Can Evaluation Promote Teacher Development?  
Principals' Views and Experiences Implementing Observation and Feedback Cycles

Matthew A. Kraft*
Brown University

Allison Gilmour
Vanderbilt University

January 2015

Abstract

New teacher evaluation systems have expanded the role of principals as instructional leaders. We study principals’ perspectives on evaluation and their experiences implementing observation and feedback cycles. Based on interviews with a stratified random sample of 24 principals in an urban district, we find that most principals viewed professional growth as the primary purpose of evaluation. However, observing all teachers multiple times undercut the depth of feedback principals could provide and resulted in infrequent in-person conversations. Expectations to provide feedback across grade-levels and content-areas led to a narrow focus on general pedagogical practices. Principals proposed four broad solutions to these challenges: strategically targeting evaluations, reducing operational responsibilities, hiring instructional coaches, and providing principal training.

*Correspondence can be sent to Matthew Kraft at mkraft@brown.edu. We would like to thank Pam Grossman, Susan Moore Johnson, Stefanie Reinhorn, and Nicole Simon for their helpful comments on the paper.
District- and state-level efforts to remake teacher evaluation systems are among the most substantial and widely adopted reforms that U.S. public schools have experienced in decades (McGuinn, 2012; Goldhaber, 2014). These reforms were motivated in large part by research documenting that teachers have large effects on student learning (Sanders & Rivers, 1996; Rivkin, Hanushek, & Kain, 2005; Rockoff, 2004), and that existing evaluation systems were perfunctory and narrowly focused on compliance (Tucker, 1997; Weisberg, Sexton, Mulhern, & Keeligng, 2009). The Obama administration has sought to strengthen teacher quality by making teacher evaluation reforms the centerpiece of its signature education initiative, Race To The Top (RTTT), as well as state-waivers to No Child Left Behind. Today, more than 40 states have enacted new legislation aimed at strengthening and expanding teacher evaluation systems in public schools (National Council on Teacher Quality, 2013).

Research on this next generation of evaluation systems has focused overwhelmingly on policy goals, program designs, and performance measures (e.g. Kane, McCaffrey, Miller, & Staiger, 2013). However, we still know very little about how these policies are interpreted and enacted by school leaders. History clearly shows that the success of federal, state, and local policy initiatives depends on the will and capacity of local actors to implement reforms (Honig, 2006). This is particularly true in the decentralized U.S. education system where local practice is often decoupled from central policy (Spillane & Kenney, 2012).

In this paper, we study the perspectives and experiences of the local actors who are primarily responsible for implementing evaluations – school principals. School principals have been supervising and evaluating teachers for well over a century (Donaldson & Papay, forthcoming). In keeping with this tradition, many states and districts have tasked principals with responsibility for conducting observation and feedback cycles, a core feature of new evaluation
systems. Among states who applied for RTTT funds, 22 specifically identified principals, administrators, or school leaders as responsible for conducting observations in their applications, while nine referenced “trained evaluators” and the remaining eight did not specify who would conduct observations. These responsibilities commonly fall to principals when no funding is available for alternative approaches.

Relying on principals as the primary evaluators raises important questions about what they perceive as the purpose of evaluation. Some scholars (Hanushek, 2009) and journalists (Thomas, Wingert, Conant, & Register, 2010) view evaluation as a mechanism for increasing teacher effort through accountability and monitoring, and for dismissing ineffective teachers. Others see evaluation as a process that can support the professional growth of teachers by promoting self-reflection, by establishing a common language and framework for analyzing instruction, and by providing individualized feedback (Almy, 2011; Curtis & Weiner, 2012). On paper, policymakers appear to privilege this latter view; nearly every state identified professional learning as the primary purpose of evaluation reforms in their NCLB waiver applications (Center on Great Teachers and Leaders, 2014). In practice, districts often hope to promote teacher development while also using evaluations for high-stakes accountability.

Evaluation system reforms have greatly expanded the role of principals as instructional leaders. For decades, principals typically completed one-time observation check-lists and then provided carbon-copies to teachers. New systems require multiple formal and informal observations using extensive rubrics, detailed written feedback, and post-observations meetings to discuss evidence and provide sometimes critical feedback (Stronge, 2005; Danielson, 2007). Principals’ expanded responsibilities as instructional leaders raise further questions about their
capacity and ability to implement observation and feedback cycles and support teacher
development through the evaluation process.

We explored these issues by interviewing principals from a large urban school district in
the northeastern United States that recently implemented sweeping reforms to its teacher
evaluation system. We conducted interviews with 24 district principals recruited to participate
using a stratified random sampling design. Our sampling framework resulted in a collection of
principals with background characteristics and school assignments that were both diverse and
broadly representative of the district as a whole. We interviewed principals in the summer after
the first year in which the district implemented the completely redesigned teacher evaluation
system district-wide. In the first year of full-scale implementation, the district did not estimate or
use any measures of teacher effectiveness based on standardized student achievement tests. The
timing of our investigation allows us to understand principals’ experiences with the observation
portion of teacher evaluations without confounding these experiences with the controversy
surrounding standardized test-based measures of teacher effectiveness.

Our study focuses on principals’ perspectives and experiences with classroom
observation and feedback because this process is the primary mechanism through which
evaluation is intended to promote teacher development. Principals’ abilities to rate teachers
accurately, to facilitate teachers’ own self-reflection, to make specific, actionable
recommendations, and to communicate this feedback effectively are central to any evaluation
process intended to improve instruction. In our view, this paper makes several contributions to
the literature while also informing the decisions of policymakers and efforts of practitioners.
First, the paper is among the first to look inside the black box of how this next generation of
evaluations systems are perceived, operationalized, and implemented by principals. Among the
principals we spoke with, the majority viewed the primary purpose of teacher evaluation as supporting teachers to improve. However, principals’ views on evaluation did not always align with how the district articulated the purposes of evaluation or how principals’ felt the system was perceived by teachers.

Second, we characterize the concerns principals expressed about their ability to support teachers’ professional growth under the district’s current approach to implementing evaluations. Many principals described how the expanded demands to observe all teachers multiple times each year undercut the quality and depth of feedback they could provide. Principals also spoke about their lack of training and the challenge of evaluating teachers outside of their grade-level and content area expertise. This lack of training combined with the expanded demands on principals resulted in infrequent in-person conversations with teachers and feedback that was often limited to general pedagogical practices. Finally, the paper summarizes several proposals principals put forth as productive ways to improve the quality of feedback teachers receive through the observation process.

**Background**

**Policy Implementation in Education**

The success of teacher evaluation reforms, as with all policies, depends critically on how reforms are implemented (Honig, 2006). Over thirty years ago, Weatherley and Lipsky (1977) emphasized the importance of the smallest-unit of implementation, the “street-level bureaucrats” who carryout policy decisions. Ultimately it is the educators inside schools that enact policies who are responsible for a policy’s success or failure. Policymakers can promote change through pressure and support, but three other factors primarily determine the success of implementation: local capacity, context, and will (Kimball & Milanowski, 2009). Educators may respond to new
policy initiatives they view as under-resourced or unrealistic by “satisficing” – focusing on compliance rather than high-quality implementation (Halverson & Clifford, 2006). Even with the necessary resources, supports, and time, educators may lack the will to implement policies as intended due to local norms and competing priorities. For example, principals must navigate local politics and maintain the trust of their staff while implementing new high-stakes evaluations (Halverson, Kelley, Kimball, 2004). Trade-offs such as these can result in policies being transformed and adapted in different ways across local contexts.

