because there is no perfect way to do this work. If we knew the answers we would tell them but we aren't sure. We will hold them accountable for outcomes and results and are trying to provide tools and guidelines to help them get there." He adds that local control is a "big part of the fabric in New Jersey," which has more than 600 school districts.

Tim Matheney, director of evaluation in the division of teacher and leader effectiveness at the state department of education, says that the implementation of evaluation reform "is a really heavy lift" and described the department as focusing on three waves of regulatory work: requirements around building district capacity, using information from the pilots to improve observation protocols, and fleshing out how the new evaluations will intersect with the new tenure law. They have also been working with private vendors to ensure that the training and evaluation instruments the vendors are providing to districts meet state expectations—2 of the 15 instruments submitted were rejected. Matheney notes, "We have to tread carefully if we are going to be really specific because we are a local control state. We want to honor the local knowledge that comes with that but we also have to have some common statewide expectations."

In the summer of 2011 the New Jersey Department of Education surveyed its 580 superintendents and found that almost three-quarters of them believed the department did not play a role in helping to improve student achievement.[36] That same summer Chris Cerf, the state's new education commissioner, initiated a radical redesign of the state education department with the expressed purpose of better enabling it to support district reform efforts. He has restructured the organizational chart and reassigned staff around four areas: academics, performance, talent, and innovation—and all four offices are focused on service delivery. According to chief talent officer Shulman: "For our priority and focus (low-performing) schools, we want to have direct intervention support." He says state education agencies have traditionally "fallen into the one-size-fits-all mantra but now we are trying to provide support at the granular school—if not classroom—level for about 250 (10 percent) of the lowest-performing schools in the state."

The state has also created seven new Regional Achievement Centers, each with a staff of 10 to 15 drawn from the state education agency who each specialize in different areas. Shulman says the "idea is to make sure that you have the right cure for

In the summer of 2011 the New Jersey Department of Education surveyed its 580 superintendents and found that almost three-quarters of them believed the department did not play a role in helping to improve student achievement.

the right ailment" and that the regional centers have created an "unprecedented opportunity for two-way dialogue." It remains unclear exactly how the regional centers will operate in practice. Shulman admits that "they are still thinking about how all the pieces connect," but the regional centers have the potential to connect the different strands of school improvement work with teacher evaluation reform. The New Jersey Department of Education has created implementation teams that provide support to districts, both those in pilot and those in the preparation phase. With only four or five staff (total) assigned to this work, however, they don't have the capacity to be in all districts so they use a "triage" approach to support those districts most in need.

The institutional reorganization has been accompanied by a major overhaul of personnel in the state department of education. Shulman notes that state education agency staff are not "widgets" and that he "needed to get the right people with the right skills on my team." He has reconstituted the office with 50 percent of the original senior staff gone—some people have been moved out or moved to other areas, while others were counseled out. "We needed to get a good combination of New Jersey educators as well as folks who are philosophically reform-minded and have an understanding of where the industry is moving," says Shulman. The state department of education hired an individual from the Harvard Strategic Data Project to provide needed quantitative expertise for the teacher-evaluation system. All of this restaffing has come at a price and Shulman admits that he has spent a great deal of time on departmental personnel matters.

Part of the challenge is the very small pool of people who are qualified and experienced in the area of teacher evaluation and the demand for their services has created a poaching problem. (Shulman, for example, led the teacher-evaluation work in Delaware before being recruited to New Jersey.) Shulman notes that the New Jersey Department of Education is trying to create a "high-performing organization"—like a Google or a McKinsey & Company—with a culture that can constantly recruit, train, and recycle good staff. He argues that human capital changes will take some time, pointing out that you can't only bring people in from the outside as there is a need for institutional knowledge. But Shulman believes that strategically bringing outside consultants onboard for a "limited engagement" is a fiscally prudent way to bring talent to the department and it avoids the need to hire a full-time employee who may not be needed or affordable over the long term. He conceives of this work as being done in stages and notes that building new evaluation systems is different than sustaining these systems and may require different skill sets.

Overall, Shulman says that he and the department have learned four key lessons in New Jersey thus far: "First and foremost we have learned about the importance of stakeholder engagement and communication. We have heard from the field how important the District Evaluation Advisory Committees have been –having this governing or advisory board has paid significant dividends in building transparency and trust—and so every district is going to have this." The second lesson, according to Shulman was the need for "training, training, and more training—high quantity and high quality." Selection of the observation instrument was left to districts (from a state-approved list) but most have chosen the Danielson Framework for Teaching. Shulman points out that the specific instrument is less important than how it is implemented. The third lesson that has emerged is the major challenge that exists around nontested grades and subjects, where standardized test scores do not exist and which require some other metric for calculating a teacher's growth or value-added score to be developed. The New Jersey Department of Education is trying to strike the right balance between giving districts flexibility while also providing guidance along with "some level of sandboxing"— establishing outer boundaries that constrain flexibility—to the process to ensure a degree of quality control. But as Shulman points out, "We are trying to get out of the compliance business and into the service-delivery business." The fourth and final lesson is the capacity issues for principals—they are being asked to spend significantly more time conducting evaluations and providing feedback and guidance for teachers but are struggling to do this along with their other responsibilities. The state education agency is trying to help principals with time management to allow them to get it all done.[37]

Matheney, the state's director of evaluation, observes that the early pilot districts were all high-capacity districts that volunteered to take part, and that these districts were "fertile soil" for implementing the reforms. He believes that most embraced the reforms and were well-prepared to implement the initiative though some struggled due to competing demands, problems with the timing and quality of training, or inadequate communication with stakeholders. Still, there are concerns about what will happen when the reforms are taken statewide. As Matheney notes, "Some districts have overburdened leadership, or don't embrace innovation and are stuck in their ways. We anticipate that there will be some recalcitrant districts that don't see as much value in this work as we do. We are concerned that they might have little intention to do this work well."

For Shulman the success of the reforms is pegged to buy-in: "We are trying to weave this into the fabric of the district, and that buy-in by folks at the local level

is very important. There is a bridge between theory and practice that needs to be crossed so that the understanding on the ground and in the field matches that at the state level. It comes down to communication." Even as the New Jersey Department of Education tries to shift toward more service delivery and support, Shulman says there is "still an element of compliance and monitoring, that doesn't go off the table completely." The department is developing new regulations and requires districts to submit two progress reports during the year. Shulman identified three key fears or challenges going forward. The first is that the reforms are a major cultural change and it takes time and the political and operational fortitude to see it through, especially in the face of gubernatorial transitions. Second is the need to better connect evaluation to professional development—to find ways to use data to identify teachers' strengths and weaknesses and drive growth. The third takeaway, according to Shulman, is removing bad teachers, but just as important is raising the bar and expanding the pipeline or pool of new teachers.

Pennsylvania

Quick facts

Number of teachers: 129,911
Number of schools: 3,269
Number of districts: 500

How Pennsylvania's teacher evaluations are structured:
Pennsylvania's new teacher evaluation system is based on traditional teacher practices and classroom observations (50 percent) and multiple measures of student achievement and growth scores (50 percent). The new system is to be implemented statewide during the 2013-2014 school year, following three years of piloting in numerous school districts. Districts are allowed to use any state-approved model.

In Pennsylvania's new teacher-evaluation system 50 percent of the evaluation comes from traditional teacher practices and classroom observations and the other 50 percent from multiple measures of student achievement and growth scores. One unique aspect of the Pennsylvania approach to implementing the new system is that they are spending a full three years piloting the new evaluations instead of the single pilot year being employed in most states. The first round of Pennsylvania's pilot program took place during the 2010-11 school year with 10 districts; the second year (2011-12) the pilot expanded to 100 districts; and in the third year (2012-13) it has gone to 300 districts. The legislature has called for implementing the new system statewide during the 2013-2014 school year. Carolyn Dumaresq, deputy secretary in the state's Office of Elementary and Secondary Education, says that as a result of the longer pilot process "we have the luxury of more time to do it right. We weren't hurried into jumping into one measure." She views losing in the first round of Race to the Top as a mixed blessing since it freed them from being forced to meet an accelerated timeline for implementation.

