Can Teacher Evaluation Provide Both Accountability and Development? Learning from Six Schools' Implementation of Evaluation Policy

Stefanie Reinhorn and Susan Moore Johnson

Project on the Next Generation of Teachers

Harvard Graduate School of Education

October 13, 2015

Abstract

Nationwide, schools are implementing ambitious evaluation policies meant to improve education by holding teachers accountable and supporting their development. Seeking to learn whether and how evaluation policy could serve both purposes, we explored implementation of standards-based evaluation in six high-poverty schools of one urban district. Teachers widely said they wanted evaluation to provide both accountability and development. However, only one principal used evaluation to achieve both purposes. Four principals responded perfunctorily to the policy and achieved neither purpose. The sixth principal used evaluation primarily to dismiss teachers. We examine differences in implementation and their implications for policy, practice, and research.

Nationwide, state education agencies, school systems, and individual schools are hard at work to reform teacher evaluation, yet debate persists about the purposes and goals that should guide the process. Some analysts argue that evaluation should, first and foremost, support professional growth for teachers (Almy & Education Trust, 2011; Curtis &Weiner, 2012). However, in the academic and popular press, others staunchly recommend that evaluation should serve primarily as an assessment and accountability tool. For example, Hanushek (2009) recommends that school officials use evaluation to identify and "deselect" (dismiss) the least effective teachers (p. 177), a stance shared by Hannaway (2009). A 2010 *Newsweek* cover similarly asserts, "The Key To Saving American Education: We Must Fire Bad Teachers." Its accompanying article denounces the decline of public education as a "national embarrassment" that should be addressed by dismissing weak teachers (Thomas, Wingert, Conant, & Register, 2010).

Goals of development and accountability in teacher evaluation often stand in tension, even though policies governing evaluation typically call for achieving both. Papay (2012) suggests, "With nearly three million teachers in the United States, rapid improvements in instructional effectiveness will not be possible by simply replacing low-performing teachers. . . . Instead, for evaluation to realize its potential as widespread instructional reform, it must work to raise the performance of all teachers" (p. 138). Further, he argues that in order for evaluation to support "transformational change"(p.125) policy makers, researchers and reformers need to refocus their attention on ongoing teacher development. Researchers from The Measures of Effective Teaching [MET] Study agree, asserting that "the real work lies ahead: understanding how to use that data to help all teachers improve their practice and the outcomes for America's young people" (Bill and Melinda Gates Foundation, 2013b, p.8).

Studies support the potential payoff of investing in teachers' learning over the course of the "career continuum" (Feinman-Nemser, 2001). Researchers (Kraft & Papay, 2014; Ladd & Sorensen, 2014) find that teachers can continue to improve for at least ten years into their career, especially when their schools provide supportive work environments. Currently, formal evaluation is the most widely adopted policy intended to improve teachers' effectiveness. However, researchers have only begun to understand how evaluation can contribute to teachers' professional growth.

It has long been established that educational policies are variably implemented depending on local institutional factors including motivation, capacity, and competing demands (Berman & McLaughlin, 1978, Berman, McLaughlin, Pincus, Weiler, & Williams, 1979; Elmore & McLaughlin, 1983; Pressman & Wildalvsky, 1973). In a recent mixed-methods study, Kimball and Milanowski (2009) sought to understand factors that influence implementation of a new standards-based evaluation policy in a large district in the West of the US. They found that, despite the detailed rubric of the evaluation instrument used (Danielson, 1996) and the training that evaluators received from the district, principals used teacher evaluations differently across schools. There is much to be learned about the various ways that evaluation policies are implemented and how teachers respond that can illuminate whether and how evaluation can support both development and accountability.

Assuming that evaluation has the potential to serve as a job-embedded learning opportunity, we sought to understand how evaluation works at the school level. We interviewed teachers and administrators in six high-poverty schools in one large urban district. Specifically, we asked what purposes the evaluation process served in practice—accountability, development, or both; how principals conducted observations and assessments, and how teachers experienced

and responded to them.