Policy implementation is a slow process even when local capacity, context, and will are aligned for success. Policymakers hope to enact change immediately, but it often takes five years or more for schools to reach high levels of implementation fidelity (Hall & Hord, 2006). The implementation process involves an initial focus on compliance, a stage of adapting policies to local contexts, and an ongoing cycle of implementation and refinement (McLaughlin, 1987). This process can result in the enactment of policies or programs in ways that are very different from the design as originally conceived by policymakers (Elmore & McLaughlin, 1988; Spillane, Reiser, & Reimer, 2002). The polemic and personal nature of teacher evaluation combined with the resources it requires suggests principals will confront considerable challenges and difficult tradeoffs when implementing observation and feedback cycles.

**Principals’ Evolving Roles**

The role and responsibilities of school principals have evolved continually over the last century in response to shifting policy landscapes and public expectations (Spillane & Kenney, 2012). Principals are at once building managers, employers, professional figureheads, supervisors, inspirational leaders, and providers of profession development. They shape the experiences of teachers and students through these multiple interrelated roles (Hallinger & Heck,
The quality of principal leadership is a strong predictor of teacher turnover and student achievement across schools (Boyd, Grossman, Ing, Lankford, Loeb, & Wyckoff, 2011; Ladd, 2011; Johnson, Kraft, & Papay, 2012). In fact, studies find that principals are second only to teachers among all school-related factors that contribute to student learning (Leithwood, Louis, Anderson, & Wahlstrom, 2004).

Principals affect students’ learning opportunities through a combination of indirect and direct channels. They affect students indirectly by supporting and facilitating teachers’ efforts. This includes creating opportunities for teacher leadership and school-wide decision making (Spillane, Halverson, & Diamond, 2004; Johnson et al., 2014). Marshalling support for collective action among teachers also requires principals to foster mutual respect and trust among administrators and staff members (Bryk & Schneider, 2002; Bryk, Sebring, Allensworth, Luppescu, & Easton, 2010). Without relational trust, teachers may be unwilling to recognize areas for improvement and engage in a process of professional growth. In addition, principals can support teachers and students by establishing school environments that are safe, orderly, and conducive to learning. Little learning takes place when schools are chaotic places where teachers are unable to focus on instruction and students are concerned for their safety (Allensworth et al, 2009).

Over the past several decades, principals’ roles have expanded to now encompass a direct role in shaping student learning via instructional leadership (Murphy, 1990). Instructional leadership can include a range of activities involving staff development, curriculum development, student assessment and analysis, and evaluation and individualized feedback (Hoy & Hoy, 2012). At the most basic level, principals shape teachers’ learning opportunities by making choices about how to allocate time and funding for professional development. Principals
can also facilitate peer learning opportunities for teachers by developing teacher teams with clear purposes, building in common planning time, and providing opportunities for peer observations and feedback (Louis, Dretzke, & Wahlstrom, 2010). Blase and Blase (1999) found that teachers perceived principals to be effective instructional leaders when they promoted teacher reflection, supported collaboration and action research among teachers, and provided feedback to teachers.

Principals are also now increasingly expected to engage with teachers directly about classroom instruction. The limited scholarly literature on the “how” of instructional leadership compounds the challenges principals face when attempting to lead instructional improvement efforts (Neumerski, 2012). Recent studies that leverage experience sampling methods and time-use logs shed light on the evolving nature of principals’ efforts to drive instructional change. Horng, Klasik, and Loeb (2010) analyzed time use data from 65 principals in Miami-Dade County Public Schools and found that principals spent less than 6 percent of their time observing, coaching, and evaluating teachers and only 7 percent developing and delivering instructional programming. May and Supovitz’s (2011) analysis of principals’ daily activity logs and teacher surveys from 51 schools revealed that principals spent an average of 8 percent of their time on instructional leadership activities, but that this average masked considerable heterogeneity. Some principals spent no time at all on activities related to instruction, while others spent over a quarter of their time leading instructional improvement; some principals allocated their instructional feedback equally across their entire staff while others chose to work with only a few teachers. While it is clear principals are taking on expanded roles as instructional leaders, we know less about how they are managing these responsibilities or the results of their efforts.

**Principals as Evaluators**
Principals’ instructional leadership responsibilities have expanded substantially as part of recent teacher evaluation system reforms to now include working one-on-one with teachers to evaluate and improve their classroom practices. There is currently little evidence of principals’ capacity to meet these expectations. Halverson, Kelley, and Kimball’s (2004) analysis of the school-level implementation of a new standards-based observation system found that the system consumed as much as 25 percent of principals’ time and resulted in satisficing behaviors such as brief observations and positive generic feedback. The absence of formative or critical feedback in written evaluations led them to conclude that “evaluators lacked the skills to provide valuable feedback, particularly with accomplished teachers” (p. 178).

Sartain et al. (2011) studied the experiences of principals and teachers in Chicago Public Schools (CPS) that were selected to pilot a new teacher evaluation rubric and observation system. The authors found that conversations between principals and teachers were dominated by principal talk and driven by low-level questions; principals spoke about 75 percent of the time during conferences and only 10 percent of their questions were higher-order questions that pushed teachers to reflect and provide open-ended responses. Sartain and her colleagues concluded that “principals need more support in engaging in deep coaching conversations” (p. 21).

Two studies of teachers’ and principals’ perspectives on next-generation evaluation systems by Donaldson further suggest that principals face substantial capacity constraints. Donaldson (2013) found that the 30 principals in her purposive sample selected from two northeastern states lacked sufficient time to implement observations as their districts intended. In her own words, “the sheer number of teachers who needed to be observed limited [principals’] ability to provide in-depth feedback” (p. 20). In a second study, Donaldson (2012) interviewed
principals, assistant principals, and teachers from 10 purposively-selected schools about their experiences with their district’s new teacher evaluation system. Very few teachers Donaldson spoke with reported that participating in the evaluation process caused them to change their pedagogy. In fact, approximately 60 percent of the teachers said they were observed less frequently under the new system as compared to the former system. Several teachers emphasized how a mismatch between their expertise and the background of their administrator greatly limited the value of the evaluation process.

Despite these challenges, there is some evidence that evaluation systems with principals as evaluators may help improve teacher effectiveness. Steinberg and Sartain (forthcoming) exploit CPS’s randomized rollout of a new pilot evaluation system, the Excellence in Teaching Project (EITP), to estimate the causal effect of evaluation on student achievement. The authors found that the new evaluation system produced significant improvements in reading achievement and positive, but imprecisely estimated, effects in mathematics. However, the authors found no effect in either subject among the cohort of schools who adopted EITP in the second year. They hypothesized that these findings were likely explained by the large reduction in training and support for principals in the second year.