Districts in the state are allowed to use any state-approved model but the state education agency only provides training and support for one model know in educational circles as Danielson, an evaluation tool created by The Danielson Group, a highly respected educational consulting firm. As of May 2012, however, no other models had been added to the list as the state department of education was waiting to review

the results of district pilots. This approach has enabled the state education agency to concentrate limited resources in providing greater depth of support for a single model than would have been possible for multiple models. The Pennsylvania department of education is providing technical assistance to support implementation of the new system in two different ways: through two-day face to face "coaching training" sessions for classroom observers, which is provided through regional "intermediate units"[38] and through the creation of a free, online professional development center. Pennsylvania's department of education is in the process of developing a module for each of the 22 components in the Danielson rubric that includes videos of proficient teaching. The videos enable teachers and evaluators alike to gain a better understanding of what they should be looking for in their own work. According to the state's department of education "these trainings will help participants 'unpack' the components of each domain and identify what evidence looks like at each level of proficiency."[39] Teachers also get professional development credit for completing the online modules. The state is drawing on information gleaned from the three years of piloting to refine their model and the associated training. Teachers were candid about their strong belief that the model used during the first pilot did not work. The willingness of teachers, principals, and state education agency staff to be open and honest about the strengths and weaknesses that emerged from the pilot was very important and led to a switch to a new model. Dumaresq also reports having learned a lot from the regional education labs, which have provided webinars and virtual learning around evaluation reform.

Dumaresq notes that a great deal of "retraining, retooling, restructuring, and reshuffling of responsibility" has gone on at the Pennsylvania Department of Education to enable it to better support the school reforms enacted in recent years. She says the need to create a multiple measures model that incorporated student achievement data required the state education agency "to bring in psychometric kinds of folks because that kind of expertise wasn't there." As a result, the state education agency is partnering with external consultants, including Mathematica and SAS Inc., to provide this expertise and to connect the state's new value-added model to the common core standards. Some of this work was funded by the $41 million federal grant Pennsylvania received in round two of Race to the Top. A small Gates foundation planning grant also enabled the state's department of education to hire three additional staff—one focused on qualitative evaluation, another on quantitative evaluation, and a third dedicated to training. Deputy Secretary Dumaresq believes they now have a "good team of internal and external folks" in place:

> *We are going to leave capacity behind here after we finish the development phase*
> *which is especially important when the development money and consultants go*

*away. We need to make sure that our own staff inside the department under-
stands the process and the program to be able to continue to support it. What we
have started to do incrementally each year is move different pieces of the work
into our departmental budgets.*

One of Pennsylvania's largest school districts, Pittsburgh, began its own evaluation
reforms even prior to the state effort thanks to a Gates Foundation initiative. One
difference between the two reform programs is that Pittsburgh is using differ-
ent value-added and multiple-measures models, the implementation of which is
supported not by the state but by the Gates grant. As a result of its earlier start it
appears that Pittsburgh is farther ahead of the state in doing this work, but it has
also given rise to some concerns, specifically whether the state education agency is
facilitating, obstructing, or simply playing a neutral role vis-a-vis the district?

 Sam Franklin, executive director of the Office of Teacher Effectiveness for the
Pittsburgh schools, sees the Pennsylvania Department of Education as having
been supportive of the district's efforts around teacher-evaluation reform. He
believes the bigger challenge is how the state education agency supports the state's
500 other districts, particularly the many smaller ones that lack resources and
are much further behind in doing this work than Pittsburgh. No other district in
Pennsylvania, he points out, has moved beyond piloting new observation models
and some haven't even done that yet. Franklin did note that it would be helpful for
the state to "move more quickly and comprehensively" on expanding the assess-
ment portfolio an area where most districts lack competency. "You don't want to
have districts across the state developing lots of different assessments at different
levels of quality. There's an advantage to having tests that are consistent across
the state where you can have comparable data. And developing new assessments
shouldn't be driven just by the need to evaluate teachers."

For her part, Dumaresq says "the state has benefited" from the Pittsburgh experience
but "has had to figure out which parts of the [Pittsburgh] pilot could be expanded
statewide and which could not." Franklin has concerns that fiscal sustainability issues
may be an obstacle to statewide adoption of his district's teacher-evaluation model,
noting that outside consultants such as "Mathematica and Cambridge Education
are not cheap." He says the Pittsburgh district is "working to determine what the
ongoing costs are" and looking for ways to "incorporate these costs into our budget."
He notes, however, that large or medium sized urban districts with more sizable
operating budgets such as Pittsburgh "may have more ability to reallocate resources
and staff time to support this work than smaller districts."

Delaware

<div style="background-color:#dce6f0">

Quick facts

Number of teachers: 8,933
Number of schools: 218
Number of districts: 19

How Delaware's teacher evaluations are structured:
Delaware has a single statewide evaluation model but local districts have the option to use an alternative evaluation model in conjunction with the state model. The new statewide evaluation system: establishes four levels of educator performance; uses multiple valid measures in establishing performance levels; requires no more than five components with one dedicated exclusively to student improvement (growth) and weighted at least as high as any other component. The new teacher-evaluation system was piloted during the 2011-2012 school year with implementation in the 2012-2013 school year.

</div>

Delaware has a single statewide evaluation model—the Delaware Performance Appraisal System— but gives districts the option of using an alternative evaluation model in conjunction with the state model. As of September 2012 no district has exercised that option. Delaware piloted their new teacher-evaluation system during the 2011-2012 school year and is implementing it statewide this school year (2012-2013). The state has three distinct advantages over other states in the implementation of teacher-evaluation reform. First, as a first-round Race to the Top winner, Delaware received a large grant ($100 million), which has enabled it to dedicate more state education agency staff and resources to the work. The second advantage is that Delaware already had a statewide teacher evaluation system in place since the 1980s. Diane Donohue, special assistant for educator effectiveness in the state's department of education, allows that as a result Delaware is "ahead of the game" compared to other states and therefore is able to focus on refining the existing system rather than starting from scratch. A third advantage enjoyed by Delaware is its small size—it only has 19 districts and about 230 schools—which has enabled the state education agency to operate almost like a local education agency, giving it more direct contact with district and school leadership than is possible elsewhere and making implementation of its evaluation program more manageable overall.

The Delaware Department of Education is very small—only 270 people—compared to other state education agencies. The department was reorganized in the wake of Race to the Top and developed a teacher leader-effectiveness unit (which has four staffers) along with a delivery unit that serves as the project manager for their Race to the Top grant. The state's department of education has used its Race to the Top funds—as well as some repurposed state funds—to develop an extensive network of supports for schools and districts to facilitate the implementation of the new teacher-evaluation system. Delaware's Department of Education also hired nine development coaches for a two-year program—in partnership with the University of Delaware's Academy for School Leadership—to work with evaluators to improve the accuracy of the evaluation system. The department also used $8.2 million of its Race to the Top funds to hire 29 data coaches to work with teachers, principals, and administrators to expand their capacity to analyze student data as part of the growth portion of the new evaluation system. The department of education contracted with Wireless Generation, a leading provider of education software, to manage this program. In addition, it partnered with the U.S. Education Delivery Institute, an innovative nonprofit organization that focuses on implementing large-scale system change in public education, to provide technical assistance to districts for planning around the Race to the Top reforms.

