

Local control of teacher evaluation reform strains the capacity of state education agencies to support their districts and spurs some states to streamline their efforts.

by Patrick McGuinn

Remaking Teacher Evaluation: A Heavy Lift for State Education Policymakers

In 2010, the Obama administration's Race to the Top competitive grant program initiated a wave of teacher evaluation reform, which scholars and policymakers have long identified as critical to improving teacher quality and student performance.¹ State boards of education (SBEs) and state education agencies (SEAs) took different approaches to these reforms, based on the strength of their state's attachment to local control of schools and varying views of the proper role of the state in education. As a result, clear tension has emerged between some states' desires to let districts select or adapt evaluation instruments that are best suited to their particular circumstances and their SEAs' limited capacity to support the implementation of a wide array of instruments.

States that won Race to the Top grants have struggled with their teacher evaluation reforms. Media coverage in 2015 noted that while many of these states have made considerable progress in rolling out their new evaluation systems, most grantees have asked to extend the timetables for completing this work.² In addition to limited SEA capacity, another obstacle to implementation stems from the traditional focus of SEAs and SBEs on compliance and accountability activities, which has made local education agencies (LEAs) hesitant to confess whether and how they might be struggling to implement reform and reluctant to seek assistance.

As US Education Secretary Arne Duncan 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."³ In that spirit, I conducted comparative case studies on the implementation of teacher evaluation reforms in six "early adopter" states: Colorado, Delaware, New Jersey, Pennsylvania, Rhode Island, and Tennessee.⁴ Their

experiences reveal some of the key challenges and adaptations in implementing new teacher evaluation systems as well as recommendations that can inform the efforts of other states going forward.

Targeting SEA Resources

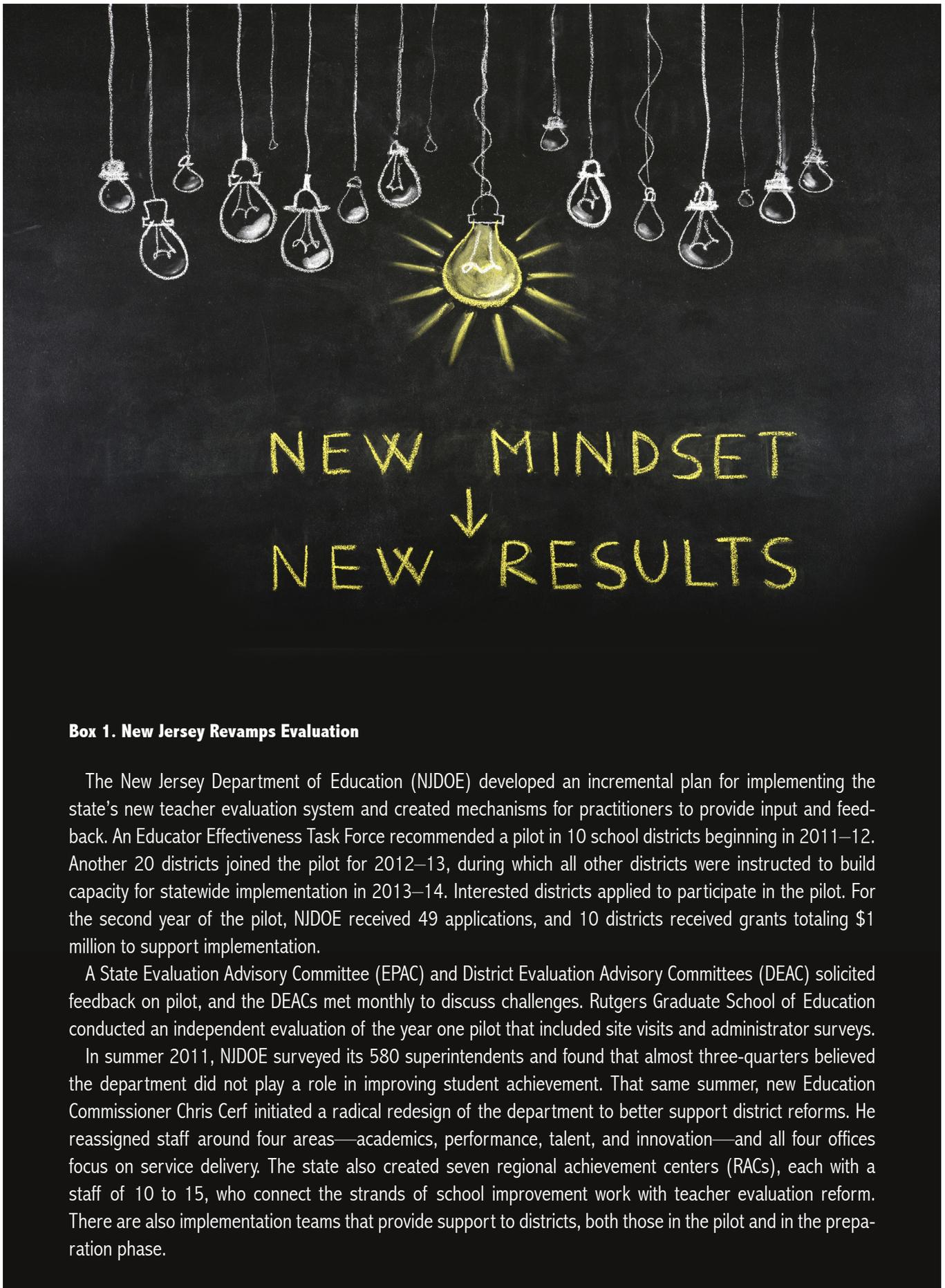
Given limited resources, state leaders have to think about how to reallocate existing SEA staff and budgets to focus on new responsibilities, build capacity, and eventually bring work that is funded by external grants on-budget. As they do so, they should consider comparative advantage and economies of scale—where the state can provide something that districts cannot. Providing technical assistance and policy interpretation, creating networks for information sharing, expanding assessment portfolios, and establishing online training modules are several areas where SEAs and SBEs could add real value.

