EDC Learning transforms lives.

Changing Cultures and Building Capacity: An Exploration of District Strategies for Implementation of Teacher Evaluation Systems

Karen Shakman, Nicole Breslow,
Julie Kochanek, Julie Riordan,
Tom Haferd

Learning and Teaching Division
Education Development Center, Inc.

Executive Summary

This paper explores the challenges to implementation that districts are facing and the strategies that early implementers have developed to address these challenges.

In recent years, researchers, policymakers, and practitioners have coalesced around educator evaluation as a critical lever for reforming teaching and learning. National and local policy changes have proliferated, and districts across the country are in the midst of reforming their systems for evaluating teachers. Old systems of evaluating educators, relying upon infrequent and unstandardized observations, are being replaced with more rigorous systems that include frequent observations with validated protocols, evidence of teacher practice and student outcomes, and measures of student learning. Given the fast-paced changes to educator evaluation policy and practice nationwide and what research suggests about organizational change in education, this paper explores the challenges to implementation that districts are facing and the strategies that early implementers have developed to address these challenges. Our researchers sampled 16 districts in 11 states across the country that have embarked on significant efforts to implement more complex educator evaluation systems in line with the changing policy environment. Interviews with district leaders indicated that they face two overarching challenges: (1) changing district- and school-level culture and (2) building the district- and school-level capacity to implement a new, more rigorous evaluation system. In response to these challenges, we identified nine strategies that many districts have employed to manage the change process and build capacity. Interviews from the 16 districts in our sample suggest that there are many common challenges and much is to be learned from their efforts.

Introduction

Teacher evaluation systems, though only one component of teacher performance management, are increasingly being recognized as a central component of instructional improvement. Rigorous performance measurement and useful feedback are essential in two specific ways: (1) to help teachers improve their practice and (2) to support personnel decision-making. In recent years, several high-profile studies have called attention to the shortcomings of educator evaluation systems nationwide (Gordon, Kane, and Staiger, 2006; Heneman, Milanowski, Kimball, and Odden, 2006; Toch and Rothman, 2008; Weisberg, Sexton, Mulhern, and Keeling, 2009). These studies have generated considerable debate about what makes an effective evaluation system. In particular, the authors argue that traditional means of evaluating and rewarding teachers, based on infrequent, unstandardized observations and paper qualifications such as degrees and certification, have shown minimal relationship to teachers' success with students (e.g. Toch & Rothman, 2008; Aaronson, Barrow, & Sander, 2007). In addition, the use of binary rating systems, where teachers receive an overall satisfactory/not satisfactory rating, has been criticized for lacking rigor, with nearly 99 percent of teachers in some districts earning a satisfactory rating (Weisberg, et al, 2009).

These critiques of traditional educator evaluation have spurred a movement to reform educator evaluation across the country. Fueled in part by the U.S. Department of Education's Race to the Top (RTT) program, states are rethinking their approach to educator evaluation and are making significant legislative and policy changes to support these efforts. In order to compete for the incentives promised in RTT, states were required to include provisions that evaluations would differentiate effectiveness using multiple rating categories, that a significant portion of educator evaluation ratings would be based on student assessment data, and that all teachers would receive annual evaluations (U.S. Department of Education, 2009). According to the National Conference of State Legislatures, 13 states took legislative action in 2010 to address evaluations of teacher effectiveness (NCSL, 2010). Additionally, many states that did not win RTT grants are moving forward with reforms they proposed in RTT applications, even without additional funds. Proposed reforms across the states include using other measures of teacher effectiveness in calculating ratings, such as observations of teachers, analysis of teacher artifacts and portfolios, information from peer review, student reflections and feedback, and participation in professional development (Learning Point Associates, 2010).

For many districts, state policymaking on educator evaluation systems has mandated significant reform to existing systems, while leaving much of the detail of how to measure educator effectiveness to the districts themselves. In many cases, the urgency to enact legislation in time to meet requirements for RTT accelerated a difficult and potentially contentious change process. District leaders have been left to make sense of new regulations within their district's particular context with only broad guidance from states about the details of what a new evaluation system should look like and the process for how it should be designed and implemented. Few studies have addressed the district experience in this changing policy environment. This investigation was designed to discover how districts are approaching the design and implementation of new evaluation systems and provide needed information that might help districts as they build evaluation systems that are responsive to policy, rigorous, and useful in supporting instructional improvement.

Study Design and Data Sources

This paper focuses on the processes and structures that a small number of districts have established to support the design and implementation of new teacher evaluation systems. The paper does not provide a comprehensive review of the features and components of the districts' systems themselves. In light of the early stages of development of these new systems, and the significant reform they entail, our goal was to understand how district leaders have chosen to approach the change process and to capture early successes and challenges with regard to implementation. Additionally, while many districts are working on parallel efforts to

roll out new evaluation systems for principals, this paper focuses strictly on district activities related to teacher evaluation.

We explore two questions:

1. What challenges are districts facing as they design and implement new teacher evaluation systems?

2. What strategies have districts developed to address these challenges?

To investigate these questions, we used a sample of 16 districts across the country in which significant efforts to implement more rigorous and useful educator evaluation systems have begun. We selected a variety of districts from urban, suburban, and rural locales and from a mix of eight RTT and three non-RTT states. Most of the selected districts were categorized as large urban or suburban. [1] Twelve of the districts were located in states that received an RTT award, reflecting the fact that districts with these characteristics may have additional resources that support significant reform of teacher evaluation systems. We selected districts that were at a variety of points in the implementation timeline, including some that were several years into the implementation of a new system and others that were only a few months into the process. We also sampled districts based on whether their focus on redesigning educator evaluation was entirely self-directed or whether it rested in part on a larger state reform effort.

We conducted semi-structured interviews with district representatives between September and November 2011. Interview respondents were selected by district leadership as those in the best position to discuss design and implementation challenges with the educator evaluation system. Respondents' roles in their districts included superintendent, assistant superintendent, human resource manager, and other district level leaders, such as a director of teacher effectiveness. Interviews were conducted with the understanding that the confidentiality of the respondents and the districts they represented would be respected in the published paper.