The Delaware Department of Education asked each school to create a "school team" that included (at a minimum) the principal, a school specialist, and two teachers (one of whom was a union representative) and then did countywide, full-day training on the new evaluations. The idea was to get the information to a team in each school who in turn would take the program information back to the rest of the building staff. The Delaware Department of Education is trying to visit every school in the state to solicit feedback and address any questions that arise. It has also contracted with an outside vendor to evaluate the evaluation system every year through surveys and focus groups.[40] As part of the monitoring process a team from the department visits every school to assess the way they are evaluating teachers. The U.S. Education Delivery Institute has praised Delaware's local education agency support program as a model for the rest of the country.[41]

Rhode Island

Quick facts

Number of teachers: 11,212
Number of schools: 325
Number of districts: 32

How Rhode Island's teacher evaluations are structured:
Rhode Island's evaluation reform is written into the regulations of the State Board of Regents for Elementary and Secondary Education which established a default state model that all districts are required to use unless they propose an alternative model approved by the state education agency. Educator evaluations must contain three components based on evidence of professional practice, professional responsibilities, and student learning. The state is silent on the percentage that each component contributes to the final rating. Statewide implementation of the full evaluation system is taking place during the 2012-2013 school year.

Rhode Island has taken a different approach to educator evaluation, one that stands out from most other states. It has established a default state model—the Rhode Island Model Educator Evaluation System—that all districts are required to use unless the state says otherwise (districts can propose alternative plans, but those plans must be approved by the state education agency). Mary Ann Snider, the chief of educator excellence and instructional effectiveness at the Rhode Island Department of Education, reports that about 80 percent of districts have adopted the state model, with six urban districts using an American Federation of Teachers model, and one district using its own locally developed model. But even the alternative models have to implement the student-learning component that is featured in the state's evaluation model. As a result all teachers must write student-learning objectives with some getting a growth score in an effort to, as Snider notes, limit variation in the student learning component. As she says, "We are trying to keep things as tight as possible these first few years so we can get a more accurate distribution of teacher effectiveness."

Unlike most states, Rhode Island's evaluation reform is not written into state statute but instead in regulations set by the state's Board of Regents for Elementary and Secondary Education that require every district to have an approved evaluation

system and to evaluate every teacher—on evidence of student growth and achievement—every year. Snider believes the board's set of regulations have been extremely helpful: "It gives us the force of law but with additional flexibility so that if we need to revise them, it is much easier than having to go back to the statehouse."

Partial implementation of Rhode Island's evaluation reforms occurred in the 2011-2012 school year, when every district was required to conduct two of the four annual teacher observations required by the new evaluation model. The state's department of education solicited significant feedback—surveying every teacher and principal in the state and conducting focus groups. The department then revised their evaluation model based on that feedback. All training on the Rhode Island educator evaluation model and the student-learning component for districts using alternative models was planned and delivered by the state's department of education, which hired outside consultants to help districts with the other parts of the system. The department of education also conducted a four-day summer academy for evaluators and participants who had to complete about 25 hours of online work to calibrate their observation ability with the rubric, including a calibration test at the end of the training session. Based on the initial results and feedback the state education agency decided to have two additional days of training for evaluators during this school year. Statewide implementation of the full evaluation system is taking place now, during the 2012-2013 school year.

In order to better support the implementation, the Rhode Island department of education was completely restructured, creating divisions to unite different units that had previously not communicated or coordinated very much. "Creating a connection between educator evaluation and curriculum, instruction, and assessment is really unusual among states but it has really helped us with alignment and preparation for the transition to the common core and common assessments," says Snider. There has been no new state money to fund the evaluation work in Rhode Island but the state was awarded a second-round Race to the Top grant of $75 million in August 2010 and they have relied on the federal funds to hire new department of education staff. "We hired a few new people but we way underestimated the lift it would take to get a new education evaluation system ready," admits Snider. She says that miscalculation has meant that the staff in the department has been "working extensive hours over the past few years." Despite the heavy workload, Snider believes that part of the success in Rhode Island is due in part to avoiding the instability in state education agency staff that has been common in other states. She says that the Rhode Island Department of Education is "proactive" about checking in with staff to enhance productivity and limit turnover.

Limiting staff turnover, according to Snider has been important because "there are not a whole lot of people out there with deep evaluation expertise." Moreover, she notes that there is a lot of fear and anxiety when outsiders are brought in. She says the "perception is that they are there to fire people." In an attempt to ease fears, when the department of education sought to hire some 20 intermediary service providers (outside consultants), who were eventually deployed in the field to support districts and work with principals and teachers on evaluation, it focused on in-state recruiting for those positions. Some of the consultants ultimately came to work full time for the department on educator evaluation. Going forward, however, the state is taking steps to build and sustain their capacity over time, in part by working with the Rhode Island higher education sector to ensure that the necessary skills are being taught in teacher and administrator certification programs.

Developing effective lines of communication and building trusting relationships with districts has been crucial according to Snider. "We have been successful in changing the mindset about the role of the SEA [state education agency]—principals and superintendents say that it has been a different ride over the past few years. We have gone out of our way to be present in the school district and to provide as much of a partnership as we can."

The Rhode Island Department of Education meets on a monthly basis with every assistant superintendent and curriculum director in the state to check in and share information and build their capacity. One example that Snider highlights is the Collaborative Learning for Outcomes process, where a team from the state's department of education visits every district and gives them a "self-reflection rubric" to prompt them to think about what good implementation looks like and to self-assess their progress. "We don't pass any judgment," Snider notes. "We just ask folks to tell us where they are and what they are struggling with and what is unclear to them." The teams that do the district site visits then write up those reports and give them to the program directors in the department of education who determine where more support might be needed. Snider notes the districts like this approach: "What we are saying is that we understand that this is hard work and if districts are open to trying to get better then part of the change is we are not shutting them down, asking for reports, monitoring them, being a big brother—we are trying to provide the support that they need."

Snider says the state education agency tends to be more "insistent" when districts pull back from the work altogether: "Our reform model depends on everyone working to continuously improve. We won't negotiate our efforts to close student achievement gaps. At this time, everyone understands that and this approach seems to be working in our state."

Partial implementation of Rhode Island's evaluation reforms occurred in the 2011-2012 school year, when every district was required to conduct two of the four annual teacher observations required by the new evaluation model.

Key lessons and challenges

Role of state education agencies: A philosophical/statutory/
constitutional debate over centralization

It is important to recognize that all state education agencies are not the same—that each state agency has a unique history and operates in a different fiscal, political, statutory, and constitutional context. In particular, states vary significantly in their attachment to local control of schools and the proper role of the state in education. This has a major impact on how state education agencies approach teacher-evaluation reform. Janice Poda, strategic initiative director for the education workforce division at the Council of Chief State School Officers, says that many western states, for example, are simply philosophically opposed to an active role by state education agencies and are resistant to the idea of standardizing the teacher-evaluation process across districts. Additionally, there are also constitutional limitations on the role of the state education agency in some states such as Colorado. Being familiar with the national education landscape, Tennessee's assistant commissioner for teachers and leaders Sara Heyburn observes, "The state role varies drastically from state to state in terms of how much local control exists. It has huge implications for what the state attempts to do or doesn't do and the kinds of support you offer at the state level versus how you facilitate the right things to be happening at the district level."

A related issue revolves around the traditional state education agency focus on compliance and accountability activities, which has made local education agencies wary of being candid about whether and how they might be struggling to implement reform, causing reluctance to seek out assistance. There has been a lot of talk about the attempt of state education agencies to shift from a compliance role to a service-delivery role, but the agencies are struggling to figure out how to fulfill both functions simultaneously. Adam Tucker of the Gates Foundation acknowledges, however, that this is hard to do and requires a different approach to implementation. Dan Weisberg, executive vice-president and general counsel for performance management at TNTP reiterates this point, noting that "there is this perception that the agency that has the ability to take money away and take other punitive action against

districts can't also be a support to the entities they regulate." He draws a parallel to the challenge that health and safety agencies such as the Occupational Safety and Health Administration have in balancing its inspection and fining function with its workplace safety improvement function. Weisberg argues, "There is nothing mutually exclusive about the two functions—in fact the dual role is absolutely critical." He adds that there " are not enough districts that are going to be able to do this work successfully without both support and accountability."