States should reorganize their education agencies (as Tennessee and New Jersey have) around discrete functions rather than funding streams. And they should create human capital offices that can integrate the recruitment, training, evaluation, and professional development of teachers. Given the distance—literal and figurative—between SEAs and LEAs, it is important to create structures—such as New Jersey's county offices and regional achievement centers and Pennsylvania's intermediate units—to provide differentiated and targeted support on a regional basis.

Providing Evaluator Training and Certification

The Rhode Island Department of Education has developed a promising approach to providing ongoing evaluator training. Every summer it runs institutes for all evaluators: a two-day session for veteran principals and a





Box 1. New Jersey Revamps Evaluation

The New Jersey Department of Education (NJDOE) developed an incremental plan for implementing the state’s new teacher evaluation system and created mechanisms for practitioners to provide input and feedback. An Educator Effectiveness Task Force recommended a pilot in 10 school districts beginning in 2011–12. Another 20 districts joined the pilot for 2012–13, during which all other districts were instructed to build capacity for statewide implementation in 2013–14. Interested districts applied to participate in the pilot. For the second year of the pilot, NJDOE received 49 applications, and 10 districts received grants totaling \$1 million to support implementation.

A State Evaluation Advisory Committee (EPAC) and District Evaluation Advisory Committees (DEAC) solicited feedback on pilot, and the DEACs met monthly to discuss challenges. Rutgers Graduate School of Education conducted an independent evaluation of the year one pilot that included site visits and administrator surveys.

In summer 2011, NJDOE surveyed its 580 superintendents and found that almost three-quarters believed the department did not play a role in improving student achievement. That same summer, new Education Commissioner Chris Cerf initiated a radical redesign of the department to better support district reforms. He reassigned staff around four areas—academics, performance, talent, and innovation—and all four offices focus on service delivery. The state also created seven regional achievement centers (RACs), each with a staff of 10 to 15, who connect the strands of school improvement work with teacher evaluation reform. There are also implementation teams that provide support to districts, both those in the pilot and in the preparation phase.

more comprehensive four-day session for new principals and those who are new to an evaluation role. It also offers “calibration sessions” during the academic year in which a team from the department works with a district’s leadership team. The sessions focus on setting student learning objectives (SLOs), observing teachers, providing feedback, and scoring learning objectives. Evaluators pass a certification test and annual recertification tests—as they do in Tennessee—to demonstrate their readiness to conduct high-quality observations and ensure interrater reliability.

In addition to training and certification for evaluators on the front end, it is also important for state boards, SEAs, and LEAs to monitor results on the back end: Are evaluators achieving a meaningful distribution of observational scores? How well do those scores align with student achievement data? Tennessee’s SEA analyzes the data to identify schools that have a pattern of misalignment and offers them optional support in the form of a coach from the SEA.

Supporting Principals

States need to think long term about how to produce a large and stable supply of SEA staff, principals, and superintendents with the training, technical expertise, and field experience to handle teacher evaluation reform. Partnering with a state’s higher education system or management consultants to devise training and certification programs that reflect the required skill sets is crucial. As the primary evaluators, school principals will in large measure determine whether these new evaluation systems succeed. However, it is a major challenge to find time to do lengthier and more numerous evaluations, to talk with teachers about the results of the observations, and to find ways to use the observations to modify and improve instruction.