Interviews lasted 30–60 minutes and covered a range of topics, including evaluation system design, implementation challenges and strategies, and the districts' next steps in refining the evaluation systems. When possible, interviews were recorded and transcribed. In addition, interviewers maintained notes from conversations that were used for analysis. Transcripts and notes were initially coded by a team of three researchers with a list of ad hoc codes developed from the interview protocol. The coding team then identified a set of codes that emerged as key themes upon

> We conducted interviews with leaders in 16 districts that are engaged in significant efforts to implement new teacher evaluation systems.

---

[1]   Designations of urbanicity are taken from the Common Core of Data designation of public school districts in the 2010–2011 school year. Large urban districts are located in or near principal cities with populations > 250,000; small urban districts < 100,000. Large suburban districts are located outside a principal city but inside an urbanized area with populations > 250,000. Rural districts are identified as a rural territory by the U.S. Census.

a second round of reading the transcripts and notes. All materials were then recoded with the full set of codes. Researchers identified a subset of codes that were employed most frequently across all the interviews and generated a set of key themes based on the output for the subset of codes. These themes were discussed and research memos were created capturing key themes and supporting examples from across the interviews. Further discussion of key themes, with emphasis on challenges and strategies, generated another read of the output data, in which researchers searched for key terms and phrases that were not in the initial codes but emerged from the coding process as critical to understanding district experiences. This iterative, collaborative process generated a set of common challenges and strategies, as described later in this paper.

This paper aims to address a need in the field by documenting some of the shared challenges districts are facing and the strategies they are using to address these challenges. While this paper does not provide significant detail about the specific measures employed or other particularities of the systems, the district leaders we interviewed provided rich description of the change process and the challenges they have faced with regard to building stakeholder support and engagement. As one respondent suggested, as the implementation process continues, the nature of the challenges may change and may be more focused on the details of the systems. However, many of the district respondents interviewed are not yet at that phase in the reform process. That said, it is our hope that this work can be used to support and promote further learning across districts, both among those furthest along in their implementation efforts as well as those just beginning the process.

How Do Educational Systems Change?

Implementing new teacher evaluation systems requires significant changes that impact many elements of an educational system. Areas of teacher support, such as professional development, induction, and mentoring; areas related to curriculum, instruction, and assessment; and human resource areas, such as recruiting, hiring, and compensation, are all affected by changes to teacher evaluation. To understand the significant reform effort districts have undertaken in developing new educator evaluation systems, we present a brief overview of some of the key literature about how educational systems change. The systems change literature provides a useful framework for understanding the nature of educator evaluation reform and suggests potential areas of focus for successful district implementation.

In short, the literature suggests that (1) strong leaders guide with a morally compelling vision; (2) relationships are established across the district to gain the trust of key stakeholders; (3) stakeholders engage in an iterative, collaborative change process with considerable involvement throughout the system; (4) complex relationships among elements of a system are addressed by making appropriate

> The systems change literature provides a useful framework for understanding the nature of educator evaluation reform.

shifts in organizational structure and processes; (5) accountability and support structures are aligned; and (6) policymakers recognize that change takes place within a particular context and that implementation must be responsive to local needs and the specific environment. These six key themes from the change literature are outlined below.

Establish strong leadership and communicate a morally compelling vision. Systems change requires leadership that establishes a compelling policy vision that is collectively clarified over time (Fullan, Cuttress, & Kilcher, 2006; Harris, 2010; Pyhalto, Soini, Pietarinen, 2011). This vision must have a clear "moral purpose" (Fullan, 2003) and revolve around drivers for change that teachers and other stakeholder groups can believe in (Bryk & Schneider, 2002; Fullan 1999). For example, policy changes that emphasize instructional improvement and professional collaboration are more likely to engage teachers and school administrators in a change process than those that emphasize negative consequences or are focused on changes outside the instructional core (Fullan, 2010). Additionally, policy changes that are initiated by the federal or state levels of the system must be reshaped or adapted by district leaders into a vision that is relevant and responsive to the local context.

Build relationships and trust among stakeholders. The development of trust among those engaged in the change process, both internally and externally, as the process evolves is considered a key ingredient for successful implementation (Bryk, Sebring, Allensworth, Luppescu, & Easton, 2010; Kochanek, 2005; Bryk & Schneider, 2002). Lasting change requires participants to question assumptions and practices that are currently part of their work. To do so, participants must feel confident of the intentions and competence of those around them and must feel that those with whom they are engaged will respect their needs and perspective (Daly & Chrispeels, 2008). Leaders of change efforts who do not attend to building relationships find that many participants withdraw from the change process and wait for the reform initiative to pass. To even greater detriment, those most threatened by the change may stand in direct opposition and erect time- and resource-draining barriers to change efforts.

Provide stakeholders with frequent opportunities to engage with new ideas over a sustained period of time. Research on organizational change emphasizes that the change process evolves over time and requires a sustained effort. Stakeholders need multiple opportunities to receive and engage with new information and ideas and to interact with others involved in the change process, both within and outside their own system (i.e., school or district) (Fullan, 1999). Participants in the change process need help making sense of new information within their own particular context to help them internalize it (Weick, Sutcliffe, & Obstfeld, 2005; Spillane & Thompson1997). This process should involve all key stakeholder groups that play a role in the policy implementation in embedded professional development (Maurer,

1996) at the earliest stage possible. One-day workshops or weeklong trainings in isolation rarely provide the necessary conditions to sustainably influence practice. Recognizing the importance of the evolutionary nature of implementing change and the necessity of an interactive, embedded, and iterative process holds greater promise for successful change (Coburn & Stein, 2010).

Address organizational structure and processes in order to support new approaches. Successful change requires attention to the organizational systems, processes, and structures to ensure they support the intended reforms. Lack of alignment between organizational systems and the intended reform may thwart the best intentions of individuals as they seek to implement the new policy (Serafin, Bustamante, & Schramm, 2008). Therefore, a policy implementation cannot be undertaken in isolation. Rather, it must be integrated into the existing system with a review for gaps, overlaps, and inconsistencies to be addressed. Additionally, change takes place within the context of a powerful organizational culture, and the structures and processes within the organization influence that culture. As suggested above, an organization's capacity to engage in a collective reform effort depends on the presence of trusting relationships. When the organizational culture does not support these relationships, a reform effort will face additional barriers to success.