Even where a state education agency may have the resources and constitutional and statutory authority to be active, it may lack the relationships and trust with district leaders that are essential to ensure effective collaboration. Tucker emphasizes this point, remarking that "district capacity and willingness" to follow the lead of the state education agency varies widely within and across states. He also believes some local education agencies are

> …ready to fully embrace the work, and there are others that are concerned about the state agency's ability to execute a program that they think is right for their district. Because there is room for interpretation in the policy, this complicates their ability to implement it effectively at times. So issue number one is establishing a definition of roles, a clarity around who is responsible for doing what and who will pay for the different pieces of the work.

State education agencies need to consider comparative advantage and economies of scale—where the state can provide a service or function that districts cannot. Pittsburgh's Franklin notes, "SEAs should pay attention to the aspects of the work where they have a unique ability to solve a problem and where there are good economies of scale and return on their investment." Providing technical assistance, communication networks for information sharing, and policy interpretation are three areas where state education agencies could add real value.

Flexibility in state evaluation systems varies greatly

States vary in the degree of centralization and standardization they have mandated—either in statute or in regulation—in the new teacher-evaluation systems, and this is having a major impact on the approach to supporting implementation by state education agencies. Sandi Jacobs, vice-president and managing director of state policy at the National Council on Teacher Quality, notes that some states have mandated a single statewide model (Delaware), some have created a "pre-

sumptive" state model with an opt-out only for districts that receive state approval (Rhode Island), and others (Colorado, Indiana, and Illinois) have adopted an "opt-in" approach with an available state model that districts can use at their own discretion. Still other states have created state-approved lists of models that districts can choose from (Oklahoma), some (Florida and Maryland) have given districts free reign to select any model that they believe suits local needs pending state approval, and other states are leaving districts completely on their own (Arizona, New York, Minnesota). The National Council on Teacher Quality, however, cautions, "… State review and approval of district evaluations may not be an adequate approach to ensuring quality and rigor… States that have left districts to their own devices without any oversight are even more worrisome. There is good reason to be skeptical that all districts in such states will have the capacity and will to implement strong evaluation systems on their own."[42]

A clear tension is emerging between states' desire to give districts flexibility to select or adapt evaluation instruments that are best suited to their particular circumstances, and the limited capacity of state education agencies to provide implementation support for a wide array of instruments. The Gates Foundation's Adam Tucker notes, for example, that New York has approved 6 or 7 different observational rubrics for districts to use and also permits districts to submit others for approval that meet state criteria. As such there is clarity in the state about the evaluation rubric but it is not clear whether the state education agency or the local education agency is responsible for certifying evaluators in the use of the rubric. Even in places that seem to have settled on a single model—often the Danielson model—several districts are modifying it to such an extent that the state education agency is unable to provide support. Such modifications will also potentially undermine or complicate the desire to produce teacher evaluation data that is comparable statewide. The Pittsburgh school district, for example, modified the state evaluation instrument to incorporate "equity" and while the Pennsylvania department of education approved the change it announced it would not provide support for the modified system, leading to concerns about the impact the changes will have on the state accountability system.

State education agencies need to consider comparative advantage and economies of scale—where the state can provide a service or function that districts cannot.

State education agency restructuring and the human capital demands

State education agencies in many states are undergoing radical restructuring and restaffing as they embrace a shift from being compliance monitors to service-delivery/school-improvement organizations.[43] This restructuring is difficult and

time-consuming work, and while necessary to carry out new responsibilities over the long term, in the short term it is creating a number of challenges. It will take some time for this organizational shakeout to be completed and for new structures to be put in place and staff to acclimate to their new roles. Many state education agencies have created new teacher-effectiveness units, but the degree to which these units have been well-integrated with other units appears to vary and long-standing concerns about agency siloing persist.

There are also concerns about state education agency capacity gaps—that the resources and staff available to support local education agencies with teacher-evaluation reform are inadequate. Sir Michael Barber, an architect of British education reform currently working with the U.S. Education Delivery Institute emphasizes the importance of what he calls the "mediating layer" in education reform—subsidiary structures that can build an "effective delivery chain" that translates state policy changes into positive change at the school level.[44] Some states such as Pennsylvania have long had regional intermediate units but are now changing their role, while other states have opted to create entirely new institutions such as New Jersey's Regional Achievement Centers.TNTP's Dan Weisberg cautions:

> *SEAs have very limited resources but the need at the district level for implementation support around these very ambitious initiatives is very large. It is not going to be possible for SEAs to provide intensive implementation support for every district or even half the districts in a state. States need to have a plan about how to use their resources strategically; that means doing a gap analysis and focusing on high-need/low-capacity areas.*

But state education agencies are struggling to staff these new structures. Sandi Jacobs of the National Council on Teacher Quality notes that historically "the teacher units" in state education agencies were " very compliance oriented." She points out that "you can't simply re-task existing staff" and have, for example, the person previously responsible for teacher certification (mostly doing transcript review) "now lead the teacher evaluation implementation effort." The Gates Foundation's Tucker argues, "Capacity is about talent—in some places this is about having enough bodies, but also about having enough bodies with the right skill set and experience to get the job done. This is complicated by the fact that there are real invention challenges around this work—you need folks who are smart and savvy enough to invent solutions to these problems and dilemmas and challenges that are new to us all. That is a different kind of talent issue."

There is a very small pool of people in the country who possess the skills and experience necessary to do this work and demand vastly outstrips this supply. As a result, a serious poaching problem has emerged with state education agencies competing with other state education agencies, local education agencies, the U.S. Department of Education, consulting firms, and foundations to retain the services of these professionals. Jacobs notes, "…there is poaching in every direction and some of the SEAs that built up great staffs have then had to deal with significant turnover." For this reason, state education agencies need to think about how to generate new teacher-evaluation talent and not just recycle existing talent.

Another challenge around human capital centers on the instability that can result from leadership turnover due to attrition or electoral change. As noted above, state education agencies across the country are undergoing major restructuring and re-staffing and this is upsetting long-established relationships and lines of communication between local education agencies and the state. In addition, elections regularly bring changes in control of state legislatures and governorships and thus also in political appointees to state education agencies, presenting a major challenge to effective reform implementation. As Jacobs observes, "We have had a huge turnover in governors and state chiefs since [Race to the Top] grants were awarded and in many states the people responsible for implementing these reforms are not the same people who wrote them and this leadership change is causing problems in some states." There are reportedly a number of states where the legislature is either unsupportive—or in some cases—actively opposed to the teacher-evaluation system being rolled out by their state education agency and this too has caused problems during implementation.

Internal vs. external capacity

In the short term, state education agencies are dealing with their internal capacity gaps by relying on two different kinds of external capacity—outside consultants and foundations. The Gates Foundation is supporting teacher-evaluation reforms in a number of state and local education agencies across the country. And the positive impact of the Colorado Legacy Foundation has been praised by U.S. Secretary of Education Arne Duncan and was the inspiration for legislation recently passed in Kentucky to set up a similar foundation. There is some concern, however, that reliance on outside consultants and foundations may preclude or delay the development of the fiscal self-sufficiency and internal capacity that can

support these systems over the long- term. Another concern is that "outsiders" do not bring the knowledge of state context and networks of relationships that can build crucial trust during difficult implementation work. Some observers worry about what will happen when the outside funding, which is making much of this external capacity possible—such as federal Race to the Top grants and private philanthropy—dries up. By contrast, others believe that the capacity demands differ over the short- and long-term, and that once the initial "heavy lift" and large "start-up costs" associated with developing and installing new teacher evaluation systems are over the role and resource needs of state education agencies will be less intense. Hiring outside consultants on an as needed basis, others say, also enables them to be more flexible in their personnel decisions and prevents the need to "bulk up" the state education agency.