Some states have tried to redefine the role of principals, reallocating some of their current responsibilities or providing external capacity to help them. One such example is the Pennsylvania Inspired Leadership Program, a statewide, standards-based continuing professional education program for school and system leaders that focuses more than traditional programs on evaluation skills and using evaluation data to improving instruction. In Rhode

Island, some districts have hired a central office staffer to help with the evaluation work. Others have created collective bargaining agreements whereby teacher leaders can help with the observations. Colorado has established a process whereby a nonprincipal can be trained as an approved evaluation provider.

Moving from Evaluation to Coaching

Once new evaluation systems are operational, states need to ensure that the new information they provide drives personnel decisions and instructional improvement. The principal evaluation system must be aligned with the new teacher evaluation system to ensure that principals are rewarded for giving priority to assessing and coaching teachers with rigor and objectivity. Pennsylvania, for example, introduced the new Framework for Leadership a year after it launched the new teacher evaluation system. Tennessee also redesigned its principal evaluation system to better align with teacher evaluation. For evaluations to inform classroom instruction, teachers need differentiated, targeted professional development that can accommodate the wide range of academic disciplines, grade levels, student demographics, and instructional specialists (i.e., for English as a second language and special education).

Teachers and principals are being asked to use data—from student assessments and their own evaluations—to create targeted interventions that can drive improvement in student achievement. But they often are not adequately trained to accomplish this task. Creating professional learning communities among groups of educators working in the same subject and/or grade level can be very helpful, as can providing principals with professional development or coaches to assist them in understanding how to analyze and use the new data.

Centralizing Data Collection and Reporting

Data collection and reporting systems are a crucial piece of infrastructure for the new evaluations. Districts and states need such systems in order to gather, analyze, and disseminate information about teacher performance: observations, student surveys, and student growth scores. This is where scale is helpful, and statewide solutions will be more efficient and reliable

than each district reinventing the wheel. Rhode Island, for example, used its Race to the Top funds to develop the Educator Performance and Support System (EPSS), a platform that helps districts schedule evaluations and collect data and provides evaluators data at individual and aggregate levels. Colorado also built a management performance platform to help districts manage data from their evaluations. Use of the system is free but optional. As they build these systems, however, states should at the same time avoid creating student data privacy laws (such as the one enacted in Louisiana) that preclude policymakers and administrators from using the student assessment and teacher evaluation data as intended.⁵

Creating a Clearinghouse of Student Learning Objectives

Most teachers work in untested grades or subjects. Figuring out how to measure student achievement or growth in their classrooms remains perhaps the biggest problem confronting the new teacher evaluation systems. SEAs can play a productive role in identifying and designing assessments that are aligned with state learning standards. In Tennessee, for example, the department of education developed alternative growth measures that are optional for districts to use, including in world languages, physical education, health, fine arts, special education, and pre-K and kindergarten.

States vary widely in the extent to which they have created SLOs and aligned measures and in how centralized the assessment process is. Pennsylvania piloted a voluntary SLO process for districts in 2013–14 that was mandated in 2014–15. The Pennsylvania Department of Education worked with an expert to design training, resources, and templates. Pennsylvania then trained their trainers and piloted the system. The state vetted the models that came from the pilot and provided the exemplars and supporting resources to districts free of charge in 2014–15.

Engaging Stakeholders

Educators have long complained about the silos in their SEAs and district central offices and their isolation from the field. These concerns underscore the need for effective lines of communication—horizontally and vertically.

Given the interconnectedness of teacher evaluations with standards, assessment, and curriculum, state boards of education and administrators in SEAs and LEAs must ensure that these different areas are aligned.

SEAs also must be accessible to teachers and principals and answer their technical questions promptly. SEAs need to actively engage them in building, piloting, and refining the new evaluation systems. Such engagement will produce a better system and also give stakeholders ownership and buy-in in the system. New Jersey's Evaluation Pilot Advisory Committee and the evaluation advisory committees in each district appear to have been effective in this regard. Operating as they do at the top of the state education governance structure, SBEs have an important role to play in communicating with parents and teachers about what the teacher evaluation changes mean and why they are necessary.

Aligning Teacher Evaluations with New Assessments

Implementing new teacher evaluation systems is a major undertaking in its own right, but most states and districts are simultaneously rolling out the new academic standards and aligned assessments. This further strains SEA and LEA capacity and emphasizes the need to think carefully about the sequencing of rollouts of new evaluation systems with interconnected reforms.