Align accountability and support structures. When those involved in a change effort are held responsible for a new or higher standard, it is essential that they be given the support they need to meet that standard (Elmore, 2002). Change cannot happen unless those who are expected to change are given the opportunity to build the capacity required to make that change. Accountability must be a reciprocal process. When improved performance is expected, there is a responsibility to make investments in developing the skills and knowledge of those who are expected to improve. Similarly, when investments in capacity are made, those who have the opportunity to gain skills and knowledge have a reciprocal responsibility to demonstrate improvements in their performance (Elmore, 2000).

Adapt change efforts to the local context. Finally, systemic change efforts must recognize that local variability is the rule, and uniformity of implementation is the exception (Coburn & Stein, 2010; Coburn, 2003). Policies that come from higher levels in the system and are intended to make change throughout the system rely on the capacity of the lower levels in the system to implement. Hence, the local context and local capacity to respond to change create unavoidable variability in how policy is enacted in practice.

These six elements of systems change provide an important context for understanding the challenges and successes districts are experiencing as they implement new teacher evaluation systems. While districts are working through the details of how to design and implement their teacher evaluation system, the challenge of system change remains central to these efforts.

Findings from Interviews with District Leaders

Two overarching challenges define the work of the district leaders we interviewed. First, districts face the challenge of changing district- and school-level culture while, at the same time, confronting anxieties associated with the new system. Second, districts face the challenge of building the district- and school-level capacity to implement a new, more rigorous evaluation system.

We identified a series of strategies that many of the districts—regardless of size, demography, or state context—have employed to respond to these challenges as they implement new evaluation systems. In the next section, we present the challenges in brief, and then describe the key strategies many of the districts we interviewed have employed in response.

Challenge 1: Changing District- and School-Level Culture

Many respondents described the challenges associated with the considerable culture shift their districts have had to make as they moved from a teacher evaluation system that was largely a pro-forma activity lacking rigor or accountability to one that demands more from all participants. New evaluation systems establish clear and specific definitions of effective teaching, where previously, a common language for discussing instruction often did not exist. As districts move toward clearer definitions of effective practice, new evaluation systems increase teacher accountability and hold teachers to a higher performance standard. The process of identifying and analyzing evidence, receiving specific feedback, and engaging in frequent conversations with an evaluator about instructional practice is a significant change for many teachers.

Resistance is a common reaction to change, and our respondents indicated that practitioners in their districts have expressed concerns about the new evaluation systems. Many teachers and administrators have expressed anxiety that the new system would hold unreasonable expectations for teachers and would emphasize the identification and elimination of ineffective teachers rather than concentrate on improving the quality of teaching in the district. As one respondent from a large urban district indicated, in the previous system, everyone was rated highly. With the new system, they've "set the bar much higher" and this requires "reorienting" teachers and principals to the new system. Another respondent from a large urban district explained that they have experienced resistance from some teachers who believe the new system sets unreasonable expectations for teachers' performance. As she explained, the new evaluation system requires a more rigorous and ongoing conversation about teaching practice between principals and their teachers, and "It's hard for principals to change the conversation with teachers they've known for years." Finally, a respondent from a large suburban district described the culture shift inherent in instituting more frequent observations of teachers. He explained,

"Typically, being observed a lot means the principal is building a case to remove you, so this is a culture shift [for teachers] to be observed regularly." District leaders we interviewed consistently referred to this challenge of changing cultural norms while building trust in the new system and engaging key stakeholders, particularly teachers, unions, and building-level administrators.

Strategy I: Focus on Effective Teaching.

The district leaders we interviewed have focused the early stages of reforming the evaluation system on developing a shared vision of effective teaching and learning. These districts have approached educator evaluation reform as an opportunity to build a districtwide dialogue about what good teaching and learning should look like in the district. Consistent with an emphasis in systems change literature on the importance of having a clear and compelling vision, district respondents indicated that this initial focus on establishing a vision has communicated to educators that the driving goal of the reform is to improve instruction. It has also addressed educator anxieties that the new evaluation system will put their jobs at risk.

The collaborative process of developing a common definition of effective instruction has provided a foundation for building the trust and engagement of teachers and principals. One respondent from a large urban district in an RTT state explained that "leading with effective teaching and support" has been the most effective strategy for making progress in building the trust and support of teachers and principals as well as for improving instruction. Getting groups together, she explained, to "have real conversations about effective practice" has been the most positive aspect of their work to date. Another respondent from a small urban district in an RTT state indicated that the work the district has already done to build a professional culture of support for improving teaching will be useful in making the shift to the new system. He explained, "One thing we've been doing for the last couple of years that helps us is that we've been going into classrooms . . . giving teachers feedback about what we see. It's not made part of their file . . . but the fact that we've been going in and keeping records, and teachers are used to having people in their rooms—[the new system] won't be a culture shock for teachers."

Strategy II: Engage Stakeholders in the Design Process.

In addition to promoting a new vision, engaging teachers and school-level leaders as partners in the design process, and inviting them to help to develop the tools that will become part of the new system, has deepened participants' engagement in the reform. Several districts have implemented design teams in which teachers, principals, and other stakeholders are developing specific aspects of the new system. This approach changes the tenor of the reform effort from something that is happening *to* teachers and principals to something they are doing themselves

"Typically, being observed a lot means the principal is building a case to remove you, so this is a culture shift [for teachers] to be observed regularly."

to improve the district. One respondent from a large urban district described the centrality of the design process in engaging key stakeholders and building critical good will and trust of teachers and principals: "The design process is at least as important as the design itself. We have been deliberate about . . . collaboration to make sure it doesn't feel like another initiative from central office. We've made a tremendous effort to engage principals and teachers and give them voice in feedback, reflection, and evolution process."

Strategy III: Involve the Teachers Union.

Beyond engaging teachers as individual contributors to the reform effort, many of the respondents described collaborative work with union representatives, and explained that the union is a critical partner in the development of the new systems. Respondents from small and large districts, representing states with RTT funding and those without, recruited union presidents and representatives to sit on the districts' steering committees and working groups. These union representatives play a key role as districts develop everything from the district's definition of effective teaching to how they intend to employ the multiple measures of teacher effectiveness. In one district, which is now in its third year of implementing a new evaluation system, the union president was an early advocate of reforming the teacher evaluation system and acted as an equal partner in the development and implementation of the new system. In another district, in which the local union leadership was not supportive of the reform effort, the superintendent went to the statewide teachers' association and collaborated with the state-level union leaders to build the new system.