Tennessee's Sara Heyburn believes that state education agencies have to rely mostly on their internal capacity noting that contracting out the work ultimately is not sustainable: "You need to embed capacity inside your department, however, there are some areas where outside support is warranted and helpful—to support the work but not drive the work." Some state education agencies such as Tennessee's are beginning to think about how to shift these costs on budget and to reallocate and retrain department personnel or to hire new staff with the competencies and experience necessary to undertake this work. One interesting development in this regard has been the shift of former foundation and consulting staff to positions inside state education agencies—in this sense, external capacity is *becoming* internal capacity. (Outside consultants like Dumaresq, for example, who previously headed up the Gates Foundation team in Pennsylvania and TNTP staff, are sometimes being hired by the state education agency at the end of consulting projects.)

Tucker of the Gates Foundation agrees that state education agencies need to figure out the kinds of activities that are best to contract out and which should stay in house. He says the notion of a state education agency "that can do everything for everyone all the time is a pipe dream … Both from a resource perspective and in terms of having the nimbleness, innovation, and entrepreneurial spirit that helps to move an agenda over time." He says it is important that state education agencies strike the "right balance," even if it is necessary to shift that balance over time in terms of where capacity exists. But he argues it's "fair to say that whether it is in-house or out of house that capacity is still quite thin in this arena." Some states such as Rhode Island, are looking to develop home-grown capacity by working with their colleges and universities to incorporate teacher evaluation into teacher preparation and administrator training programs.

Some observers worry about what will happen when the outside funding, which is making much of this external capacity possible—such as federal Race to the Top grants and private philanthropy—dries up.

Funding streams and the fiscal cliff

Janice Poda of the Council of Chief State School Officers reports that their survey of state education staff revealed a great deal of concern about the lack of capacity to implement these reforms at state agencies. The survey also revealed that most states are purchasing "off-the-shelf" evaluation systems rather than developing their own due to time pressures, cost concerns, and limitations in their technical capacity. For states that did not win a Race to the Top grant or secure foundation support (which is the majority of states) the resource and staffing issue around teacher-evaluation reform is even more pronounced. Given the current tight fiscal climate, most states have been unable or unwilling to allocate new money to support the implementation of these reforms. As a result state education agencies have had to do one of the following: push costs to the district level; focus on low-performing and/or low-resources districts; adopt a "train-the-trainer" model; or reallocate staff and funds from other activities. South Carolina, for example, was identified as a state that simply has "no resources" to design and implement new teacher-evaluation systems. Iowa, on the other hand, has devoted 15 state education agency staff to oversee its implementation, which is positive though is still considered inadequate.

An additional issue centers on the fact that teacher-evaluation reforms are currently supported by multiple funding streams—state appropriations, local education agency budgets, federal grants, national and state foundation grants—and figuring out who should pay for what needs to be resolved over the longer term. The timing and sustainability of "external" funding to state and local education agencies is also an issue as grant-application processes often are poorly aligned with the implementation timetables out in the field. The federal Race to the Top program, for example, encouraged states to move forward rapidly with the roll out of evaluation reforms but it took a considerable amount of time for those federal funds to reach state coffers and much of that money remains unspent.

Many of the people interviewed for this paper expressed concern about the looming "fiscal cliff," when external funds from Race to the Top or private foundations runs out. TNTP's Weisberg sees this as a huge concern: "Making a sustainability plan is very important given the one-time nature of the funding—you have to openly confront how you are going to reallocate stable funding sources to support implementation when the funding cliff arrives." State education agencies appear to vary widely in the way that they have spent external funds, the degree to which they are dependent on them, and the extent to which they have begun to bring these expenses on budget. As a consequence the "fiscal cliff"—when it hits—is likely to affect states in different ways.

Evaluating the evaluators

One of the primary activities of state education agencies in supporting local educa-
tion agencies with teacher-evaluation reform has been providing training to the
administrators who will be conducting the new observations. States vary widely in
their approach, however, for both philosophical and capacity reasons. Some state
education agencies, Tennessee for example, are directly training all evaluators, some,
including Colorado and Pennsylvania, are adopting a train-the-trainer model, and
others such as New Jersey are leaving the training entirely up to districts. Poda, of
the Council of Chief State School Officers, cautions that there has been "too much
focus on evaluations themselves and not enough on the evaluators who will be using
them and how they are trained, especially in terms of giving feedback to teachers
for improvement." Most states are relying—entirely or in part—on principals to
evaluate teachers. The new evaluation systems require a much greater quantity and
quality of teacher observations, along with enhanced feedback, and principals are
clearly struggling to find enough time to complete these tasks along with their other
responsibilities. In addition, some observers question whether principals can bring
the necessary objectivity to the evaluation process given their often close relation-
ships with teachers. Others note that given the lack of incentives and sanctions that
principals have to motivate teachers, evaluations have emerged as "principal pay-
back," and a way for school leaders to reward or punish staff for their loyalty, referred
to as "management by Santa Claus." A related issue centers on whether states should
implement teacher evaluation before, after, or alongside the rollout of new principal-
evaluation systems. Colorado is introducing new principal evaluations before new
teacher evaluations, New Jersey is doing them simultaneously, while Pennsylvania
is not putting its new principal-evaluation system in place until the year after its new
teacher-evaluation system is in place.

There are two very different ways that state education agencies are attempting to
support districts with the implementation of new teacher-evaluation systems—
actively and passively. Some agencies are actively reaching out to districts and
schools and providing face-to-face training, monitoring, and technical assistance,
though the amount of this kind of outreach is contingent on the amount of staff and
financial resources available (which as previously noted varies widely across state
education agencies). While other state education agencies—particularly those lack-
ing the capacity to provide active outreach—are developing online resources that
can be accessed by teachers and principals for free, anywhere, anytime. The types
and quality of online materials appears to differ from state to state but this appears
to be a relatively low-cost and effective way for state education agencies to support

districts. A frequent criticism by teachers and evaluators alike has been the lack of accessible examples of what constitutes "effective" teaching under the new standards.

Implementation timetables and sequencing

Individual states need to think carefully about the work required to implement a new teacher-evaluation system and assess the capacity that they will need to do the work, how the evaluation work should be sequenced with other related reforms, and design implementation timelines accordingly.[45] Most state reform statutes have established rapid timetables for the installation of new teacher-evaluation systems. While all states are struggling to meet these timetables, it is becoming clear that some states are struggling more than others largely due to the fact that states vary in terms of their experience with statewide evaluation systems. Some states such as Delaware and Tennessee have long experience with such systems while other states have little or no experience. It is important to recognize that this varying experience will undoubtedly impact the speed and effectiveness with which new systems can be implemented. Michelle Exstrom, an education program principal for teaching quality and effectiveness at the National Conference of State Legislatures, notes, "State legislators should consider the timelines that are best for their state context – what their data system can and can't do and the capacity of their SEA—rather than simply adopting other states' legislation and timelines."

There is a clear tension between the desire to do this important work quickly and the many obstacles that need to be surmounted to do the work well—a process that takes time. As Jacobs of the National Council on Teacher Quality states: "Finding the sweet spot between the real urgency and need to do this as fast as we can, and recognizing that we can't do it so fast that we make a mess of this—there is a huge tension there and a real danger of undermining teacher confidence in the system." Snider of the Rhode Island Department of Education adds that it is "foolhardy to think that you are going to roll out a system in the first couple of years that is absolutely valid and reliable." She stresses that there are incredible consequences tied to educator evaluations, saying if a teacher as is going to be labeled as ineffective then it is imperative that "we are absolutely sure they are ineffective and that we have a preponderance of evidence to say they are ineffective." She adds

We are really clear that we can't make that claim between the 'developing' and 'effective' categories right now—it is something we are working towards. We can tolerate some looseness for now and will tighten it up over time—we just think

it is impossible to have all of those categories so well defined and implemented so thoroughly right away that you don't have some misclassifications. That was an important lesson that we learned.