Taylor and Tyler (2012) analyzed an evaluation program in Cincinnati Public Schools in which teachers were observed by peer evaluators three times and by principals once. Peer evaluators were high-performing teachers from other schools in the district who completed intensive training on the new evaluation system and who were released from their teaching responsibilities to focus exclusively on conducting observations feedback cycles. The authors found that frequent observation and feedback cycles with expert evaluators as well as principals raised student achievement in mathematics, but found no effect on reading achievement.
Taken together, these studies suggest that there is potential for high-quality observation and feedback cycles to promote teacher development, but that it remains unclear whether principals have the time, training, and support necessary to implement these cycles effectively. We build on and contribute to this body of literature by exploring the following research questions about principals’ experiences implementing evaluation system reforms:

1) **What are principals’ views on the purpose of teacher evaluation?**

2) **How do principals balance their expanded roles as instructional leaders with their other responsibilities?**

3) **What are principals’ experiences implementing observation and feedback cycles?**

4) **Do principals feel they are able to promote professional development through the evaluation feedback they provide to teachers?**

**The Former and Current Evaluation Systems in our District**

The former evaluation system used by the district we studied was typical of those characterized in The Widget Effect report (Weisberg et al., 2009). The system stipulated that administrators should rate new teachers annually and permanent teachers biannually using a rubric with a binary rating scale: satisfactory or unsatisfactory. Teachers received ratings on eight different dimensions of professional practices as well as an overall rating. Principals were required to write an individualized improvement plan for any teachers receiving an overall rating of unsatisfactory. If the teacher failed to improve, the principal was required to write a second improvement plan and could initiate the dismissal process. Moving towards dismissal meant following a strict timeline of interim observations that could take up to two years to complete.

Studies of the former evaluation system suggest that it was more a perfunctory process than a useful tool for promoting teacher development or dismissing ineffective teachers. A survey of principals and teachers in the district found that evaluations were superficial and
infrequent; many teachers went unevaluated and schools often failed to submit the required evaluations to the district. Principals complained that the extensive checklist was too complicated with almost 20 behavioral statements and 70 indicators that did not lend themselves easily to observation or measurement. In light of these weaknesses, the district implemented a new evaluation system in 2011 that was built on the state’s new evaluation regulations and adapted for the district’s context in partnership with the local teacher’s union.

This new evaluation system currently used by the district was “designed first and foremost to promote leaders’ and teachers’ growth and development.” The current system is centered on a continuous cycle of assessment using a detailed rubric that captures measureable and observable standards related to teaching effectiveness. Teachers are active participants in the evaluation process; they initiate each cycle by self-assessing their own work and designing action plans to achieve professional practice and student learning goals. Evaluators conduct between one and four formal unannounced observations of each teacher throughout the year, depending on a teacher’s prior evaluation rating, and provide formal written feedback after each observation. In addition, evaluators are encouraged to conduct frequent informal observations lasting 15-20 minutes and hold face-to-face post-observation conversations with teachers.

Evaluators are responsible for providing teachers with a mid-year formative assessment and end-of-year summative assessment consisting of an overall rating on a four-point scale, as well as ratings on each rubric standard. Evaluators use evidence from classroom observations and artifacts submitted by teachers documenting their progress towards professional practice and student learning goals to inform their ratings. Teachers rated in the top two categories continue this cycle of self-directed growth while those in the lower rating categories are placed on more

---

1 Source redacted to protect the identity of the district.
2 Source redacted to protect the identity of the district.
structured and directed evaluation plans, which, after several repeated low evaluations, can result in dismissal. In the year leading up to the full-scale rollout of this current system, the district was explicit about its intent to shift the purpose and perception of evaluation from compliance to teacher development. Our interviews with principals focused on their perspectives and experiences implementing this current teacher evaluation system in its first year of district-wide implementation.

Research Methods

Sample

The district we studied is an urban district in the northeast that serves a racially and linguistically diverse student population. Hispanic and African-American students make up approximately three fourths of the district student body, while the remaining 25 percent of students are predominantly white and Asian. Over 70 percent of students in the district are eligible for free or reduced price lunch and nearly half speak a language other than English as their first language. We defined our target population of inference as all principals in the district that oversaw schools serving students in main-stream classes across grades K-12. This included traditional district schools, exam schools, and other semi-autonomous school types including within-district charter schools, but excluded early childhood centers, vocational and technical schools, and alternative schools for students with disabilities.

Early in the summer of 2013, we recruited a subset of 46 randomly selected principals to participate in the study in order to capture views that were broadly representative of principals across the district as a whole. In order to reduce chance sampling idiosyncrasies that might skew our results, we identified potential participants using a stratified random sampling framework. We chose two school characteristics, school size and level, on which to stratify our sample.
Specifically, we categorized all principals into 6 different strata: three school types (elementary, middle, and high) and two school sizes (390 students or more, less than 390 students). We then contacted up to nine randomly selected principals within each strata by phone and email to invite them to participate in our study, assuring them of the confidentiality of their participation.

Our sampling procedure resulted in a diverse collection of interview participants with demographic characteristics and school assignments that were broadly representative of the district as a whole. Twenty-four out of the 46 principals we contacted agreed to be interviewed, a participation rate of 52 percent. Ten of the participating principals were African-American, eight were white, two were Asian-American, two were Hispanic and two were of mixed race. Figure 1 Panel A illustrates the range of prior teaching experience among the sample. All principals except one had prior experience in the classroom with an average of just under ten years across the sample. Administrative experience varied across the sample which consisted of novice, early career, and veteran principals with an average of just over ten years of total experience as administrators. However, Figure 1 Panel B illustrates how most principals were relatively new to the schools where they currently worked. Nine of our participants were in their first or second year as principal at their current school, eight were in their third or fourth year, and seven had been at their school five years or more.

The principals we spoke with worked across the full range of school types, levels, and sizes in the district. Our sample included principals of 15 traditional district schools, six semi-autonomous schools, two exam schools, and one in-district charter school. These schools varied considerable by levels and size: five small and six large elementary schools, three small and three large middle schools, and two small and five large high schools. The student populations these schools served ranged widely and closely mirrored the distribution of student body
characteristics across all schools in the district. For example, the percentage of students scoring proficient on mathematics state exams in 4th through 8th and 10th grade ranged from as low as 16 percent to as high as 95 percent. In Figure 2 Panel A, we plot the distribution of the percentage of students scoring proficient in mathematics in our sample (solid line) and across the district as a whole (dashed line). These distributions track each other closely suggesting that our sample is broadly representative of the distribution of schools across the district. Panel B presents corresponding distributions for the percentage of students eligible for free or reduced price lunch (FRPL) and provides further evidence of how the schools in our sample map onto the full distribution across the district.

We conducted a series of t-tests to confirm that our stratified random sample of participating principals is representative of principals across the district. In Table 1, we provide the average demographic characteristics and school characteristics of all principals in the district we interviewed and those we did not. We find no statistically significant differences across any measures, strong evidence that our sample is broadly representative of the district as a whole.

**Data Collection and Analysis**

Interviews with principals lasted between 45 and 60 minutes and gave principals the opportunity to share their perspectives about teacher evaluation generally as well as their experiences implementing the districts’ former and current evaluation systems. The authors and a research assistant conducted each interview individually in person, or by phone, based on principals’ availability and preferences. We used a semi-structured protocol (see Appendix A) to ensure that each interview touched upon a common set of topics and reduced interviewer effects and bias (Patton, 2001). We audio-recorded each conversation and later transcribed the interviews to facilitate data analysis. Our research team then composed structured, thematic
summaries (Maxwell, 2005) of each interview and used these summaries to develop a set of codes that captured the common themes and topics raised by principals.

We coded interview transcripts for central concepts (Strauss & Corbin, 1998) using a hybrid approach to developing codes (Miles & Huberman, 1994). We generated codes based on our review of the principal leadership, coaching, and teacher evaluation literatures as well as common topics that were reflected in our thematic summaries. We then iteratively revised and refined our codes as new ideas emerged from the data. We analyzed our interview data by organizing codes around broad themes and reviewing interview passages associated with the codes. We then wrote analytic memos that outlined the range of perspectives and experiences that principals shared, and reviewed the characteristics of principals and their schools to situate quotes within context. Once the evidence on each theme was organized into an extended analytic memo, we returned to the interview transcripts to search for disconfirming evidence and counterexamples.