Respondents explained that involving the union from the ground up in this reform effort has built good will among teachers and helped the districts to more effectively communicate with teachers about how the evaluation system will work. One respondent explained that the collaboration with the union has helped to ensure that teachers and principals hear a consistent message about the new system: "Every PowerPoint that goes out to any of those audiences . . . we work together with the union to make sure we're on the same page. If they heard my version it would be the same message as if they heard the union version. We've tried to make it so that there is a united front." The district leader indicated that reaching agreement on the final message is not always easy, but by the time the message goes out to teachers in the district, the union and the district leaders have all approved the message. Thus, just as the literature suggests, the districts' ongoing engagement with their teachers' unions as key stakeholders in the reform effort may prove critical to the sustainability of the reform over time.

"The design process is at least as important as the design itself. We have been deliberate about . . . collaboration to make sure it doesn't feel like another initiative from central office."

Strategy IV: Align Professional Development.

As some leaders' aim to move the conversation in their districts from a specific focus on evaluation to a broader emphasis on improving instruction, they are leveraging teacher evaluation reform as an opportunity to align other key aspects of their human capital systems. Specifically, some district respondents described efforts to align professional development with the new evaluation system.

A small number of the district representatives we interviewed described significant financial and material resources that their districts have invested in connecting the evaluation system to professional development. One respondent from a large urban district explained that the district examined the professional development offerings as they aligned to the new teaching framework and then spent approximately $500,000 to fill gaps in the types of professional development offered so that the district could provide teachers with the specific knowledge and skill development that a teacher's evaluation might suggest. For example, if a teacher performed poorly on one element of the framework, such as his or her ability to meet the needs of English language learners, the district wanted to be sure that the professional development offerings were available for the teacher to improve in that specific area.

Similarly, a respondent from a large urban district described how the district is able to look school by school at how teachers in a specific school are rated on the different components of their evaluations and to determine in what areas a school might benefit from additional professional development. For example, if teachers in a specific school are, on average, rated poorly on one of the items in the rubric, such as differentiating instruction, the district has the capacity to generate this information and respond with additional targeted professional development to improve teachers' practice in that area.

In addition to aligning the offerings, two large urban districts that are considered leaders in the area of teacher supervision and support are in the process of developing digital libraries of effective practices, aligned with the new evaluation rubrics. These digital resources allow teachers to view other in-district teachers exemplifying the practices described in the evaluation rubric. The digital libraries are designed to support the personalized approach to the professional development described above, such that if a teacher is weak in a particular area, the digital resources are available to help support instructional improvement in that area. As one respondent noted regarding her district's investment in professional development that was aligned to the new, more rigorous evaluation system: "with heightened accountability comes heightened support."

While our data indicated that some districts are carefully coordinating the rollout of new teacher evaluation systems with aligned professional development, other

As one respondent noted regarding her district's investment in professional development that was aligned to the new, more rigorous evaluation system: "with heightened accountability comes heightened support."

respondents in our sample said that their districts do not have the time or resources to focus on aligning professional development right now. One respondent from a large urban district explained that the district determined that if they waited to align professional development before rolling out the new evaluation system, they would never get the system up and running. Instead, the respondent suggested that the evaluation system is "shining light" on the district's weak areas and illuminating types of professional development that might be critical to improving instruction. However, the task of building the professional development offerings would take time and resources not currently at the district's disposal. Our data do suggest that whether the investment in aligning professional development is front-loaded or not, district leaders are paying attention to the relationship between evaluation and professional learning and are seeking ways to build bridges between these two elements of the system.

Strategy V: Establish Differentiated and Direct Communication Strategies.

Along with the emphasis on a new vision of effective teaching and the engagement of key stakeholders, many of our respondents stressed the importance of a strong communication strategy as an essential component of effective reform. In fact, some districts learned this lesson with early missteps from which their leaders have had to recover. In one large urban district, a survey of teachers indicated that 30–40 percent were not satisfied with the pilot of the new evaluation specifically because they did not feel they had received adequate communication about the system. As a result, the district has focused more attention on how messages about the new system get to teachers. In another district, in an effort to start gathering the three years of data necessary to effectively use student growth as a measure of teachers' performance, the district started collecting test data right away. The respondent described this as a "communication challenge" for the district as many teachers perceived this action by the district as evidence that the leadership was not following through on the goal of multiple measures of teacher performance and appeared to be focusing only on student growth data. The district has had to "step back and be more strategic and tactical in terms of taking the work forward" especially regarding how they present the goals of the evaluation system to their teachers.

In order to establish an effective communication strategy, many of our respondents indicated that the district must be aware of the various stakeholder audiences for whom messages about the new system will be crafted. One respondent spoke specifically of the need to differentiate communications for various audiences: "We have lots of different audiences and we are acutely aware of who we are speaking to." In this district, they have created two separate PowerPoint presentations for internal audiences of teachers and administrators and a third presentation for external audiences.

Furthermore, as described above, districts have found that messages about the new system are not reaching teachers, or in some instances, when the messages do get to teachers, they have been distorted. As one respondent from a large urban district explained, "[I]t's like a game of telephone," in which the message from central office is garbled by the time it finds its way to the teachers. As a result, several districts have looked for more direct contact between the central office and the teachers so they may reduce the amount of misperception and misunderstanding about the new system. One district has responded to this challenge by bringing groups of 30 teachers and principals into the central office to participate in the training rather than relying on key messages to pass directly from central office to principals and then on to teachers.

In short, districts have approached the challenge of changing the culture in surprisingly similar ways. The strategies we have identified reflect the systems change literature in that district leaders are focusing on promoting a vision of effective teaching and building the trust and engagement of stakeholders as they develop a culture that supports the new vision. The strategies outlined above are the districts' best efforts to win the trust of teachers and administrators across the districts and garner their support in embarking on significant change. At the heart of this work is a focus on effective teaching, and the efforts of districts to engage stakeholders in defining practices that reflect a collective vision of effective teaching. Then districts must communicate consistently and regularly about the goals of the reform and about the specific details and requirements of the system to ensure that educators philosophically support the reform and have the knowledge to maintain fidelity to the system. However, even when these strategies succeed in changing the culture, districts require considerable capacity to implement this vision. This brings us to a second key challenge that emerged from our findings.