The pressure, however, to implement these new systems is particularly acute in the Race to the Top states, which generally embraced ambitious timetables in order to enhance their applications for federal grants but which Jacobs of the National Council on Teacher Quality says are now being "forced to make less than ideal implementation decisions." The Gates Foundation's Tucker observes that "many state chiefs see this as an unprecedented opportunity to move the needle on reform in ways that they historically haven't been able to and so are loathe to move as slowly as their capacity tells them they should. To go slow or slower for some folks feels like a real risk."

A related challenge centers on the extent to which evaluation reforms are or are not being connected to the implementation of other reforms. Anthes of the Colorado Department of Education notes, "…we see all of this work—student standards and assessments, principal evaluations—as intricately interconnected. We need to figure out how to integrate these different pieces seamlessly so that districts don't just see it as another add on but rather see how it is all coming together." But Poda observes

The policies and emphasis right now are really only focusing on the evaluation piece in isolation…we need to think more broadly about how pieces of the entire instructional management system fit together. We're really missing the boat by not providing more feedback to teachers and helping them use the evaluations to improve instruction.

States also need to think about sequencing as several local and state education agency staff complained that the timelines for the implementation of different reforms are not well-aligned. States are moving at different speeds on different reforms and some are doing a better job than others of thinking about how and when they should connect (learning from the Gates Foundation integration grants will be very important in this regard—highlighting lessons that can be used across the state). States, for example, are taking different approaches to the introduction of new principal evaluations and new common core standards and assessments with some implementing them before the new teacher evaluations, some simultaneously, and some after. State education agencies don't have the capacity—time, energy, space, or expertise—to think through all of this right now. In addition, state flexibility around timing and sequencing of implementation is constrained

by the promises and requirements contained in their Race to the Top and No Child Left Behind waiver applications.

Value-added/growth scores for teachers in nontested subjects

Perhaps the single biggest technical challenge in implementing new teacher-evaluation systems concerns the development of teacher value-added scores that incorporate student growth measures. The National Council on Teacher Quality reports that many states do not yet have data systems with the capacity to link student test score data to individual teachers, as new evaluation laws often require.[46] In addition, the majority of teachers do not teach in tested subjects or grades and as such standardized student achievement data is not available to be used in their ratings. This is an enormous problem and it is clear that many states are struggling to address it.

A number of states are relying on school-building scores as value-added scores but this is problematic and unfair as it doesn't capture the individual teacher's contribution to student learning. Several districts are working independently to develop their own student learning objectives but the quality of the results appears to be very mixed and messy both within and across states.[47] There would seem to be an important role for state education agencies to play here but as Franklin, Pittsburgh's executive director of teacher effectiveness, cautions, "Just because districts don't have the capacity to do this work doesn't mean that states do."

Networks, policy learning, and politics

The rapid implementation of new educator-evaluation systems has been extremely difficult for state education agencies, both technically and politically. Policy learning and continuous improvement requires that local and state education agencies and the U.S. Department of Education be transparent and forthcoming about what is working and what is not, and that lessons learned be regularly shared within and between states. Tucker of the Gates Foundation remarks that a clear and important role for state education agencies is to serve as a "knowledge manager" between and among their districts, noting that there should be "some level of coordination and facilitation."

The reality in the field, however, appears to be that not enough communication and sharing of information about what works and what does not work is happen-

ing. Lopez of the Colorado Legacy Foundation, for example, says that "there is great desire but only sporadic opportunity for this" at present. This same sentiment was reiterated by others, including one state education agency official who said that "there are lots of multistate networks but no time to learn from them." Some initial efforts to create national networks are visible: the Gates Foundation has helped to facilitate information sharing among its grantees and the Department of Education has made some effort with its Race to the Top Network to bring together states to share information about their efforts to implement teacher-evaluation reforms. Snider with Rhode Island's Department of Education participated in these meetings along with key stakeholders from the state. "Being able to have a couple of days out of state without political posturing moved us ahead light years. We now call that group 'the moving forward team' and it continues to meet on its own outside of D.C. meetings and present a united front about the evaluation reforms," says Snider. These efforts, however, do not yet appear to be providing enough sustained communication or reaching enough states. One observer notes that there is a vision of creating "online resource banks" to serve this purpose but the reality is that the folks doing this work are operating on short timelines and have no time to contribute to, read, or watch these materials.

Politics and compliance issues further complicate the push for transparency and information sharing. There is a great deal of talk about how state education agencies and the U.S. Department of Education are trying to move from being compliance monitoring organizations to being service-delivery organizations but the reality is that they remain—and will always remain—both. Local education agencies understand that state education agencies have the power—and the statutory responsibility—to ensure compliance with legislative mandates and that divulging information about their implementation struggles can get them into hot water and bring sanctions. State education agencies face a similar dynamic in their dealings with the U.S. Department of Education, as candid reports of their challenges around evaluation reform may bring unwanted attention or intervention from the federal government. A state education agency leader who has taken part in the Race to the Top Network meetings, for example, recently observed that "sometimes we have been fully transparent and honest about our challenges and where we were struggling and then we got screwed when U.S. ED responded with heightened oversight and reporting requirements." Harmonizing their support and compliance monitoring functions will continue to require a delicate balancing act for state education agencies and the U.S. Department of Education, but getting the balance—and the communication— right will be crucial to the evaluation reform effort going forward.

Recommendations

The lessons derived from these challenges form the basis for the following recommendations:

- Individual states need to think carefully about the work that needs to be done to implement a new teacher-evaluation system, assess the existing capacity that is present and/or unavailable at the local and state education agency levels, and define an appropriate role for the state education agency that is commensurate with state constitutional and statutory provisions.

- Given their limited resources, state education agency leaders have to think carefully about how best to reallocate existing staff and budgets to focus on new responsibilities and build capacity and eventually bring work that is funded by external grants on-budget. Federal regulations and state budgeting and civil service requirements that constrain the ability of state education agencies to do so should be revised with an eye toward permitting greater managerial flexibility.

- State education agencies need to think about comparative advantage and economies of scale—where the state can provide something districts cannot. Providing technical assistance and policy interpretation, creating communication networks for information sharing, expanding assessment portfolios, and establishing online training modules are several areas where state education agencies could add real value.

- State legislatures and state education agencies should tailor their implementation timelines to the unique needs and resources of their particular state. They should also determine how the evaluation work ought to be sequenced with and connected to the roll out of other related education reforms, particularly those reforms around teacher preparation, professional development, principal evaluation, and common standards and assessments.

- States need to think long term about how to produce a large and stable supply of administrators—state education agency staff as well as school principals and district superintendents—with the training, technical expertise, and field experience to address their current human-capital challenges around teacher-evaluation reform. Partnering with a state's higher education system or with management consultants to devise new training and certification programs that reflect the different work and skill set required out in the field is crucial.

- The learning curve for local education agencies, state education agencies, and the U.S. Department of Education during the implementation of new teacher-evaluation systems will be steep and mistakes will inevitably be made. But it is crucial that the work be transparent and that information about what is working and what is not be shared up and down the education delivery chain. State education agencies and the U.S. Department of Education need to create a safe space where practitioners within and across states can be candid about the mistakes they are making and the support they need without fear of triggering punitive oversight or interventions by a higher authority.

Conclusion

It is clear that state education agencies are working hard to realign their organizations with the many new responsibilities that have been thrust upon them in the wake of No Child Left Behind and Race to the Top. State efforts to implement new teacher-evaluation reforms offer excellent examples of the ways that state education agencies are adapting to their new role as well as the ways in which ongoing capacity gaps continue to impede their work. Improving teacher quality has become the centerpiece of the Obama administration's education agenda and of the contemporary school reform movement. And this effort in turn, is dependent on the development of new teacher-evaluation systems with multiple measures of performance rooted in student achievement that can provide reliable data around levels of effectiveness and allow states to better support teaching and leading throughout the cycle of an educator's career from preparation to practice.

The many challenges that have already emerged also highlight the difficulty of this work and how it is further complicated by short timelines and limited state education agency staff and funding.[48] As Tucker of the Gates Foundation observes, "SEAs are making progress and doing their best in pretty difficult circumstances and high demand from districts for support but many still need to build capacity to do all that they need and want to do."