In 2006, Walker City School District (WCSD)[1], where we conducted this study, adopted a standards-based evaluation tool and a single policy for teacher evaluation to be used district-wide. Not surprisingly, we learned from teachers that their experience with evaluation differed substantially from school-to-school, suggesting that the district's policy was being adapted at the school level (Lipsky, 2010). In fact, we found that principals used the evaluation process for different purposes. In four of the six schools, teachers described the evaluation process largely as a ritualized, bureaucratic activity that prompted procedural compliance, but failed to support either increased accountability or professional growth. In contrast, the principals at the other two schools invested heavily in evaluation in order to improve the quality of teaching in their school. However, even these two principals viewed the purpose of evaluation quite differently. One relied on evaluation to identify and document the shortcomings of teachers who did not meet standards and then to dismiss those who failed to improve. The other used evaluation to achieve both purposes—holding teachers accountable and supporting them for improvement—which teachers widely praised.

In what follows, we begin by reviewing literature on teacher evaluation, situating our study in this larger body of research. After providing background about teacher evaluation in WCSD and describing our research methodology, we present and discuss our findings. We conclude by discussing the study's implications for policy, practice, and research.

Teacher Evaluation in the United States

The process by which administrators observe and assess teachers' instruction against a set of standards has a long history in this country (Cubberley, 1915). For nearly a century, observers

---

[1] We use pseudonyms for the district, schools, and administrators.

relied on relatively simple checklists of topics to evaluate teaching. However, over the past

decade districts widely adopted or adapted Charlotte Danielson's (2013) model for evaluation,

with its 4 domains of teaching responsibility and 22 components. Danielson's framework

includes a rubric with detailed descriptions of four levels of effectiveness for each of its 76

elements. The system was designed to apply to all grade levels and disciplines and to support

formative and summative assessment of teaching practice. An overall rating that summarizes

scores from the domains can be used to make decisions about teachers' re-employment, tenure,

or dismissal. Evaluators can also use evaluation to support teachers' learning and development

by recommending changes in practice and activities such as peer observations, instructional

coaching, or outside workshops. Danielson asserts that, although evaluation is necessary to

identify ineffective teachers, "we must create educative systems that actually result in learning,

that are worth doing from the standpoint of teachers" (Griffin, 2013, p. 29). Other researchers

agree: "[I]t's a waste of effort to use measures of teaching only for high-stakes decisions.

Multiple measures provide rich information to help teachers improve their practice" (Bill and

Melinda Gates Foundation, 2013b, p.7).

Past Failings and Current Hopes for Teacher Evaluation

Researchers and policymakers agree that, whether evaluation is intended to dismiss

ineffective teachers or promote improvement in all teachers' performance—or to do both—very

few school districts currently identify and respond to variations in teachers' effectiveness

(Donaldson, 2009; Toch & Rothman, 2008). In *The Widget Effect,* Weisberg, Sexton, Mulhern,

& Keeling (2009) report that in 12 districts from 4 states, only 1 percent were judged

unsatisfactory; yet 81 percent of administrators and 57 percent of teachers reported that at least

one ineffective tenured teacher worked in their school. Donaldson (2009) describes U.S. school

districts as suffering from "the Lake Wobegon effect" because almost all teachers receive positive ratings, even though teachers within schools vary widely in their effectiveness. Furthermore, school districts rarely use evaluations to dismiss teachers or inform decisions about awarding tenure (Donaldson, 2011; Honowar, 2007; Tucker, 1997). Thirty principals of traditional and charter schools interviewed by Donaldson reported being constrained by time, insufficient opportunities to observe teaching, inadequate evaluation instruments, and a school culture that did not support the process. Researchers and practitioners generally agree with TNTP's (2010) assertion that teacher evaluation in the United States has been largely a "perfunctory compliance exercise that rates all teachers good or great and yields little useful information" (p.1).

Nevertheless, many policymakers and practitioners believe that evaluation can be used to reduce the uneven quality of instruction and to increase overall effectiveness. This is particularly important in high-poverty schools, where students are more likely than their peers in higher-income schools to have less effective and less experienced teachers (Borman & Dowling, 2008; Hanushek, Kain, & Rivkin, 2004). Rivkin, Hanushek, & Kain (2005) report that having a "high-quality" elementary school teacher "can substantially off-set disadvantages associated with low socio-economic background" (p. 419). Thus, students stand to gain if teacher evaluation improves instruction for all.