Challenge 2: Build District- and School-level Capacity

In addition to the challenge of transforming the culture, the district leaders we interviewed described the significant demands their new teacher evaluation systems place on those responsible for designing, managing, and implementing them. They discussed several capacity challenges that districts face at both the school and central office levels. At the school level, teachers need support and training to understand the components of these new systems, including new definitions of effective teaching, expectations for professional practice, and processes related to the evaluation. Principals and evaluators also need to become familiar with new definitions of effective teaching and must to be able to consistently evaluate teachers using the tools provided. At the district level, the new, more complex evaluation systems require coordination and information sharing across domains, such as across the academic departments, human resources, and professional development offices, which were previously independent. New areas of expertise are also needed to develop and manage various components of the system. In short,

As one respondent from a large urban district explained, "[I]t's like a game of telephone," in which the message from central office is garbled by the time it finds its way to the teachers.

"In the past, we have had a performance management team that has worked with principals, but the focus has been primarily on when you have an unsatisfactory teacher . . . Now it's focused on all teachers with the notion of growth and supporting the growth and learning of teachers no matter where they are."

these capacity challenges indicate the need, reflected in the change literature, for changes in organizational structures and processes that support the new system over time. While districts are taking steps to build capacity and transform organizational structures at the district level, they have considerable work to do and limited resources to accomplish these changes.

Strategy VI: Invest in Training Evaluators and Teachers about the New System.

In response to these capacity-building challenges, districts are investing significant time and resources to train and support teachers and evaluators. Training on the new evaluation systems for evaluators in these districts addresses a range of needs, including helping the evaluators to understand the teaching rubric; to improve their skills in collecting evidence; to give effective formal and informal feedback; and to use data to make decisions. Additionally, some districts discussed the need to reorient principals, who act as the evaluator in many of the districts in our sample, to the purpose of the evaluation system; whereas previous systems typically focused on the lowest performing teachers, new systems require principals to think about how to support the improvement of *all* teachers. As one respondent from a large urban district in an RTT state explained, "In the past, we have had a performance management team that has worked with principals, but the focus has been primarily on when you have an unsatisfactory teacher . . . Now it's focused on all teachers with the notion of growth and supporting the growth and learning of teachers no matter where they are. We are shifting the support to how do we now create improvement plans for each teacher."

Respondents frequently discussed the challenge of establishing consistency in how evaluators are rating teachers across the district. Inter-rater reliability was the most commonly mentioned topic of professional development related to the evaluation system for evaluators. A superintendent from a large suburban district explained, we "have invested a huge amount of time and financial resources in making sure the evaluators are calibrated to a standard . . . that has been the crux of leadership training over the last two years." Respondents reported using professional development sessions for principals to view videos, participate in norming exercises, and conduct evaluator team classroom visits. A combination of face-to-face and online training formats is employed in these districts. Other strategies were also described to increase inter-rater reliability. In one district, each principal is paired with a trained external evaluator to conduct evaluations. In some districts, the state requires an evaluator certification process to ensure calibration of the evaluators's use of instruments.

In addition to evaluator training, respondents also said they spent considerable time and effort on training for teachers. Training is being offered to help teachers understand the district's definition of effective teaching and the rubric used to judge teachers' practice. Districts are also providing teachers with training on the

evaluation process and how and when to use the various tools and forms. Finally, districts are providing support to help teachers identify what student data is important and how to use it to inform instruction and professional development. As discussed earlier, the new evaluation systems represent a huge culture shift as districts focus on providing more precise feedback and differentiating practice in finer detail, and teachers' professional development in some districts has aimed at helping teachers adjust to this new culture.

Strategy VII: Establish District-Level Cross-Functional Teams.

In addition to the school-level training that a new system demands, the leaders we interviewed reported that the new teacher evaluation systems have required changes in the way the central office works. Consistent with the systems change literature's emphasis on the need for changes to organizational structure to support reforms, district-level cross-functional teams have been established in many districts to facilitate collaboration across traditional boundaries within the central office. In one large urban district, the cross-functional team includes staff from the instructional superintendents' team, human resources, chief academic officers, and data and assessment staff. A leader from another large urban district explained, "We've tried to ensure that we're all sitting at the table at the district and having conversations cross-functionally. This means that the various offices and departments at the central office are really making an effort to coordinate the work. We're not always successful, but we're making a real effort to do so."

In addition to supporting better collaboration, these cross-functional teams help to distribute the work across multiple district offices and ensure that the reform is a coherent and systemic effort. Although it can be challenging to make decisions when multiple departments are involved, with various and not always complementary interests, there are many benefits to this approach. According to a respondent from a large urban district, "We've deliberately designed work to have multiple owners across multiple parts of organization because we didn't want it to be put upon [one department] but rather to be a transformation effort. As a result, we move slower and it's harder to make decisions, but there is broader support and collective ownership."

Strategy VIII: Identify New Roles and Fill Gaps in Expertise.

As suggested above, managing new evaluation systems places significant human resource demands on district central offices. Respondents identified a variety of new roles that have been established to manage the different system components. For example, one large urban district hired a new "Director of Human Capital Strategy" who will be "growing a team" to run all of the data the new system will demand. Another small urban district hired an additional data person to focus on "pulling . . . assessment data . . . to support conversations with groups of

> "We've tried to ensure that we're all sitting at the table at the district and having conversations cross-functionally. This means that the various offices and departments at the central office are really making an effort to coordinate the work."

teachers and principals." One large urban district has hired six new positions to oversee evaluation system components, such as special education metrics, student achievement data, database management, and the appeals process. Another large urban district has created a new "talent manager" position that is responsible for using evaluation data about teacher effectiveness to create "a more holistic cycle of talent management." These talent managers are examining issues such as the relationship between teacher effectiveness ratings and teacher preparation programs and how to develop a pipeline for highly effective teachers into leadership development pathways. It is worth noting that the majority of districts (though not all) that described new hires associated with the evaluation system are districts with RTT funding. Of course, the new roles that districts have created and the ways in which these positions are structured are undoubtedly influenced by the size of the district and their financial resources, such as whether they are receiving RTT funding. As one district leader explained, "[W]e have had some support through RTT money; we wouldn't have a team internally without these resources."