It is important to recognize that the early adopter states discussed in this report are not a random or representative sample of states—by choosing to apply for a Race to the Top grant, they selected to undertake teacher-evaluation reform and (because they won) demonstrated a greater initial ability to deliver compared to other states. As a result, states that subsequently undertake this work may well struggle more than the six states discussed here. It is hoped that this analysis has highlighted some of the key lessons and challenges in implementing new teacher-evaluation systems that have emerged from the work of the early-adopter states and that their experiences can inform the efforts of other states going forward.

About the author

Patrick McGuinn is an associate professor of political science and education and chair of the Political Science Department at Drew University. He holds a doctorate in government and a master's degree in education policy from the University of Virginia. Patrick previously held fellowships at the Institute for Advanced Study at Princeton University, the Taubman Center for Public Policy at Brown University, and the Miller Center for Public Affairs at the University of Virginia, and was a visiting scholar in the Education and Politics program at Teachers College, Columbia University. His first book, *No Child Left Behind and the Transformation of Federal Education Policy, 1965-2005* (Kansas, 2006) was honored as a *Choice* outstanding academic title. He is the editor (with Paul Manna) of *Education Governance for the 21st Century: Overcoming the Structural Barriers to School Reform* (Brookings Institution Press, 2012). His work has been published in *Perspectives on Politics, Journal of Policy History, Publius: The Journal of Federalism, The Public Interest, Teachers College Record, Educational Policy*, and *Governance*. He has contributed chapters to a number of books including: *Carrots, Sticks, and the Bully Pulpit*; *Educational Innovation and Philadelphia's School of the Future*; *Judging Bush*; *Conservatism and American Political Development*; *No Remedy Left Behind*; *Educational Entrepreneurship*; and *The Great Society and the High Tide of Liberalism*. Patrick has also produced a number of policy reports for the American Enterprise Institute, the Center for American Progress, and the Thomas B. Fordham Institute, and is a regular commentator on education in media outlets such as Education Week, *The New York Times, The Wall Street Journal*, and *The New Jersey Star Ledger*. He is a former high school social studies teacher.

Acknowledgements

Appendix

The Interviews that were a part of the research for this paper were conducted during August and September of 2012 and include the following:

Katy Anthes, executive director of educator effectiveness, Colorado Department of Education.

Diane Donohue, special assistant for educator effectiveness, Division of Teaching and Learning, Delaware Department of Education.

Carolyn Dumaresq, deputy secretary, Office of Elementary and Secretary Education, Pennsylvania Department of Education.

Michelle Exstrom, education program principal, teaching qualifications and effectiveness, National Council of State Legislatures.

Sam Franklin, executive director, Office of Teacher Effectiveness, Pittsburgh Public Schools.

Tim Gaddis, assistant superintendent for teaching, learning and assessment for Williamson County Schools. Former director of educator evaluation, Tennessee Department of Education.

Sara Heyburn, assistant commissioner, Teachers and Leaders, Tennessee Department of Education.

Sandi Jacobs, vice president and managing director of state policy, National Center on Teacher Quality.

Nina Lopez, vice president, Colorado Legacy Foundation and former commissioner's representative to State Council for Educator Effectiveness.

Tim Matheney, director of evaluation, division of teacher and leader effectiveness, New Jersey Department of Education.

Janice Poda, strategic initiative director, education workforce, Council of Chief State School Officers.

Adam Tucker, senior program officer, Bill and Melinda Gates Foundation.

Peter Shulman, chief talent officer/assistant commissioner of teacher and leader effectiveness, New Jersey Department of Education. Former director of the teacher and leader effectiveness unit at the Delaware Department of Education.

Mary Ann Snider, chief of educator excellence and instructional effectiveness, Rhode Island Department of Education.

Dan Weisberg, executive vice president and general counsel, performance management, TNTP.

Endnotes

1 For more on Race to the Top see Patrick McGuinn, "Creating Cover and Constructing Capacity: Assessing the Origins, Evolution, and Impact of Race to the Top" (Washington: American Enterprise Institute, 2010).

2 John Tyler, "Designing High Quality Evaluation Systems for High School Teachers" (Washington: Center for American Progress, 2011), available at http://www.americanprogress.org/wp-content/uploads/issues/2011/11/pdf/high_school_evaluation.pdf; See also Steven Glazerman and others, "Passing Muster: Evaluating Teacher Evaluation Systems"(Washington: Brookings Institution, 2011), available at http://www.brookings.edu/research/reports/2011/04/26-evaluating-teachers and The New Teacher Project, "Teacher Evaluation 2.0" (2010), available at http://tntp.org/ideas-and-innovations/view/teacher-evaluation-2.0.

3 D. Doyle and J.G. Han, *Measuring teacher effectiveness: A look "under the hood" of teacher evaluation in 10 sites.* (New York: 50CAN; New Haven, CT: ConnCAN; and Chapel Hill, NC: Public Impact, 2012), available at http://www.conncan.org/learn/research/teachers/measuring-teacher-effectiveness

4 Arne Duncan, "The Tennessee Story," Huffington Post, July 23, 2012.

5 Edward Crowe, "Getting Better at Teacher Preparation and State Accountability" (Washington: Center for American Progress, 2012), available at http://www.americanprogress.org/issues/education/report/2012/01/05/10941/getting-better-at-teacher-preparation-and-state-accountability/

6 See Michael Usdan and Arthur Sheekey, "States Lack Capacity for Reform," *Education Week*, May 12, 2012. available at http://www.edweek.org/ew/articles/2012/05/16/31usdan_ep.h31.html; See also, Nancy Kober and Diane Stark Rentner, *More to Do, But Less Capacity to Do It: States' Progress in Implementing the Recovery Act Education Reforms* (Washington: Center on Education Policy, 2011).

7 For more on the potential impact of teacher evaluation reform on student achievement see: Robert Gordon, Thomas Kane, and Douglas Staiger, "Identifying Effective Teachers Using Performance on the Job" (Washington: Brookings Institution, 2006).

8 For more on state implementation of No Child Left Behind see David K. Cohen and Susan L. Moffitt, *The ordeal of equality: Did federal regulation fix the schools?* (Cambridge, MA: Harvard University Press, 2009); Paul Manna, *Collision course: Federal education policy meets state and local realities* (Washington, DC: CQ Press, 2010; and Patrick McGuinn, *No child left behind and the transformation of federal education Policy, 1965-2005* (Lawrence: University Press of Kansas, 2006).

9 For more information on state teacher evaluation systems prior to Race to the Top, see Patrick McGuinn, "Ringing the Bell for K–12 Teacher Tenure Reform" (Washington: Center for American Progress, 2010), available at www.americanprogress.org/issues/2010/02/teacher_tenure_reform.html See also: Arnold Shober, "From Teacher Education to Student Progress: Teacher Quality Since NCLB" (Washington: American Enterprise Institute, 2012), available at http://www.aei.org/papers/education/k-12/teacher-policies/from-teacher-education-to-student-progress-teacher-quality-since-nclb/ and Daniel Weisberg, and others "The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness" (New York: New Teacher Project, 2009).

10 National Council on Teacher Quality, "*State of the States 2012: Teacher Effectiveness Policies*" (2012), available at http://www.nctq.org/p/publications/docs/NCTQ_State_of_the_States_2012_Teacher_Effectiveness_Policies.pdf; See also: Sara Mead, "Recent State Action on Teacher Evaluation," Bellwether Education, 2012, available at http://bellwethereducation.org/recent-state-action-on-teacher-effectiveness/

11 National Council on Teacher Quality, "*State of the States 2012: Teacher Effectiveness Policies.*"

12 Richard Elmore, "Unwarranted Intrusion," *Education Next* 2 (1) (Spring 2002); See also Kerstin Le Floch, Andrea Boyle, and Susan Bowles Therriault, "Help Wanted: State Capacity for School Improvement" (Washington: American Institutes for Research, 2008), available at http://www.air.org/expertise/index/?fa=viewContent&content_id=613.