There is a crucial role here for SBEs as they set state policy; it is imperative that core education policies are well aligned and stable over time. Teachers and administrators in the field can become disillusioned when major policies become disjointed or unexpectedly changed in the middle of being implemented. Tennessee, for example, announced that it would not implement the PARCC assessments at the end of the 2014–15 school year as planned but that it would continue to implement the Common Core State Standards. The state's standards and assessments are therefore misaligned, and educators believe they cannot be fairly evaluated on the new standards with old tests.⁶

In a Colorado program, 13 “integration districts” assign teacher leaders as “integration liaisons” to implement new evaluations using a systems thinking approach that integrates new

academic standards, assessment, and evaluations. With the help of their SEA, the district leaders met at several professional development gatherings during the year. By piloting the new teacher evaluation systems in advance of “going live” statewide, implementers have been able to identify and resolve problems that emerged and give teachers and principals time to adjust to the new system and their roles within it.

Learning from Others’ Successes and Struggles

While the design of new teacher evaluation systems varies considerably from state to state, states can learn much from one another. For this learning to happen, LEAs, SEAs, and state boards must be forthcoming about what is working and what is not. In reality, such information sharing is scant. Balancing support and compliance monitoring is a delicate balancing act for SEAs, but getting the balance—and the communication—right will be crucial to successful evaluation reform.

Conclusion

States are working hard to realign education policies, institutions, and personnel in the wake of No Child Left Behind and Race to the Top.⁷ Their efforts to reform teacher evaluation offer excellent examples of how SEAs are adapting to the new roles thrust upon them as well as ways in which ongoing capacity gaps continue to impede their work. SBEs have a vital role to play in setting the broad policies that guide this work and in helping SEAs develop the necessary capacity to implement them successfully. Alignment, consistency, transparency, and communication by SBEs will dramatically increase the odds that these new evaluation systems will improve teaching and learning.

Improving teacher quality has become the centerpiece of the Obama administration’s education agenda and of the contemporary school reform movement. The past few years have highlighted how difficult this work is and how short timelines and limited staff and funding complicate it further. In particular, states are struggling with the incorporation of student test scores into teacher ratings, how to measure student growth for teachers in nontested grades and subjects, adapting professional

development to the new evaluation process, and achieving meaningful differentiation in teacher ratings. It is important to recognize that the early adopter states discussed here are not a random or representative sample of states. By choosing to apply for a Race to the Top grant, they both self-selected into doing teacher evaluation reform and (because they won) demonstrated a greater initial ability to deliver on it compared with other states. As a result, states that subsequently undertake this work may well struggle even more than these six. But other states can benefit from a close study of the challenges the early adopters encountered in reforming teacher evaluations, and this analysis can inform their efforts going forward. ■

¹For 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, DC: American Enterprise Institute, 2010). 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, DC: Brookings Institution, 2006).

²See Michelle McNeil, “Race to Top Reports Detail Winners’ Progress, Challenges: Teacher Evaluation Puzzle Proving Difficult to Crack,” *Education Week*, March 27, 2015.

³Arne Duncan, “The Tennessee Story,” *Huffington Post* (July 23, 2012).

⁴Patrick McGuinn, “The State of Evaluation Reform: State Education Agency Capacity and the Implementation of New Teacher Evaluation Systems,” (Washington, DC: Center for American Progress, November 2012), <https://www.american-progress.org/issues/education/report/2012/11/13/44494/the-state-of-teacher-evaluation-reform/>. This research consisted of a review of the scholarly and think tank research on SEA capacity and teacher evaluation systems; analysis of reports and data from the SEA websites and from organizations such as the Council of Chief State School Officers; a study of media coverage of the reform efforts in the case study states; and 28 semistructured interviews with national experts on teacher evaluation reforms and SEA and LEA staff in each state. See also “Evaluating Progress: State Education Agencies and the Implementation of New Teacher Evaluation Systems,” Center for American Progress, 2015.

⁵Susan Dynarski, “When Guarding Student Data Endangers Valuable Research,” *New York Times*, June 13, 2015.

⁶Lauren Camera, “Tennessee Teachers Chafe at Common-Core Uncertainty: Mixed Signals Given after Years of Training,” *Education Week*, December 10, 2014.

⁷For 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, KS: University Press of Kansas, 2006).

Patrick McGuinn is associate professor of political science and education at Drew University and senior research specialist at the University of Pennsylvania’s Consortium for Policy Research in Education.