Findings

Principals’ Views of the Purpose of Teacher Evaluation

As the primary observers, principals were the face of the teacher evaluation system in the district we studied. Principals’ own perspectives on evaluation directly shaped how they chose to implement the evaluation system, and ultimately, how teachers experienced the evaluation process. We found that there existed a range of perspectives among principals about the primary purposes and value of teacher evaluation systems. We also found that principals’ views on what the district evaluation system should be used for did not always align with how the district articulated the purpose of the system or how the system was perceived by teachers.
Helping teachers improve. Among the principals we spoke with, the vast majority viewed teacher evaluation as a system that should focus on helping teachers improve their practice. This view was shared by principals with a wide range of prior teaching and administrative experience and who led schools at every level. For example, one principal described the purpose as follows:

I think it’s to get feedback to our teachers on the work that they’re doing, and how to, number one, how to make sure they know that you’re there to support them – but to also let them know where they need support and help, and then help us identify the help that they need to be better teachers.

This view was echoed by many of his colleagues who saw evaluation as a process where principals worked with teachers to identify their areas for growth and supported them to improve via direct feedback. Direct feedback “that would help them get better. Feedback that [is] specific and actionable, and that comes from a place of knowledge and experience on the part of the administrator,” as another principal explained. Other principals agreed with the overall focus on teacher improvement, but saw teacher self-reflection as the primary mechanism for improvement rather than their own feedback. “I think ultimately the goal is for teachers to self-reflect on their teaching and become better teachers and realize the areas that they need to work on as teachers,” stated an elementary school principal with 22 years of classroom experience. Although principals did not always agree on the mechanism through which the evaluation system would improve teachers, all but a few shared the belief that the primary objective was to improve teachers’ instructional practices.

Dismissing low-performing teachers. Several of the administrators we spoke with agreed that evaluation systems should support the vast majority of teachers to improve their practice, but also highlighted the importance of dismissing teachers who were ineffective
educators. One principal characterized the dual objective as “to support that teacher to become better. That would be the first goal. The second alternative, not a goal but an alternative, would be to remove that teacher from the profession.” This view was most often expressed by more experienced principals. These principals often framed the purpose of the evaluation system in terms of raising student achievement, a goal that could be accomplished via professional development and the selective dismissal of low-performing teachers. For example, a principal described the purpose of evaluation as follows:

It’s to improve teacher instruction in order to improve student achievement, to raise student achievement. That’s the purpose. If the person isn’t meeting a certain standard, then they need to be removed, because we only want the best for our students, only the best teachers in front of our students.

Not all principals agreed that the role of evaluation should be, in part, to support teacher improvement. We spoke with one principal who viewed evaluation more narrowly as a process for identifying underperforming teachers and removing them from the profession. She stated plainly, “I think the purpose of evaluations should be to weed out those that aren't doing their job.” The principal went on to describe that she invested little time evaluating teachers that were meeting her expectations and focused on evaluating out low performing teachers.

**Principals’ Views of the Implementation of Teacher Evaluation**

**Perceived purpose of the former system.** When asked about the purpose of the former binary evaluation system in their district, principals explained that although the system was intended to support teacher improvement, in practice it became a perfunctory exercise that was on rare occasions used to dismiss low-performing teachers. Principals spoke about how the evaluation system’s focus on “strict compliance” and their own selective implementation undercut the potential for the former system to support improvement. As one principal explained:
For stronger teachers, I would try to spend—I would also try to give them written feedback using the tool, but I wouldn't say that was my primary way of giving them support. I kind of then shifted to just using it as a way of evaluating teachers out or sending a very strong message to a teacher that I felt needed to improve.

The binary rating scale and focus on paperwork did not provide a system that principals found useful for supporting teacher improvement. Not surprisingly, principals largely abandoned using the former evaluation system for professional development. “If someone was strong I would evaluate them in October and never come back in [to their classroom],” admitted a middle school principal. This perception that the former evaluation system became narrowly about “get[ing] a document in” and focused on dismissal was widespread. As a third principal explained:

Improvement, that, ultimately, theoretically [was] the goal, but really it was . . . unwritten—target the teachers who were low performing and obstinate toward the school culture, and who were just bad for kids. We just needed to get them out.

With the focus of their evaluation efforts on low-performing teachers, principals perceived that teachers became wary of the evaluation process. “I think that there's a reputation from folks within the teachers' union and even some administrators too. It's like, ‘We're going to use this as a tool to terminate folks' appointment,’” explained a principal who had worked in the district for six years. This perception among teachers that the former evaluation process was focused on teacher dismissal posed a challenge to principals and the district as they transitioned to the current system.

**Perceived purpose of the current system.** According to most principals, both they and the district were working hard to shift the culture of teacher evaluation they inherited from the former system. The new evaluation system’s design helped to make this possible by providing a rubric that “engage[d] teachers around what high quality teaching looks like” and a process that directly involved teachers in the evaluation goal setting and evidence gathering process.
Principals had mixed opinions about whether these design features, and the districts’ efforts, were changing the overall culture around evaluation. One principal said, “I think there's definitely less of a feel around, this is going to be used as a tool to terminate teachers.” As another principal put it, “The new evaluation system does not have an ‘out to get you’ impression.” However, other principals characterized the current evaluation process as “still very formal” and teachers as being “still very paranoid,” and “a little bit edgy.” In the view of an elementary school principal, her staff felt the current system was still a “gotcha” system. Principals described quite positive interactions with some teachers, but for others, “once you got to the evaluation part they froze because they had had such a bad [prior] experience.” It remains unclear whether the current system will be successful at shifting teachers’ perceptions of the purpose of evaluation over time.

The Expanded Role of Principals and its Effect on Feedback

Principals experienced a variety of challenges implementing the current evaluation system as is expected when any organization rolls out a large-scale reform. These included a variety of technical challenges such as coordinating observations times, navigating the new online evaluation system, and meeting the deadlines and requirements of the current system. Principals were quick to recognize that these were transitional costs that would become less of a burden once they had developed new routines and become familiar with the new technology and requirements. However, they were much less optimistic about their ability to address the challenge that they most frequently and fervently pointed to – “the biggest challenge is time.” Principals commonly described the process of evaluating all teachers in their schools as “a nightmare” or “nuts.” As one principal shared, “It’s too much. It almost killed me to try to do all of it.”
Instructional and operational responsibilities. Principals expressed grave concern about their ability to meet the demands of the evaluation system while continuing to manage their many other responsibilities. This view was held by principals of all levels of experience who worked in both smaller and larger schools. The district evaluation plan substantially expanded the role of principals in teacher evaluation without releasing them from any of their other responsibilities. One mid-career elementary school principal likened this experience to sitting down to dinner at a family-style Italian restaurant:

It’s like going to Sorentos. Sorentos is the kind of place where they pride themselves on Italian tradition, right? Educators pride themselves on Italian tradition. That tradition is we’re going to keep piling on your plate until it falls over. We’re not going to remove anything. If you want to remove something off your plate you’d better eat it. If not, here comes the food. It keeps coming.3

Several other principals, including two principals of small elementary schools with few other administrative staff, explained that if they had dedicated themselves fully to the evaluation process “their building [would] fall apart.” “You have a lot of other things to do when you’re running a school.” A large elementary school principal asked rhetorically, “What about your buses? What about your cafeteria? What about your parents who want to meet with you? What about your district people who are calling you for this or that?” A principal of a large high school spoke about instances when the Department of Children and Family or the Police would show up at his school about a student who was removed from his home or placed in the juvenile justice system. These events required the principal’s immediate attention, and as the principal put it, “There goes that observation.” Unexpected situations required principals to be “out and about, and available.” These types of interruptions made it difficult for principals to protect the blocks

3 The name of the restaurant is a pseudonym
of time they needed to observe teachers, craft well-written evaluation feedback, and hold post-observation conferences.