Current Efforts to Improve Evaluation

Recently, federal policymakers used financial incentives in Race to the Top and waivers from No Child Left Behind to compel states to reform their teacher evaluation system and improve accountability. They required school districts to link evaluation to high-stakes decisions, including teachers' dismissal, tenure, and compensation. They also required states to incorporate

student achievement into teacher evaluation. By September 2012, thirty-three states and Washington D.C. had received waivers from NCLB in exchange for making significant changes to their evaluation processes. Given the federal requirement that evaluation be used for high-stakes decisions, districts adopted new tools for observing and assessing teaching practice. They then concentrated on training principals to make judgments about classroom observations that would be valid and reliable (Papay, 2012).

In a convincing study that links teacher evaluation and student learning, Taylor and Tyler (2011) found that, when Cincinnati used "multiple, highly-structured classroom observations" conducted by experienced peer evaluators and administrators, mid-career math teachers' effectiveness, as measured by their students' achievement, improved. Importantly, the gains persisted and were even stronger several years after the evaluation cycle. The authors suggest that the evaluation process provides teachers with detailed feedback and the incentive to improve. However, because their study does not explain how the evaluation process yielded these results, further study is crucial.

The validity and reliability of an evaluation process is also a function of the evaluator using it (Bill and Melinda Gates, 2103a; Sartain et al., 2011). When researchers studied a Danielson-based evaluation instrument being piloted in Chicago, they found ratings based on classroom observations by trained evaluators to be valid and reliable (Sartain et al., 2011). In the same study, students showed the greatest growth in classes where teachers consistently received the highest evaluation ratings, while students showed the least growth when their teachers received the lowest ratings. These researchers conclude that better evaluation tools can support school leaders in assessing teachers' instructional practices and engaging them in reflective conversations that support improvement, but only when principals have strong knowledge of the

instructional framework, well-developed skills for instructional coaching, and are highly engaged in the process. In a related study, O'Pry and Schumacher (2012) surveyed 121 new teachers in Houston about their views of the district's standards-based performance appraisal system. In order to understand teachers' experiences, researchers then interviewed those whose views of evaluation were most positive or negative. Their findings, consistent with those of Sartain and colleagues (2011), reveal that teachers' perceptions of evaluation were determined less by the tool itself than by how it was used. One factor that most consistently influenced teachers' perceptions of the process was the value that the teachers thought their principal placed on the process.

This research base establishes clearly that developing and implementing an effective standards-based evaluation policy is challenging. Not only must district officials create, select, or adapt an instrument that can be used to validly assess teaching practice, but they also must support evaluators in practicing the process so that they reliably observe and rate teachers' instruction. Having achieved both validity and reliability, administrators still must determine how to use data from evaluations to support teachers in improving their instructional practice.

If reformers hope for evaluations to serve as "engines of professional improvement" (Toch & Rothman, 2008, p.13) throughout the school, much more is required. At a minimum, principals must value the process and convey that to teachers. Yet, we still know little about what individual principals do as they effectively implement new evaluation processes. What goals do they have for evaluation and what practices do they engage in to achieve those goals? Is there convincing evidence that teachers experience evaluation as a process that achieves accountability *and* supports their learning? These are important next questions in this emerging line of inquiry. For, if we do not understand how the evaluation process can promote learning among all

teachers, it will be used—at best—to make decisions about reappointment, tenure, and dismissal. In the process, scarce resources will be wasted and the important goal of improving instruction by promoting all teachers' learning and development will be left behind.

Conceptual Framework

Early writing and scholarship on teacher supervision and evaluation described both "formative" and "summative" stages of the process. A formative assessment was expected to provide feedback to guide improvement, while a summative assessment rendered a final performance appraisal. In some ways, the espoused purposes for today's standards-based evaluation process parallel those of both formative and summative evaluation. Yet, importantly, they are understood to be potentially separate and not necessarily sequential. In fact, pressures from policymakers and the public give priority to the summative assessments that can be used for accountability. School districts and principals who seek to use evaluation to promote teachers' learning and development typically must do so in tandem with documenting performance for accountability. Given the practical challenges and bureaucratic demands of conducting valid and reliable observation-based assessments, it seems likely that goals for learning and development will often be ignored during the evaluation process or incorporated into a different set of practices, such as team-based professional development.