In some of the districts, in addition to or in lieu of developing new positions, external partners, primarily representing a small group of nonprofit organizations, have been brought into the district to help build capacity and fill gaps in expertise. Outside partners have played various roles, including helping with tool development, training central office and/or school level staff, calculating growth or value added scores, providing a technology platform, and evaluating pilots of the system. One respondent from a large urban district described a multi-year partnership with an external organization that has embedded employees into the district as an extension of the district staff. The leader explained, this has "been huge in terms of resources and building our capacity and providing expertise . . . and has allowed us to do this [work] so fast."

While many districts have focused attention on addressing human resources needed to manage the new evaluation systems in the central office, the increased demand for school-level human resource as districts take this work to scale is proving more difficult to address. Many respondents indicated that the new systems require more of people who already have many demands on their limited time. Several districts are struggling to find the personnel needed to conduct all of the observations that the new evaluation systems require and to provide timely and effective feedback to teachers. As one district leader explained, "Principals can't do it all. Each observation takes significant time. There are pre- and post-conferences involved and time required to write notes, review the rubric, and assign scores." A leader in another district described human capital concerns that were specific to certain subject areas: "As we get into areas like the performance arts and physical education where we want to make it a performance-based assessment, we're challenged with having the manpower to take it to scale on an annual basis . . . I have less worries about infrastructure but concerns about having the human capital to pull it off."

Two of our respondents, both from large urban districts, did indicate that the creative deployment of teacher leaders to conduct some of the evaluations has helped to ease the burden on principals and other school level administrators. However, the majority of our respondents indicated that this remains a significant challenge that will require creative solutions in an era of limited resources and staffing cuts. As we heard from some districts, the challenge of principals' capacity also raises important questions as districts address the development of new principal evaluation systems alongside the new teacher evaluations. As districts design and implement new principal evaluations, attention to the principal's changing role and increased responsibilities related to teacher evaluation will be a critical piece of the design of these new systems.

Strategy IX: Sequence Implementation Activities.

Beyond additional training and staffing, districts have also been strategic about how they sequence the rollout of the new system. The careful sequencing of implementation is intended to serve two purposes: to address limitations in district capacity while at the same time engendering support by carefully selecting which elements of the new system will be most palatable in the initial phase of reform. In other words, this strategy responds to the two overarching challenges we identified earlier in this paper: it supports efforts to *change the culture* while also allowing the district to *build capacity* over time to respond to the increasing demands the new system presents.

In terms of sequencing the rollout of the new system, one common strategy was to begin implementing changes in the way observations are conducted before implementing (or completing the design of) the student data components of the system. This is beneficial from a capacity standpoint because many districts are struggling to develop or refine their approach to using student data and often need additional time to work through the challenges it poses. It has the added benefit of being strategic for building trust and engagement. As discussed above, the process of collectively defining effective teaching and creating tools for measuring instructional practice provides an opportunity for stakeholder participation. Putting this work first helps to garner support and buy-in from stakeholders, while delaying the student data component, which often meets with anxiety and resistance from teachers. Delaying the high-stakes use of student data helps practitioners stay focused on the important work of understanding the new definition of effective teaching and how it can be used to improve practice while also giving the reform some momentum. As one district leader explained, "It's been helpful that we're phasing in different elements of the system, and we've tried to be really smart about that. By delaying the student performance piece we've been able to really get people to focus on the observations and getting good at giving feedback. By rolling the system out in phases, we're making sure people don't get overwhelmed and

"By delaying the student performance piece we've been able to really get people to focus on the observations and getting good at giving feedback. By rolling the system out in phases, we're making sure people don't get overwhelmed and shut down in response to all of the changes. We're giving people a chance to really understand everything."

shut down in response to all of the changes. We're giving people a chance to really understand everything. This has been really important."

A second sequencing strategy districts are using is to initially implement new evaluation systems without consequences as a way to ease the change process and help people understand the new system. This approach also helps districts address capacity limitations, as it buys additional time to ensure the quality of the data that is being used to make personnel decisions. According to a leader in one large urban district, "There's a lot of ambiguity around how and when to use data from a pilot . . . and . . . also questions around the authenticity of that data because it's a pilot." As a result, the respondent explained that, "This year, the system is being implemented without consequences so that it was a safe way for people to get used to a new system and also recognize that the system isn't perfect yet and there is still room for feedback and continual improvement . . . There was a shared belief that if we could work collaboratively on the development of the system and delay some of the impact on decisions like dismissal, then we could build a shared confidence in the platform."

Finally, some leaders in districts we interviewed reported piloting new systems in low-performing schools as an initial phase of implementation. This strategy, although often not one the districts select themselves but rather one that is mandated by the state, allows districts to tap into additional funds to support implementation. As one respondent from a large urban district in a state with this requirement explained, "We are trying to build a bridge between where [the district] has been and the new system . . . Our plan is to spend this year as we implement this in turnaround schools as a pilot and revise in time for full implementation next year." Another large urban district's pilot includes implementation of the new evaluation system, with consequences, in schools that were identified by the state as underperforming, while implementing the system without consequences in other schools. It is not clear yet whether this strategy will be a benefit to those schools or the districts at large. However, respondents indicated that they expected to learn from the work they have done implementing these changes in a subset of their schools prior to rolling it out districtwide.

In summary, despite district leaders' strong commitment to leverage the teacher evaluation reforms as an opportunity to improve teaching and learning in their districts, it was clear from our interviews that the new evaluation systems present a real challenge for districts in terms of building the capacity to get the systems up and running. The new systems require considerable capacity at both the school and district levels. Regardless of how districts support these changes, they require significant investments of time and money to train and support teachers and administrators and to build the infrastructure they need to manage the more rigorous systems. While districts have made some strides in the early stages of implementing these new systems, including developing trainings to support the use

of the new systems and building new teams and positions to manage the demands of the systems, it is not yet clear how these changes will be supported and sustained over time.