**Sacrificing depth for breadth.** Several principals expressed concerns that they were unable to provide the depth of feedback they viewed as necessary for supporting teachers’ professional growth because of the sheer number of teachers they were required to evaluate.

From the perspective of one principal, if feedback cycles for improvement are “done right, it’s a weekly to monthly thing that you do with teachers.” Instead, it was all that most principals could do to observe and write the formative and summative evaluations for each teacher in their school.

The high ratio of teachers to evaluators was of particular concern for one principal:

> A leader—or in this case an instructional leader—can only be effective if the feedback and support that they provide is high quality. We know from research in the private sector that a supervisor or manager can only be effective supervising up to 12 people. Once you go beyond 12 people, you’re not able to provide the time and attention and support and feedback to those people as you can if you have 12 or fewer. I know there are some buildings where nobody is evaluating more than 12 people and then there are buildings like mine where I’m evaluating 48 people. I know there are other principals that are evaluating 30 something and 40 something teachers. I really worry about myself as an instructional leader, because am I really providing quality feedback and quality time and quality supervision to that many people? I personally don’t think so.

A principal of a large middle school expressed similar concerns. “In years past I would spend, with maybe a dozen teachers, I would spend a tremendous amount of time. I [would] sort of be very superficial with the rest. This year I was sort of deeper with 40 but not able to get nearly as deep with a few.” The infrequent evaluations and limited oversight under the former evaluation system allowed some principals to provide more in-depth feedback to the teachers they felt needed the most support.

**Limited time for feedback conversations.** Even principals who were able to hold their time dedicated to observations as “sacred” struggled to complete the feedback cycle by holding
post-observation conferences. One principal broke down the time he dedicated to the evaluation process as follows:

I would say writing it up is the majority of the time. Evaluation shouldn’t be mostly writing, but I think that I would say that it’s meeting with teachers that is probably the least amount of time. I’d say that’s probably five to ten percent of it. Observation is probably ten to 15, and then the rest is devoting to writing it.

While the exact breakdown of time varied considerably across principals, this pattern where the least amount of time was spent on in-person conversations with teachers was quite common. “The actual face-to-face conversation is not where I wanted it to be,” was a common sentiment expressed by principals with varying levels of experience.

The responsibility of drafting written evaluation feedback that was submitted via an online system and entered into a teacher’s permanent record caused principals to prioritize this step in the evaluation process. The electronic system increased the visibility and permanence of the write-up compared to the old carbon-copy evaluations that were filed away and often lost in the paper shuffle. This also served to increase the pressure on principals to draft carefully worded feedback that balanced accurate assessments with the ability to motivate teachers. An experienced middle school principal with no teaching experience explained his anxiety:

I fell into this trap where I would go in and do an observation for 20 minutes and then it would take me an hour and 20 minutes to write feedback for the teacher because I was trying to write the perfect piece of feedback where they wouldn’t be offended but they would be inspired; where it was authentic and constructive and it wasn’t judgmental; where they would follow through on what I was writing in the feedback and they wouldn’t just dismiss it as either, “He isn’t going to follow-up with me on this,” or “I disagree with him.” … I was spending no time conferencing with people.

A high school principal echoed these sentiments when she explained that, in an “ideal situation,” she would want her written and verbal feedback “to be equal.” However, the district did not
mandate in-person meetings and had no way of tracking them, making principals unaccountable for holding these meetings.

**Challenges to Implementing the Current Evaluation System**

The current evaluation system demanded a wide range of skills from principals in order to implement the evaluation process successfully. Principals were required to 1) accurately differentiate teachers on a new four point scale, 2) support their ratings with low-inference evidence, 3) communicate these ratings effectively, and 4) prescribe specific, actionable feedback for teachers on how to improve – all across a range of grade levels and subject areas.

The principals we spoke with identified three main challenges to implementing these steps successfully: their limited training, navigating difficult conversations with teachers, and providing feedback outside of their expertise. Many principals dealt with these challenges by narrowing the focus of their feedback to general pedagogical practices.

**Limited training.** In the district we studied, evaluator training was focused on familiarizing principals with the expansive rubric and logistical requirements, and calibrating principals to be reliable and accurate raters. Still, principals experienced real challenges differentiating among teachers, particularly at the upper and lower ends of the rating scale. A veteran principal of a large elementary school told us, “I think we really have a very, very fine line in between exemplary and proficient. This is the part that I have difficulty with and my APs have difficulty with.” Another experienced administrator described that he and his peers struggled with identifying “the difference between a genuinely bad teacher, who isn’t trying to improve, versus a teacher who just doesn’t have the skills in place that they need, and could improve, if they were given the right supports and feedback.” The current evaluation system
required principals to distinguish between ratings that, in the experience of some principals, required very nuanced assessments.

In addition to assigning accurate ratings, there was a critical “human component,” as one principal described it, that they had to learn on their own. “It’s an area that isn’t emphasized,” the principal lamented. A principal of a large high school explained how, under the current system, principals were expected to know how to teach adults as well as children.

The way that the role is described, the role of the principal, it says “instructional leader” and you’re told to give feedback, but I don’t think that there’s been a lot of training and resources provided on what that looks like and how to do it well, and how to do it even in challenging difficult relationships.

She had previous experience as a manager in a non-profit organization where she learned to manage people and provide feedback. For principals who transitioned into administration directly from the classroom, the only option was “learning when you get into the job,” as one principal explained. These challenges could be even greater for administrators who had no classroom teaching experience. A principal of a large high school with over 100 teachers lamented that “some of our administrators haven’t taught, so that’s a challenge.” These administrators lack of a “teaching background” and an “instructional lens” that evaluators need meant they gave “very different evaluation responses” than other member of her team.

**Difficult conversations.** The process of evaluating teachers in a way that supported their professional growth required principals to differentiate among teachers who had been told they were satisfactory for many years. Explaining ratings and communicating specific recommendations for improvement proved to be a difficult task for many of the principals we spoke with. As one administrator described, “The most difficult part of the job is probably to deliver those difficult messages, and not everyone is capable of that.” Another principal shared:
People would be crying or, ‘I can’t believe you think that. Needs improvement, I’ve never been needs improvement.’ I wanted to say, ‘Well, of course you’ve never been needs improvement, it hasn’t existed before.’

A third principal felt that some of his peers would “shy away from difficult conversations.” The focus of the evaluation process on improving teachers’ practice meant principals also had to navigate a dual role as supervisor and instructional coach. Another principal explained that her biggest challenge was, “Finding a balance” where you say to people, “I need you to do something really different from what you’ve been doing. Don’t be afraid to make mistakes. Oh, but by the way, I’m your evaluator, so I’m watching what you’re doing all the time.”