With that in mind, we have developed a conceptual matrix to guide our school-based analysis of evaluation practices (Figure 1). It depicts the purposes of evaluation, as experienced and reported by teachers, along two dimensions—Accountability and Support for Learning and Development. This framework guides our analysis of how teachers experience evaluation processes with respect to both accountability and development. Arguably, the evaluation process would be most effective if it were perceived by teachers to be high on both dimensions

(Quadrant B) and least effective if it were low on both (Quadrant C). Although the range of experiences described by teachers is far more nuanced than a two-by-two table suggests, this tool is useful in considering what it means to use evaluation for both development and accountability.

Background: Teacher Evaluation in WCSD

In 2005, WCSD developed a standards-based tool for teacher evaluation, which then was incorporated into the teachers contract. The policy stated: "True performance evaluation involves analysis of an employee's strengths and weaknesses, resulting in diagnosis and prescriptions which lead to desired professional growth" (WCSD, 2010, p. 1). Administrators are required to conduct classroom observations and assessments each year for probationary teachers and biannually for tenured teachers (WCSD, 2006; WCSD, 2010). They can rate teachers either as "meeting" or "failing to meet" expectations for each of eight dimensions, ranging from "instructional planning and implementation" to "partnership with family and community." Evaluators assign each teacher an overall rating of "meeting" or "failing to meet" expectations. District policy documents squarely emphasized procedures to use the evaluation process as a tool for holding teachers accountable to meet minimum professional expectations. The guidelines describe only the support that must be provided for teachers who fail to meet expectations, not what might be provided to those who already meet them. Furthermore, in a memo to WCSD principals about the new evaluation policy, the superintendent briefly mentioned using evaluation to support professional growth, but provided detailed explanations about how to properly execute the process, especially in response to ineffective teachers (WCSD, 2010).

Like many urban districts, WCSD came under significant pressure from the state to improve its evaluation systems. When we conducted our study in 2010-2011, administrators in all six schools described their efforts to comply with the district's policy by meeting deadlines

for submitting teachers' evaluations. Importantly, we collected data before the state required

districts to include student achievement data in evaluation ratings. Therefore, our findings about

the evaluation process focus entirely on the observation, feedback, and assessments that were

carried out by school-based administrators—without being complicated by the controversial

addition of student achievement data.

Study Methods

This study builds on earlier quantitative research examining the importance of a teacher's

work environment in one state. Teachers' views about their principal, colleagues, and school

culture strongly predicted teachers' professional satisfaction, career plans, and their students'

achievement (Johnson, Kraft, & Papay, 2012). Our subsequent interview study, from which the

current data are drawn, examined teachers' views of their social working conditions—

organizational culture, school leadership, and collegial relationships—in six high-poverty

schools of one large, urban district. In this analysis, we focus on teachers' experience with

evaluation.

Our research questions are:

1. *From the perspective of both teachers and principals in six high-poverty, urban schools, what purpose does teacher evaluation serve at their school?*

2. *How do principals implement evaluation in their schools?*

3. *How do teachers experience and respond to the evaluation process?*

Sample Selection

WCSD, a large urban district on the East Coast, served approximately 60,000 students in

120 schools in 2010. Nearly 90 percent of those enrolled in WCSD were students of color and

approximately 70 percent were from low-income families.

In selecting our sample, we first identified schools that fell above the district median in the proportion of students who qualified for federal free- and reduced-price lunch. We then identified a group of schools exhibiting different levels of student achievement growth (a state-determined measure) and different levels of teacher satisfaction with the school's work environment, as reflected in responses to a statewide survey. From this group, we selected six schools that varied on a range of other measures, including grade level and organization, location, student demographics, and the principal's race, gender, and administrative experience. Our final sample includes two traditional elementary schools, one K-8 school, one middle school, and two high schools. Basic information about these schools is included in Table 1. All would be labeled "high-poverty" schools according to the Institute of Education Sciences' criteria (>75 percent low-income). Each also enrolled large proportions of minority students (>90 percent). Median student growth percentiles across the schools ranged from as low as the 20th and 35th percentiles in mathematics and English language arts to as high as the 65th and 60th percentiles respectively, but were generally clustered around the 50th percentile.