Discussion

As we look across the responses from district leaders, we recognize many approaches indicated in the systems change research. Our data show that efforts to implement new teacher evaluation systems are consistently being approached as a system change problem, and districts are applying lessons from the literature in a variety of interesting ways. In many cases, district leaders are taking the mandated educator evaluation reform and reframing it to lead with the instructional support and improvement purpose of the system rather than its accountability purpose. Districts are focusing on engaging stakeholders in the change process, and building trust and relationships to engender the support of key constituents. They are sequencing the reform effort, rolling out elements of the system over time, and creating opportunities for stakeholders to provide feedback and input on the system as it is implemented. They are providing training and support to make the transition to the new system. Finally, they are making efforts to reorganize the system, from aligning professional development to supporting new roles and team structures. These strategies are allowing districts to successfully address some of the challenges of building trust, transforming a district culture, and responding to the new demands on teachers and administrators.

However, districts continue to face considerable challenges and have not yet identified strategies to address them all. Specifically, our respondents consistently identified three challenges that remain unresolved:

1. New systems place increased demands on principals' time, and in most cases, district leadership has not yet determined how to ensure principals have the time they need to implement the system well.

2. District leadership has not yet determined what student data to use as a measure of teacher effectiveness and have many unanswered questions about how these measures should be used in their system.

3. Some districts are experiencing a lack of alignment between their local work and the state's efforts to reform teacher evaluation, and finding that this discontinuity may interfere with the change process.

Districts have engaged stakeholders and created many new roles and cross-functional teams at the district level; however, they still face the challenge of human capacity at the school level. Although some elements of organizational change are underway, much of the entrenched systems and processes remain

untouched. The reforms to organizational culture and structure described in the change literature have primarily targeted the central office. The role of school-based administrators, in most cases those responsible for conducting the evaluations, has not changed in most contexts, despite the increasing demand of the new evaluation systems. While districts have engaged principals in the design process, many districts have not fundamentally changed the nature of principals' work to make room for the increased demands the new evaluation systems present. As one respondent, a principal from a large suburban district explained, "[I]n order to do it well, something has to come off our collective plates. That's going to be the real challenge."

> As one respondent, a principal from a large suburban district explained, "[I]n order to do it well, something has to come off our collective plates. That's going to be the real challenge."

As indicated above, the use of student data as a measure of teacher effectiveness has raised more questions for districts than they have answers. Although the postponement of the consequential use of student growth measures is a strategic decision that seems to help districts build trust and engage stakeholders, this delay may also be a result of districts' uncertainty about the student measures and how they will be employed in the educator evaluation systems. In a sentiment that was echoed by many of our respondents, one district leader said, "The student performance piece is the newest and most complicated element." While much has been written about the value of these measures, districts have many unanswered questions about how to proceed, what measures to use, and how to measure teachers who do not teach in the core tested subjects. One district leader described the challenges her district is facing in using student data: "The biggest design challenges are around student learning measures. Last year we used teacher created assessments where teachers worked with administrators to co-construct assessments and to measure student progress against these goals. We found that the capacity of teachers and administrators to design high quality assessments without significant support was limited. We also struggled with the state requirement [that measures be] comparable across grades and subjects . . . we are trying to strike the right balance between having [assessments] be rigorous and comparable across grade levels and also instructionally meaningful."

The majority of the respondents said their district was still in the early stages of grappling with questions about how to use student achievement data in their teacher evaluation systems. Three districts were somewhat further along in developing specific approaches, and two districts were in the process of piloting their student growth measures. Only one district was currently implementing student growth measures for stakes as part of their system. Districts have a long road ahead of them and will need considerable support to figure out the measures to use, develop the capacity to employ student measures, and maintain the good will they have built among teachers in the early phases of implementation.

Finally, we heard from several districts that the work they are doing is not always aligned with the work of their states, with regard to either their timeline or

approach. As a result, in some instances, they find themselves holding back from changes they may be ready to make as they await further direction from the state. One district leader explained, "One of the things we struggled with is to not move ahead of the state, but try to move forward at the same time. We don't have much guidance from the state right now . . . Timing is the real challenge here. We don't want to get far and then have to back track when guidance comes out."

In other cases, the leaders we spoke with reported feeling considerable pressure to make reforms that are rushed and perhaps not well considered in order to meet the demands of state regulations. For example, one superintendent of a large suburban district that has developed a new evaluation system ahead of the state expressed concern about the demands of the new state regulations on other districts that have not yet made the progress her district has. She explained, "My colleagues are at square one, and they have to devise, negotiate, and implement the tool this year. We are implementing a tool that was devised over a period of years . . . these superintendents are in an incredibly difficult position. Time is their enemy."

While some respondents expressed frustration with the lack of clarity of state guidance, other districts preferred that the state offer less specificity, leaving more flexibility for local interpretation. As one respondent from a large urban district explained, "Sometimes the language that comes out from [the state] isn't so helpful. When they come out with an interpretation of the [regulations] and say you need to have this form and that's not as helpful as a menu of options that districts can choose from. We believe we don't have to have everything spelled out from the state. Although there's a need for guidance, some guidance is difficult." While states are struggling with many of the same questions as districts, it may be difficult for states to stay steps ahead of their districts and provide guidance for implementation that addresses districts' different needs.

Future Directions

The districts we interviewed and the states in which they reside—in addition to many other districts and states across the United States—are breaking new ground as they envision, design, and implement new, more rigorous evaluation systems. This exploratory study was designed to learn about the bumps along the implementation road and the strategies districts are employing to traverse those bumps. The results indicate promising early practices in transforming district culture and building the human capacity and infrastructure to support the new systems. However, it is also clear that research is needed to learn about districts' design approaches and implementation challenges and strategies over time and the impact of the new systems on school and district structure and culture, instructional practices, and student outcomes. Our suggestions for future research fall into three general categories: studies of implementation, studies related to the specifics of

measurement, and studies of impact. We highlight some topics within each of these general areas below.

- Implementation Studies: These might include longitudinal studies of the fidelity of implementation, comparative studies of implementation across districts, studies of systems change and infrastructure development, and tracking the development of principal and administrator evaluations.

- Measurement Studies: These might include studies of the use of multiple measures, including reliable observation protocols; use of student growth data, including the student assessments used for measuring teacher performance and percentages assigned to student measures; and student data sources for teachers of untested subjects and grades.