**Providing feedback outside their expertise.** The most consistent challenge principals identified was their responsibility to provide detailed and specific feedback to teachers across subjects and grade levels. Principals described how they relied on their own teaching experiences as a primary source of ideas for supporting teachers. When they evaluated teachers in subjects and grade levels they had not taught, principals felt less comfortable and confident in their abilities to evaluate instruction accurately or provide meaningful support. Elementary school principals typically characterized this challenge in terms of grade levels. A principal who taught second grade explained that his “weaker point would be the upper grades.” In order to compensate, he would often rely on two assistant principals for these evaluations. A young principal of a new elementary school explained, “I feel a little bit more comfortable in the upper grades,” as he had only taught fifth grade. A third elementary school principal who had also taught fifth grade expressed similar sentiments, “[I] feel a lot more comfortable in grades two through five . . . The kindergarten world is like a different world.”

For middle school and high school principals, evaluating teachers across different subject areas presented more of a challenge than grade-level differences. A principal with five years of
experience teaching history and English told us, “history, I do, science and math are a little bit of a challenge.” She explained that she preferred to observe math teachers with the math coach whenever possible. A high school principal laughed at the notion that she was responsible for evaluating foreign language teachers. “What do I know about Spanish and French?” she exclaimed. One middle school principal we spoke with had taught English language learners for 32 years, and stated simply, “I am not a math person.” To compensate for this, she had hired a “math interventionist” to lead instructional improvement.

**Focus on pedagogy.** Lack of content expertise led many secondary principals to narrow the focus of their evaluation to general instructional practices and strategies. A veteran high school math teacher who had just become the principal of her high school explained how she adapted her feedback across subjects. “I just find that, for myself, whenever I’m evaluating a math teacher, it’s very easy to give content suggestions, and I give pedagogy, but not content [feedback], in the other areas.” A high school principal with five years of experience said that her peers recommend a similar strategy:

> The advice that I got was to really, for content areas that I did not teach, to really focus in on just the instruction. To not worry about the content unless there was just something egregious.

Another high school principal even went as far as to focus exclusively on pedagogy in the evaluation process. As she put it, “It’s not about the subject. You know what good teaching is and it doesn’t matter what content it is.”

This focus on general pedagogical practices allowed principals to feel confident in their ability to evaluate teachers across subject areas. One principal explained that regardless of the subject, “I can walk into a class and see that there's a good delivery system; I can walk into a class, and see it's well managed.” Another principal we spoke with who had no prior teaching
experience approached evaluation by looking for general practices that he felt were beneficial for
cstudents. During observations he would ask:

How is the teacher planning to ensure all students are engaged? How is the teacher
planning to use their time wisely and to be efficient with time? How is the teacher
planning in terms of differentiating instruction? How is the teacher planning in terms of
using groups?

This principal also described how teachers at his school had raised the issue of his lack of
content expertise at a faculty meeting. His approach was to be “honest with [teachers]” that they
“are more of experts in each of the content areas than I will ever be.” Instead, he explained, he
chose to “defer to district experts” when it came to questions about implementing curriculum.

**Principals’ Proposals for Improving Evaluation Feedback to Teachers**

While principals were candid about the limitations of the current evaluation system as it
was being first implemented in the district, all principals cited meaningful ways in which the
current system was an improvement over the former binary system. Many principals felt that
transitioning from a system of infrequent evaluations with a focus on low-performing teachers to
a new system where all teachers were evaluated regularly had begun to shift the “gotcha” culture
around evaluation. Principals perceived this structural change as beginning to increase teachers’
willingness to engage with the evaluation process. Several principals also spoke positively about
the way the current system changed teachers’ role from passive recipients to active participants
in the evaluation process by requiring them to set student learning and professional practice goals
and assess their own progress.

The principals we spoke with also cited a variety of structural changes that made their
efforts to support teacher development more likely to succeed under the current evaluation
system. For example, the shift from binary checklists to rating scales with multiple categories
allowed principals to differentiate among teachers, rate them more accurately, and provide more specific feedback. Many principals appreciated the rubric’s focus on data and supporting artifacts to determine teachers’ ratings. These principals felt that ratings based on observable data helped teachers understand why they received certain feedback and how to respond to that feedback, making the evaluation process seem less subjective. In addition, several principals, who described their leadership approach as focused on school-wide goals, valued how the current rubric provided a common language to talk about improving teaching as a school community.

However, we heard time and again that placing the full responsibility of observing and coaching teachers on principals and their administrative teams would not result in major improvements in teachers’ practice without substantial changes to the implementation design. Principals warned that the amount of time they could spend with each teacher was completely insufficient, although they recommended different potential approaches to resolving this limitation. Four broad solutions to these structural challenges emerged from our conversations with principals: strategically targeting evaluations to reduce the evaluation load; relieving principals of their operational management responsibilities; hiring dedicated instructional coaches; and providing principals with more support and guidance on how to provide high-quality feedback to teachers.

**Reduce the evaluation load.** As described above, the large majority of principals we spoke with said that their multiple responsibilities prevented them from being able to dedicate the time necessary to support teachers’ improvement through evaluation feedback cycles. In their view, there needed to be a core structural change to the evaluation system if the district was really committed to improving teachers’ practice through the evaluation process. One experienced high school principal asked rhetorically, “[Do] you want . . . really good specific
evaluations, or do you want just something to cross off so that people got some sort of feedback?” Several principals highlighted the core challenge of the time needed to support the growth of teachers who were in need of improvement. One veteran middle school principal explained:

To really improve someone who’s been doing this for 10, 15 years who is mediocre, which is a big portion of [teachers at] any school. If you really want to improve them you have to spend a lot of time with them.

In their view, the current system where principals were often responsible for evaluating between 20 and 40 teachers did not allow for in-depth feedback cycles that were necessary to support meaningful improvements in teachers’ practice.

Interestingly, three different principals suggested they could not work with more than a dozen teachers at a time and be expected to make any real difference in teachers’ practices. In addition to the views of principals described earlier, one middle school principal said, “High quality implementation would’ve been me working with 12 people.” A principal of a large high school argued that the district needed to “come up with a system where they could portion off who works [with who] so that you’re not evaluating 20 people plus.”

**Reduce operational responsibilities.** A second potential solution to principals’ limited time that several principals proposed was to narrow their primary responsibilities to focus on instructional leadership. Principals commonly described instances when their efforts to focus on instructional improvement were undercut by unexpected operational issues or constrained by their other building responsibilities. “We spend a lot of time doing a lot of operations work, following up on phone calls, following up on emails; time, and time, and time again. Which pulls us away from the classroom, or having conversations with teachers,” lamented one principal. A second principal saw these operational responsibilities as directly limiting her evaluation
Transitioning between both instructional and operational responsibilities presented a real challenge for some principals. As one put it, “Fixing the bathroom and working with teachers, they’re just two very different thought processes. It’s very difficult to mix the two.”

For several principals, the solution to this challenge was clear. As an experienced teacher who was now the principal of a small elementary school explained:

If I could change anything, I would just make that my sole job, nothing else. Just to be the instructional leader, and I say my sole job, I mean to do the evaluations, but also to, to connect it back to the after-school activities we plan for families, the coaching collaborative cycles, what is going to be on the agenda for team meetings. I would just make my sole job anything that’s just instructional. Connect it back to professional development for teachers, workshops for parents. How are we going to get partners to help us? That’s it. That’s all I would want to do, nothing else. Take away all the other operational stuff.

This idea of restructuring the role of principal was shared by several others. A veteran principal told us:

If this is all I was doing, going in, observing teachers, giving feedback, working through with plans . . . it would be fantastic. It would be absolutely great if I didn’t have to deal with the operational side and the budget side.

A middle school principal also saw dividing her responsibilities among multiple administrative positions as a logical solution. “If they want the principal to be an instructional leader, taking as much of the operations out of their purview as possible is probably what needs to happen.”