Teacher and Administrator Interviews

Six researchers designed this study and participated throughout data collection and analysis. Two- and three-person teams conducted interviews at each site and the principal investigator participated at all six. We developed the interview protocols based on relevant literature and findings from prior research. (See Appendix A for sample protocols.)

We first conducted a two-hour, semi-structured interview with the principal of each school and then interviewed a wide range of teachers and, where present, other administrators. We sought to interview a broadly representative sample of teachers within each school. (See Table 2 for descriptive statistics). The racial composition of teachers and administrators included

in our sample was broadly representative of the schools and the district as whole. Interviews with teachers lasted approximately 45 minutes and included questions about their experiences with hiring, instruction, evaluation, discipline, the administration, and other factors of the school environment.

Interviews are the main source of data for this study, but we also learned about the evaluation process from documents such as the teachers contract, memoranda from the district superintendent, and the evaluation instrument. Our purposive sampling precludes generalizing about school-based evaluation practices throughout the district, or beyond.

Data Analysis

After each interview we wrote a structured, thematic summary (Maxwell, 2005) highlighting the views and information provided by the respondent on a standard set of topics. During the data collection process, we wrote memos capturing emerging themes by school and across schools.

We coded interview transcripts for central concepts (Strauss & Corbin, 1998) and used a hybrid approach in developing codes (Miles & Huberman, 1994). We coded a small sub-set of the transcripts, individually and together, in order to calibrate our understanding and use of the codes, as well as to refine the codes and definitions. (See Appendix B for a full list of codes and definitions.) We then coded each transcribed interview using ATLAS-TI.

After coding interviews, we engaged in an iterative and collaborative analytic process, relying particularly on data-analytic matrices (Miles & Huberman, 1994). For example, we created cross-site matrices to identify patterns in teachers' accounts of evaluation across schools. Throughout, we sought to understand variation as it became apparent both within and across schools. As we developed tentative findings, we often returned to the data to review our coding

and to test our explanations against the full range of interviews. Throughout the process, we checked our emerging conclusions for rival explanations or disconfirming data (Miles & Huberman, 1994). Having completed this analysis, we could consider where each school fell on our conceptual matrix.

Findings

In WCSD, as in many school systems in the US, district officials implemented a new standards-based evaluation process with the stated intent of supporting teachers' development and holding them accountable for meeting professional standards of practice. In this sample, teachers in four schools described a perfunctory, bureaucratic evaluation practice that actually was used neither to achieve high levels of accountability nor high levels of support for teachers' learning (Quadrant C). In the other two schools, teachers described evaluation as an intense process that was intended to support school improvement. However, at Thoreau High School, evaluation focused largely on accountability with little support for teachers' learning (Quadrant A), while teachers at Giovanni Elementary School reported that evaluation both supported their learning and development, while also holding them accountable for their instructional practice (Quadrant B). (See Figure 2)

We begin our description and discussion of findings by reporting on teachers' expectations for evaluation, which were very similar throughout all schools. We then turn to report the school-to-school differences, which reveal how evaluation actually worked.

Teachers in All Schools Seek Accountability and Support

Most teachers across all schools said they wanted an evaluation system that provided both assessment of teaching practice and support for improving their practice. Where their evaluator attended to both purposes, teachers consistently expressed support for evaluation;

where their evaluator did not, teachers expressed concern that one or both elements were being ignored or used poorly. Somewhat surprisingly, many teachers said that they wanted administrators to hold both them and their peers accountable for providing effective instruction. They also frequently said that evaluation had the potential to support their professional learning, even when it did not currently do so in their school.

Teachers thought teachers should be held accountable. Teachers and administrators in all six schools were concerned about student learning. Most said they taught at least some students who were below grade level and/or learning English. They often also said that poverty affected their students as learners. They did not present these factors as excuses, but rather as the real challenges of working in high-poverty schools.

Teachers explained that, given these demands and the high stakes of academic failure for students, they could not afford to have under-performing teachers as colleagues. A mid-career, teacher at Angelou Elementary School explained, "there needs to be a little bit more push… about people who are doing their jobs and people that aren't. ... [W]e're in such a crisis of [possibly] losing the school...we need to be a little bit harsher." A Thoreau High School teacher offered a similar perspective: "There are some incredible teachers, and there are some people who have absolutely no business in the classroom…."