- Impact Studies: These might include studies of the impact of the new evaluation systems on teacher practice and student outcomes; on school culture; on personnel decisions; and on other aspects of human capital at the school or district level.

In addition to developing a body of research, our data also suggests that those responsible for developing and sustaining new educator evaluation systems at the district-level would benefit from participating in a community of practice. While the districts in which our respondents work represent suburban, urban, and rural areas, RTT-funded and not, the district leaders we interviewed were eager to learn from other districts. In particular, as they embark on the next phase of their work, they are interested in sharing strategies and engaging in collaborative problem solving with other districts to address the complex questions about how to use student data in teacher evaluation systems. As one respondent explained, "This is a huge undertaking and we need support . . . the regulations are thoughtful, but they're complicated and there's more left to the district." Another respondent explained the district leaderships' interest in collaboration with other districts: "We're proud of what we've done and we know there's learning; the more we can engage in dialogue with others, the more it will push us. We see it as part of our role. It can be scary to be a district out in front . . . . we want to share what we've learned and learn with others."

In general, our research illustrates that districts have more in common than not, and that the opportunity for districts to collaborate and share resources would benefit all. This interest suggests an opportunity to build a diverse and dynamic community of practice, organized around substantive issues of implementation. Without such a community, we are likely to see many duplicative efforts, and many lessons learned over and over again by districts across the country. As learning is fundamentally a social process, districts' participation in a community of practice has the potential to move the field forward in a way that learning in district silos cannot (Lave and

As one respondent explained, "This is a huge undertaking and we need support . . . the regulations are thoughtful, but they're complicated and there's more left to the district."

Wenger, 1991; Wenger 1998). Establishing a structure to support districts, and representatives in various roles and at different levels of leadership, in sharing their knowledge and hard-earned wisdom would be tremendously valuable. Building a community of practice among districts, along with a parallel effort to develop a body of research that follows districts' work, would provide much needed support to districts in their efforts to implement more rigorous and useful evaluation systems and improve teaching and learning for all.

References

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, 25(1), 95–135.

Bryk, A., Bender Sebring, P., Allensworth, E., Luppescu, S., & Easton, J. Q. (2010). *Organizing schools for improvement: Lessons from Chicago.* Chicago, IL: University of Chicago Press.

Bryk, A. S. & Schneider, B. (2002). *Trust in schools: A core resource for improvement.* New York: Russell Sage.

Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher*, 32(6), 3–12.

Coburn, C., E. and M. K. Stein, eds. (2010). *Research and Practice in Education: Building Alliances, Bridging the Divide.* Lanham, MD: Rowman and Littlefield Publishers.

Daly, A. J., Chrispeels, J. H. (2008). A question of trust: Predictive conditions for adaptive and technical leadership in educational contexts. *Leadership and Policy in Schools*, 7(1), 30–63.

Elmore, R. F. (2002). Building a new structure for school leadership. Washington, DC: Albert Shanker Institute.

Elmore, R. F. (2000). *Bridging the gap between standards and achievement: The imperative for professional development in education.* Washington, D.C.: Albert Shanker Institute.

Fullan, M. (2010). *All Systems Go: The Change Imperative for Whole System Reform.* Thousand Oaks, CA: Corwin.

Fullan, M. (2003). *The moral imperative of school leadership.* Thousand Oaks, CA: Corwin.

Fullan, M. (1999). *Change forces: The sequel.* Philadelphia, PA: Falmer Press.

Fullan, M., Cuttress, C., & Kilcher, A. (2006). 8 forces for leaders of change. *National Staff Development Council*, 26(4), 54–64.

Gordon, R., Kane, T. J., & Staiger, D. O. (2006). *Identifying effective teachers using performance on the job* (Discussion Paper 2006-01). Washington, DC: Brookings Institution Press.

Harris, A. (2010). Leading system transformation. *School Leadership & Management*, 30(3), 197–207.

Heneman, H.G. III, Milanowski, A., Kimball, S.M., & Odden, A. (2006). *Standards-based teacher evaluation as a foundation for knowledge- and skill-based pay* (CPRE Policy Brief RB-45). Philadelphia, PA. Consortium for Policy Research In Education.

Kochanek, J. R. (2005). *Building trust for better schools: Research-based practices.* Thousand Oaks, CA: Corwin Press.

Lave, J., & Wenger, E. (1991). Situated learning: Legitimate peripheral participation. Cambridge, England: Cambridge University Press.

Learning Point Associates (2010). *Evaluation teacher effectiveness: Emerging trends reflected in state phase 1 Race to the Top applications.* Naperville, IL.

Maurer, R. (1996). *Beyond the wall of resistance: Unconventional strategies that build support for change.* Austin, Texas: Bard Books.

National Conference of State Legislatures. (2010). Strong leaders, strong schools: 2009 state laws. Denver, CO.

Pyhalto, K., Soini, T., & Pietarinen, J. (2011). A systemic perspective on school reform: Principals' and chief education officers' perspectives on school development. *Journal of Educational Administration*, 49(1), 46–61.

Serafin, R., Bustamante, C., & Schramm, C. (2008). *What is current practice in evaluating cross-sector partnerships for sustainable development?* (TPI Working Paper No 1/2008). London, England: The Partnering Initiative – International Business Leaders Forum.

Spillane, J., & Thompson, C. (1997). Reconstructing conceptions of local capacity: The local education agency's capacity for ambitious instructional reform. *Educational Evaluation and Policy Analysis, 19*(2), 185–203.

Toch, T. & Rothman, R. (2008). *Rush to judgment: Teacher evaluation in public education.* Washington, DC: Education Sector.

U. S. Department of Education (2009). *Race to the Top: Executive summary.* Washington, DC: U.S. Department of Education. Retrieved November 17, 2010, from http://www2.ed.gov/programs/racetothetop/index.html.

Weick, K. E., Sutcliffe, K. M. & Obstfeld, D. (2005, July-August). *Organizing and the process of sensemaking. Organization Science, 16*(4), 409–421.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness.* Brooklyn, NY: The New Teacher Project.

Wenger, E. (1998). *s* (pp. 3–15, 72–102). Cambridge, England: Cambridge University Press.