**Hire instructional coaches.** When asked about the best way to improve teachers’ instructional practice, the most frequent answer principals replied with was “coaching” and “peer feedback”. These findings were consistent across all types of principals. Several principals suggested that the current evaluation system, despite its important improvements on the former one, was not implemented in a way that could affect large-scale change in teachers’ practices.
The demands to evaluate all teachers even undercut opportunities for coaching. A veteran principal described this unintended consequence. “I found that to get my evaluations done I could not spend a lot of time coaching.” He described how instead, he hired coaches to work closely with his teachers. A young principal of a large high school saw evaluation and coaching as completely separate processes.

An evaluation system is not coaching. Coaching is actually talking to someone and listening to them and responding to what they say and what you say; it's more immediate than an evaluation system. I mean sure, if you’re really, really diligent, you could be observing constantly and running back to your computer and typing up the notes and delivering them within a few minutes, and then going back to the teacher, seeing what they thought of the notes, and then writing that down in your evaluation book. That’s not realistic.

Several principals saw the need for coaches who were content experts to supplement the general instructional feedback they could provide. “I'm advocating that the district actually put together a network of content leaders . . . Let's have them also take some responsibility in evaluating depth and knowledge of content,” said a veteran high school principal. Similarly, another principal told us, “let's have some direct evaluation of real understanding of content by people who are district-wide specialists.”

**Train and coach principals.** A fourth common response we heard from principals when asked about the support they needed was that they wanted help to improve the quality of their feedback and the strategies they used to coordinate observations. Principals thought the district could do more by “providing more models of how to structure a regular meeting with teachers [and] how to lay out your calendar effectively.” Many of the principals were eager to work together or receive coaching on how to be better evaluators. “Ideally, we should be getting feedback about our feedback. That really didn’t happen this year,” said a veteran teacher and principal. “Feedback is a huge universe . . . that we should spend time thinking about, and talking
about, and learning from each other about,” another principal urged. A younger principal of a
large middle school echoed these sentiments, “I’m always interested to do a better job at
providing people feedback ... The “Good job, keep it up,” feedback doesn’t go very far, you
know? You want be more specific about teaching and teaching strategies that you can give to
them.”

Principals suggested “having mentors that will go into the classroom with you” and
videotaping their post-observation conferences to review with colleagues. A principal of a small
elementary school explained her ideal scenario:

I would love for somebody who knows teaching and learning to observe a teacher with
me, over the course of time, so it can't just be one drop in, and then figure out with me
what feedback to give that person. I would love if somebody observed me, or if I was
able to videotape a conference with a teacher and then have somebody say, “How come
you didn't push more on this?” or “Why didn't you say this?” or “This was effective when
you did this. You should try this with a different teacher.”

Principals recognized that they were being asked to develop and deliver feedback in a
way that was new and more demanding than many had experience with. Still, not every principal
wanted to improve their feedback. Some principals were more focused on using the current
system to dismiss teachers. One principal of a large elementary school saw better training for
navigating the dismissal process as the most pressing need. “We all need to know how to remove
a teacher who is unsatisfactory and you know they're hurting children. That's what we have to be
good at. That's where we need the support because that's what's going to require the time.”

Conclusion and Policy Implications

Over a quarter century ago, Popham (1988) wrote about the “dysfunctional marriage” of
formative and summative teacher evaluations. In his view, successful evaluation systems could
help teachers become more effective, or dismiss inept teachers from their positions, but not both.
Today, teacher evaluation systems are undergoing sweeping changes in order to increase their rigor and reliability for high-stakes decisions, as well as to provide teachers with actionable feedback to support improvement. It remains an open question whether these next-generation evaluation systems are capable of reconciling the marriage of teacher development and dismissal in one single system.

The urban school principals we spoke with emphasized that how an evaluation system is implemented ultimately determines whether it will be successful at achieving either of these goals. They described a variety of challenges associated with implementing observation and feedback cycles that limited their ability to promote teacher development through the evaluation process. Differing perceptions about the purpose of evaluation among principals, teachers and the district sometimes undercut the trust and buy-in required for meaningful conversations about instructional improvement. Pushing all teachers to recognize and address their own areas for improvement after most had been told they were satisfactory for many years made for challenging conversations. Many principals also described how the expanded demands to observe all teachers multiple times each year constrained the quality and depth of feedback they could provide. Expectations to provide detailed feedback to teachers outside of principals’ grade-level and content-area expertise resulted in a focus on content-free pedagogical practices. Finally, the district’s focus on compliance – submitting written evaluations for all teachers – also caused principals to prioritize written feedback over in-person conversations to discuss feedback and make improvement plans.

Principals offered several potential solutions to the design and implementation challenges they faced as the primary evaluators. One possible solution to principals’ limited capacity would be a triage system that focuses on those teachers who need the most support. However, other
principals warned that such a solution could easily erode the progress the district has made in shifting the culture around evaluation. If teachers perceive that the evaluation process is primarily used for collecting evidence to justify dismissals, they will be unlikely to engage in open conversations about how to strengthen their practice. Requiring all teachers to participate equally in a rigorous evaluation process sends a strong signal that the process is not exclusively focused on dismissal. A triage implementation strategy would require high levels of trust between administrators and teachers and full transparency about the primary purpose of evaluation if it is intended to promote teacher.

A second option several principals suggested would be to restructure the role of principals to focus less on operations management and more on instructional leadership. Research on non-traditional leadership models has found that, in practice, such approaches rarely follow these recommendations, but instead have principals share all responsibilities jointly or sub-divide responsibilities by grade levels (Grubb, & Flessa, 2006; Wexler, 2006). However, several charter school networks provide examples of co-leadership models where principals specialize in either instructional leadership or operations management (Frumkin, 2003). This type of task specialization among administrators is promising given the increasing demands on principals to be expert instructional leaders and the core importance of operations management.

A third proposal we heard was to shift the responsibility of evaluating teachers to trained instructional coaches. This would allow teachers to be matched with instructional experts in their content area, but would require substantial financial investments in a time of already tight budgets. The Peer Assistance and Review (PAR) system is one of several examples of how districts can enable their own expert teachers to conduct rigorous observations and provide detailed feedback that supports professional growth. Research shows that the PAR evaluation
process can increase teachers’ impact on students achievement (Taylor & Tyler, 2013) and can be cost effective (Papay & Johnson, 2012), but requires effective labor-management relationships and collaboration.

Finally, principals argued that if they were to maintain primary responsibility for evaluating teachers, they would need substantially more training on strategies for identifying actionable feedback for all teachers – from novices to experts across grades and subjects – as well as on how to communicate this feedback in way that causes teachers to be open and receptive. Providing additional training to principals would be relatively low cost and easy to implement compared to the other proposed solutions. However, we know little about the content and potential success of such training programs (Peterson, 2002).

The perspectives and experiences of these principals tasked with primary responsibility for evaluating teachers can inform the ongoing efforts of districts and states as they implement their own evaluation system reforms. These perspectives, however, only capture a snapshot of principals’ experiences in one district at a single point in time. Principals’ perspectives will vary depending on the nuances of the evaluation systems adopted in their districts and the district’s specific stage of implementation. It will be important for future studies to explore these potential differences across diverse contexts and implementation phases. The remaking of teacher evaluation systems across U.S. public schools has the potential to promote teacher improvement on a large scale. Delivering on this promise will depend, in large part, on how these reforms are implemented on the ground by administrators and educators. Our findings suggest that the default approach of many districts and states to use principals as the primary evaluators is unlikely to realize this promise without thoughtful strategies to address the potential limitations of this implementation approach.
References


Ready to lead, but what role will they play? Teachers’ experiences in high-poverty urban schools. *Teachers College Record, 116*(10), 1-50.