Teachers appreciated when evaluation increased accountability in their school. Another Thoreau teacher reflected: "[The principal has] made an effort to make teachers and administrators more accountable, which I pride myself on. I believe I do my job, and I think I do a good job." Other teachers noted the lack of respect they experienced when administrators did *not* take the time to observe and assess their practice. An early-career teacher at Stowe Middle School, described her disappointment when she was not evaluated during her third year: "[A] lot

of people . . . said, 'Oh, that should make you feel good.'. . . It doesn't, though." This teacher's

frustrations were echoed by others across the sample who had been evaluated sporadically.

Teachers want support for their own growth. Many teachers across the sample

reported wanting to improve their practice and they saw evaluation as potentially supporting that

goal. A mid-career, middle school teacher explained, "I want to be evaluated. Come into my

room. You know, here's a teacher that wants to be evaluated. And I feel like I do a good job, but

I know I could do better." Similarly, a high school teacher said he felt "The more, the merrier.

Give me some ideas. Ideally if you're evaluating me…jot down a note and tell me what you

thought, so I can learn from it."

Teachers across all schools valued receiving meaningful feedback and objected when

there was no response beyond good ratings or when feedback was, as one said, "fluff." One

teacher's comments echoed those of many others: "[T]he feedback is very slow in coming. …

and then it's just not all that helpful. . . . It's just kind of vague." In contrast, those teachers who

received specific, relevant feedback explained how they integrated it into their practice. One

said, "Definitely very helpful. . . .He suggested that I model more or give an example of what I

want my outcome to look like. . . and I have started using that."

Many teachers said that they wanted their evaluator to spend sufficient time observing

their classes and talking with them about their practice. When administrators stopped by only

briefly, teachers doubted that the appraisal reflected an accurate understanding of their work and

they complained that the process lacked credibility. For example, one middle school teacher said:

> It's a little weird, I mean, I've only been observed once, maybe twice, … she looked at
>
> the board, watched the instruction, watched the "do-now," and she was really only there
>
> for fifteen minutes and then she was gone. And that was really all it was, to be honest.

But the next thing you know, [I receive] the [written] evaluation.

In addition, many teachers asserted that they would gain more from their evaluation if they could discuss the feedback with their evaluator. An elementary teacher with more than 10 years of experience regretted the missed opportunity. "[I]t would be nice to … sit down for 15 minutes instead of when-you're-passing-in-the-hall kind of feedback. . . Having a dialogue would be helpful." A middle school teacher explained the potential benefit of an interactive evaluation process. "[If the principal] would have talked to me and asked me what my problems are, I would have told her flat out what my weaknesses are. . . and maybe I would have gotten a recommendation or advice or some help for them."

Although teachers throughout the study voiced similar hopes for regular observations and timely, substantive feedback and discussion, whether that happened depended largely on the school where they taught. Of the six schools we studied, teachers in only one said that they could count on it.

Perfunctory Implementation: Morrison, Angelou, Stowe and Whitman

At four of six schools—Morrison Elementary, Angelou Elementary, Stowe Middle and Whitman Academy High School—teachers said their administrators generally complied with the district's expectations for evaluating teachers, but that the process remained peripheral to their work. Very few teachers suggested that the evaluation was helpful to them. In fact, most in these schools described evaluation as an empty requirement with little or no benefit for their learning.

Procedure-Focused. When we asked teachers in these four schools about evaluation, most described its bureaucratic procedures rather than its substance. For example, a new teacher at Stowe reported, "[it said that] I met or exceeded everything and then I had to sign it. He gets a copy. I get a copy. And I guess they do that every year." Teachers at all four schools offered

similar descriptions. Some, including this teacher at Whitman, focused on the steps in the process: "You have a pre-observation meeting, then she observes and then you have the post-observation meeting where you go over the evaluation. And if you agree, you sign. If you don't agree, you don't sign, and you write a response."

An elementary teacher at Angelou explained that because her principal was required to hand in his evaluations by March, he subsequently stopped evaluating because the formal evaluations after that date "don't count or something like that." Teachers at these four schools consistently described the evaluation process as procedural and perfunctory.