13 Center on Education Policy, "More to Do, But Less Capacity to Do It" (2011). See also: "After the Stimulus Money Ends: The Status of State K-12 Education Funding and Reforms," and "State Education Agency Funding and Staffing in the Education Reform Era." All available at www.cep-dc.org

14 Frederick, "Louisiana Schools Chief Paul Pastorek Reflects on RTTT," Rick Hess Straight Up, August 27, 2010.

15 Sarah Sparks, "Race to the Top Winners Face Data System Challenges," *Education Week*, September 14, 2010.

16 General Accounting Office, "Race to the Top: Reform Efforts Are Under Way and Information Sharing Could Be Improved" (2010), available at http://www.gao.gov/new.items/d11658.pdf.

17 Michele McNeil, "More Race to the Top Winner Push Back Promises," Politics K-12 blog, July 13, 2011.

18 Alyson Klein, "District-State Tension an Issue in Race to the Top," Politics K–12 blog, January 5, 2010.

19 The New Teacher Project, "Smart Spending for Better Teacher Evaluation Systems" (2011), available at http://tntp.org/ideas-and-innovations/view/smart-spending-for-better-teacher-evaluations.

20 Cynthia Brown and others, "State Education Agencies as Agents of Change" (Washington: Center for American Progress, 2011), available at http://www.aei.org/papers/education/state-education-agencies-as-agents-of-change-paper/.

21 Sara Mead, Andrew Rotherham, and Rachael Brown, "The Hangover: Thinking about the Unintended Consequences of the Nation's Teacher Evaluation Binge" (Washington: American Enterprise Institute, 2012),available at http://www.aei.org/article/education/k-12/the-hangover-nations-teacher-evaluation-binge-has-unintended-consequences-release/.

22 TAP was launched in 1999 as a comprehensive school reform that according to its website "restructures and revitalizes the teaching profession by providing teachers with powerful opportunities for career advancement, ongoing professional development, a fair evaluation system and performance-based compensation." The program is run and supported by the National Institute for Excellence in Teaching. For more information see: http://www.tapsystem.org/

23 "Tennessee teacher evaluation model: Information webinar," available at http://www.tn.gov/firsttothetop/docs/Teacher_Evaluation_Information_Webinar_Powerpoint.pptx

24 See for example: Mike Morrow, "Education Committee Hears from Supporters, Critics of New Teacher Evaluation System," TN Report, November 3, 2011. available at http://tnreport.com/2011/11/03/education-committee-hears-from-supporters-critics-of-new-teacher-evaluation-system/.

25 A summary of the U.S. Education Delivery Institute report can be found here: http://www.deliveryinstitute.org/deliveryinaction/reviewing-tennessee%E2%80%99s-initial-capacity-deliver

26 State Collaborative on Reforming Education, "Supporting Effective Teaching in Tennessee: Listening and Gathering Feedback on Tennessee's Teacher Evaluation" (2012), available at http://www.tnscore.org/wp-content/uploads/2012/06/SCORE-Evaluation-Report.pdf.

27 Tennessee Department of Education, "Teacher Evaluation in Tennessee: A report on Year 1 Implementation" (2012), available at http://www.tn.gov/education/doc/yr_1_tchr_eval_rpt.pdf.

28 Ibid.

29 Ibid.

30 Tony Gonzalez, "TN Education reform Hits Bump in Teacher Evaluation," The Tennessean, July 17, 2012. available at http://educationviews.org/tn-education-reform-hits-bump-in-teacher-evaluation/.

31 A complete list of the approved instruments and information on how they were selected can be found here: http://www.nj.gov/education/EE4NJ/providers/.

32 For more on this process see: John Mooney, "School Districts Comparison Shop for Teacher Evaluation Systems," NJ Spotlight, October 17, 2012, available at http://www.njspotlight.com/stories/12/10/17/school-districts-comparison-shop-for-teacher-evaluation-systems/.

33 The contract specifies that Rutgers will communicate their progress to the NJDOE through monthly phone calls and interim reports in March, June, and September 2012. A final report is scheduled to be released by December 31, 2012.

34 Renee Kiriluk-Hill "14 Districts Band Together to Save Big Money on State Mandated Training," NJ Star Ledger, October 5, 2012, available at http://www.nj.com/hunterdon-county-democrat/index.ssf/2012/10/14_nj_school_districts_band_to.html#incart_river_default.

35 Prior to coming to New Jersey Shulman was the director of the teacher and leader effectiveness unit at the Delaware Department of Education.

36 Winnie Hu, "New Jersey Superintendents Call State Agency Ineffective," The New York Times, August 15, 2011.

37 Shulman elaborated on these lessons in a presentation to the State Board: State of New Jersey Department of Education, "Excellent Educators for New Jersey Evaluation Pilot Update: State Board Meeting" (2012), available at http://www.njspotlight.com/assets/12/0801/2127.

38 Established in 1971 by the Pennsylvania General Assembly, intermediate units operate as regional educational service agencies to Pennsylvania's 501 public school districts and serve as liaison agents between the school districts and the Pennsylvania Department of Education.

39 "Teacher and Principal Effectiveness Project," April 18, 2012. XX UNCLEAR WHAT THIS IS XX

40 "Delaware Performance Appraisal System Second Edition," available at http://www.doe.k12.de.us/csa/dpasii/files/2012DPASEvalRprt.pdf

41 Sharon Gaskins, "Partnering for Student Success: How States and School Districts Collaborate to Innovate" (Washington: U.S. Education Delivery Institute), available at http://www.deliveryinstitute.org/publications/partnering-student-success; See also: Chris Unger and others, "How Can States Education Agencies Support District Improvement? A Conversation Amongst Educational Leaders, Researchers, and Policy Actors" (Providence, Rhode Island: The Education Alliance at Brown University, 2008), available at http://www.alliance.brown.edu/pubs/csrqi/symposium.pdf.

42 National Council on Teacher Quality, "State of the States 2012: Teacher Effectiveness Policies."

43 For a detailed analysis of how SEAs are changing see Patrick Murphy and Lydia Rainey, "Modernizing the State Education Agency: Different Paths toward Performance Management" (Washington: Center on Reinventing Public Education, 2012), available at http://www.crpe.org/publications/modernizing-state-education-agency-different-paths-toward-performance-managment.

44 Mona Morshed, Chinezi Chijioke, and Michael Barber, "How the World's Most Improved School Systems Keep Getting Better" (McKinsey and Company, 2010). available at http://www.mckinsey.com/locations/madrid/recentreports/pdf/Informe_World_Improved.pdf; See also Sharon Gaskins, "Partnering for Student Success: How States and School Districts Collaborate to Innovate."

45 For a useful guide to the major questions that states need to address in planning this work see the National Governors Association and CCSSO, "Teacher and Leader Evaluation Roadmap," available at http://www.nga.org/files/live/sites/NGA/files/pdf/1112EFFECTIVENESSTLEROADMAP.PDF

46 National Council on Teacher Quality, "State of the States 2012: Teacher Effectiveness Policies."

47 See for example, Lisa Fleisher, "Teacher Grading Off to Uneven Start," The Wall Street Journal September 9, 2012, available at http://online.wsj.com/article/SB10000872396390444554704577641950548987704.html.

48 Phil Oliff and Michael Leachman, "New School Year Brings Steep Cuts in State Funding for Schools" (Washington: Center on Budget and Policy Priorities, 2011).

The Center for American Progress is a nonpartisan research and educational institute dedicated to promoting a strong, just, and free America that ensures opportunity for all. We believe that Americans are bound together by a common commitment to these values and we aspire to ensure that our national policies reflect these values. We work to find progressive and pragmatic solutions to significant domestic and international problems and develop policy proposals that foster a government that is "of the people, by the people, and for the people."

Center for American Progress