Neumerski, C. M. (2012). Rethinking instructional leadership, a review: What do we know about
principal, teacher, and coach instructional leadership and where should we go from here? 
*Educational Administration Quarterly, 49*, 310-347.


Tucker, P.D. (1997). Lake Wobegon: Where all teachers are competent (or, have we come to terms with the problem of incompetent teachers?). *Journal of Personnel Evaluation in Education, 11*(2), 103-126.


## Tables

**Table 1: Principal and School Demographic Information**

<table>
<thead>
<tr>
<th></th>
<th>Interviewed</th>
<th>Non-Interviewed</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Principals Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>0.46</td>
<td>0.39</td>
<td>0.54</td>
</tr>
<tr>
<td>White</td>
<td>0.38</td>
<td>0.44</td>
<td>0.60</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.08</td>
<td>0.16</td>
<td>0.32</td>
</tr>
<tr>
<td>Asian American</td>
<td>0.08</td>
<td>0.01</td>
<td>0.06</td>
</tr>
<tr>
<td>Male</td>
<td>0.42</td>
<td>0.28</td>
<td>0.21</td>
</tr>
<tr>
<td>Age (years)</td>
<td>47.52</td>
<td>47.21</td>
<td>0.90</td>
</tr>
<tr>
<td><strong>School Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elementary</td>
<td>0.46</td>
<td>0.41</td>
<td>0.66</td>
</tr>
<tr>
<td>Middle</td>
<td>0.13</td>
<td>0.06</td>
<td>0.27</td>
</tr>
<tr>
<td>High</td>
<td>0.17</td>
<td>0.21</td>
<td>0.65</td>
</tr>
<tr>
<td>Traditional</td>
<td>0.63</td>
<td>0.69</td>
<td>0.58</td>
</tr>
<tr>
<td>African American (%)</td>
<td>34.76</td>
<td>34.75</td>
<td>1.00</td>
</tr>
<tr>
<td>Hispanic (%)</td>
<td>41.47</td>
<td>44.46</td>
<td>0.48</td>
</tr>
<tr>
<td>White (%)</td>
<td>11.54</td>
<td>12.46</td>
<td>0.76</td>
</tr>
<tr>
<td>Asian (%)</td>
<td>10.05</td>
<td>5.52</td>
<td>0.06</td>
</tr>
<tr>
<td>Independent Education Plans (%)</td>
<td>17.03</td>
<td>19.12</td>
<td>0.18</td>
</tr>
<tr>
<td>English Language Learners (%)</td>
<td>29.00</td>
<td>29.55</td>
<td>0.89</td>
</tr>
<tr>
<td>Low Income (%)</td>
<td>70.06</td>
<td>71.02</td>
<td>0.77</td>
</tr>
<tr>
<td>Proficient in English language arts (%)</td>
<td>49.29</td>
<td>46.99</td>
<td>0.64</td>
</tr>
<tr>
<td>Proficient in mathematics (%)</td>
<td>42.57</td>
<td>41.80</td>
<td>0.86</td>
</tr>
<tr>
<td>Observations</td>
<td>24</td>
<td>86</td>
<td></td>
</tr>
</tbody>
</table>

Notes: P-values are derived from two-sample t-tests of the mean difference in a given characteristic across interviewed and non-interviewed principals. Proportions of schools that are elementary, middle, and high school do not sum to one because of schools with non-traditional grade configurations.
Figures

Panel A

Panel B

Figure 1: Histograms depicting distributions of the total number of years of classroom teaching experience and total number of years of administrative experience at current schools for interviewed principals.
Figure 2: Distributions of the percent of students who are proficient in mathematics (Panel A) and who are low-income (Panel B) across the full target population of schools in the district and schools represented in the interview sample.
Appendix

Appendix A: Interview Protocol

Narrative about Research Project and Framework of Interview (Read to Interviewee):

Hi my name is XXXXXX and I’m a member of a research team from Brown and Vanderbilt studying the experiences of principals in implementing new evaluation systems. We are interested in your opinions about, and experiences with, the new Educator Evaluator System in BPS. I’ll ask you a series of questions meant to give you the opportunity to share your thoughts about the transition from the old to the new Educator Evaluation System. We particularly hope to learn about whether this change has made a difference in your work. We are also interested in how you decide which ratings to give to teachers under this new system and whether/how the new system supports professional growth and development among teachers. The interview should last approximately 50 minutes.

The information you share is completely confidential. No individuals or schools will be identified in any written reports or presentations. This information will be the basis of a scholarly article and a set of recommendations we provide to the BPS Office of Educator Effectiveness on how to improve the Educator Evaluation System.

I would like to record the conversation so I can focus on what we discuss rather than taking detailed notes, is that ok with you?

Personal & School Background:

Step 1: Briefly review the information on the demographic questionnaire to be sure it is correct.

1. What makes your school unique compared to other schools in BPS?
2. What is the biggest challenge you face as a principal at your school?

Evaluation Background:
1. Were you responsible for evaluating teachers under the old Educator Evaluation System? If yes . . .
2. What do you think was the primary purpose of teacher evaluation under the old system?
3. What did you view as the strengths and weakness of this old system?

Current Evaluation System:
1. Do you think the primary purpose of teacher evaluation has changed under the new Educator Evaluation System? If so, how and why?
2. What are the strengths and weaknesses of the new evaluation system?
3. What are the opportunities and challenges associated with being both a supervisor and instructional leader/coach as part of the new Educator Evaluation System?
4. In your experience, does your relationship with a teacher affect how you deliver feedback and what feedback you provide? If so, can you please provide an example?
5. Are there certain grades or subjects in which you feel more comfortable evaluating teachers? If so, why . . .

6. Research studies suggest that teachers receive positive evaluations even when their performance is unsatisfactory or in need of improvement.
   a. Did this happen in Boston under the old Evaluation System? If so, why . . ?
   b. Does this happen in Boston under the new Evaluation System? If so, why . . ?

7. In your experience, does the new Educator Evaluation System make it easier or more difficult to rate a teacher as unsatisfactory (or needs improvement)? Please explain . .

8. Are there ever situations when you were unable or unwilling to give a low rating? Can you give an example?

9. Were any teachers at your school rated as unsatisfactory? If so, why?

10. Were any teachers at your school rated as needs improvement? How do you use this rating?

11. Did rating a teacher as unsatisfactory or needs improvement affect your relationship with other teachers in the school? If so, how?

12. What proportion of your time completing evaluations do you spend observing & collecting data vs. writing & entering in feedback, vs. meeting with teachers to discuss feedback?

13. Do teachers rated as proficient or exemplary receive the same amount and type of feedback (written vs. in person) as those rated as unsatisfactory or needs improvement?

14. In your experience, does the feedback teachers receive via the evaluation process help teachers improve their practice? How? Please provide a specific example.

15. Research suggests that teachers can be reluctant to act on the feedback they receive, what strategies do you use to communicate feedback effectively and build teacher’s buy-in?

Evaluation & Improvement:

1. What do you think the primary purpose of teacher evaluation systems should be?

2. What training and support would be most useful to you to help you improve your ability to provide feedback to teachers about how to improve their practice?

3. If you could change anything about the Educator Evaluation System, how would you change it?

4. What do you think is the best way to improve instruction at your school?

Closing Question:

1. Are there any other issues or points you would like to raise before we conclude the interview?