A routine that teachers endured, but did not value. In addition to saying that evaluation focused on procedures, many teachers in these four schools viewed it as inconsequential. When asked if it was helpful, most concurred with a teacher at Stowe: "Not helpful, no. Honestly, not at all." Another said, "They all say "meets and exceeds," … but it really didn't inform me how to improve, or where exactly I could improve." A teacher at Whitman agreed. "It's not threatening, but to me it is not useful also." When asked if the feedback from her evaluation was helpful, a Morrison teacher in her fourth year of teaching explained, "I know I'm doing what I'm supposed to be doing so I was like, 'whatever.'" Even teachers who responded more favorably described a routine that had minimal impact on their practice. A second-year teacher at Morrison responded with indifference: "They'll just walk through and they'll look around and they'll be like, 'okay,' and they'll leave and then I'll get some sort of feedback later, or not."

Intended to Improve Teaching Quality: Thoreau and Giovanni

Thoreau High School and Giovanni Elementary stand in stark contrast to the other schools. At both, teachers and administrators described the process as central to their principal's

strategy for improving teaching quality and student learning. However, these administrators used the process in markedly different ways and their teachers' views also differed notably.

Thoreau: Evaluating all teachers in order to dismiss the least effective. At Thoreau, a comprehensive high school where approximately 70 teachers taught 900 students, Principal Thomas and her team of six administrators invested considerable time in fully implementing the district's evaluation policy. They reported evaluating all teachers at least every other year. However, most teachers suggested that administrators concentrated on documenting the weakness of the least effective teachers so that they could be dismissed.

Teachers responded with mixed emotions to this intense effort. They expressed appreciation for the increased levels of accountability, but also described the anxiety and insecurity that the process generated among other teachers who they thought were doing a good job.

Administrators' views: "A wake-up call." Administrators at Thoreau spoke at length about the evaluation process, which contrasted starkly with reports from school leaders at Stowe, Whitman, Angelou and Morrison. Principal Thomas, a former mathematics coach, enthusiastically explained the scope of their effort: "We've done a massive amount of evaluating. . . . We're evaluating about 70 percent of the building because that's what you need to be doing in a two-year cycle." Another administrator was hired to oversee the process, although Thomas remained highly involved. Her administrative team met bi-weekly to discuss a book on evaluation, to calibrate ratings and standardize practices used to debrief the observation. For example, they discussed and critiqued a videotaped session in which an administrator discussed an observation with a teacher who had received an unsatisfactory rating.

Throughout their lengthy explanations of the process, these administrators said that

evaluation at Thoreau was intended to increase accountability. Thomas explained,

> So those are the two teachers that . . . we needed to get out and I think what we're going to do with the other four or five that are really [weak]. . . . [T]hey'll probably all get [an overall rating of] 'meets' with 'does not meet' in two or three areas. But we're reserving the option to [give] them a 'does not meet' by the end of the year.

> This, she said, was intended to be a "wake-up call," warning them that they would be

evaluated again in January "and that's how much time you have to get better." The administrators' strategy for improving the overall quality of instruction at Thoreau rested on using evaluation to prod the weakest teachers to improve or to dismiss them.

Teachers appreciated increased accountability. Thoreau teachers confirmed their administrators' accounts. Although they all acknowledged that they had been evaluated, most said that administrators concentrated on a small number of underperforming teachers. Many expressed relief that that the problem of incompetent teachers finally was being addressed. One, who explained that "there had been a lot of teachers here who were horrific . . . and no one was doing anything about them," said that the fact that "certain teachers are being evaluated out" was "in many cases very justified." Another teacher commented that "weeding out the incompetents was the writing on the wall for other teachers. . . . So there is more accountability, more buy-in into the fact that at some point … you will get examined in a real way that makes a real difference."

Evaluation processes fuel fear and distrust. At the time of this study, three veteran Thoreau teachers had been dismissed in the three years after Principal Thomas arrived and several others were under close scrutiny. Few teachers at Thoreau had reason to think that they might be fired. For them, evaluation was largely a routine, empty, bureaucratic process.

However, many expressed concern about the anxiety and distrust generated among the faculty by the focus on dismissal. As one said, "some people feel targeted." Another described "a real sense that they were out to get you. They had like a preconceived notion of who the not-so-good teachers were, and they kind of insensitively went after them, which kind of made other people say, 'Whoa.'" Many teachers said they knew who currently were on what one called "the list" but were apprehensive about who would be next. A teacher speculated that administrators might think,"'[N]ow that they are gone, we can sort of start looking at the next tier.'" This, she said, "started to make people very nervous."

Lack of attention to growth. Teachers at Thoreau also said that evaluation failed to support professional growth for most teachers. A veteran teacher said that "for teachers that are, like, harmful to kids, it's about time." However, "there's a lot of middle ground between a teacher who is awesome every day and a teacher who is harmful to kids" and that teachers' in this "middle ground" were being ignored. As she explained, administrators weren't asking, "'How could we support you to become a better teacher here?'" She believed that the evaluation process might have more support "if that middle ground were actually used to help teachers become better teachers."

Teachers and administrators at Thoreau consistently described an evaluation process that was used primarily to identify and, when necessary, dismiss underperforming teachers. Many appreciated the increased levels of accountability, but raised concerns about how little the process provided for most teachers. They described a demoralizing climate of fear and distrust, which they attributed to their administrators' use of evaluation solely for accountability.

Giovanni: Evaluating all teachers to support growth. Giovanni was a mid-size elementary school with 45 teachers serving 450 students. Principal Gilmore went beyond the

district's requirements and integrated the evaluation process with his school improvement work,

holding teachers responsible for high levels of professional practice, while also supporting their

development. This approach to evaluation exemplifies the growth-oriented approach

recommended by reform groups including the Aspen Institute (Curtis, Wiener, & Aspen

Institute, 2012) and Education Trust (Almy & Education Trust, 2011). With great consistency

across interviews, the principal and teachers described an evaluation process supporting learning

and growth for all teachers, regardless of their competence or years of experience. Teachers often

described evaluation with pairs of descriptors such as "nerve-wracking" and "helpful" or "tense"

and "useful." Teachers widely said that evaluation contributed to a school culture where teachers

felt that they were both accountable for their teaching practice and supported in reaching high

expectations. This case, which differed notably from others in the study, illustrates how

evaluation can simultaneously serve the goals of accountability and development. However, the

fact that this occurred in only one of six schools suggests that doing so is no simple matter.

Gilmore had become principal after considerable experience as a literacy coach and was

skilled in helping teachers improve. He conducted a comprehensive evaluation process with

each teacher at least every other year. Notably, his explanation of the process matched teachers'

accounts. His first priority was to use the evaluation process as a developmental tool. "We can

talk to each other about teaching and learning … [W]hen I do evaluations … I'm not after

anybody; I'm really after . . . how do we improve?"

In order to use evaluation for professional growth, Gilmore had added components to

the district's basic requirements. Teachers prepared lesson plans for their formal observations

and completed a written self-assessment, including professional goals and a record of

professional development experiences. He also required them to provide evidence of their

communication with families, use of student data, and involvement in school-wide teacher

leadership opportunities. Gilmore then met with each teacher before and after the two formal

observations. He said that the process was very demanding for him, but worthwhile because it

ensured that he spent time in classes and actively supported teachers' improvement. Although

teachers described the process as time-consuming and stressful, they also said that it was very

productive.

      Accountability: "He does have strict expectations." Although Gilmore and Giovanni's

teachers focused on development in evaluation, accountability still mattered. Teachers

described their school as a place where they and their students were expected to achieve; they

saw evaluation as one means to communicate and reinforce those expectations. One explained:

> I don't feel as nervous ... knowing that he is there to help you grow. ... He does have
>
> strict expectations. . . . you do have fear. I do respect him, in terms of like that fear, but
>
> I don't fear him in the way that it's going to make my teaching go down. It's to help me
>
> support me teaching, so that's great.

In her opinion, the pressure was productive because Gilmore's purpose was to support her

growth in order to meet high expectations.

      Promoting learning and development. At Giovanni, many teachers described how

Gilmore supported their professional learning and development in the evaluation process. A

teacher with five years of experience, said: "[Mr. Gilmore] came in with the presence

[conveying] that 'I wasn't here to judge you necessarily but to see what you can improve.' And

that makes a big difference for me." She also explained that he wasn't just looking to assign a

grade to her performance but was interested in helping her grow and might even teach with her

during the observation. "'How do you get there?' So he would show you and if you don't know

. . . he'll try to find ways to explain." Many teachers noted Gilmore's willingness to model what he expected teachers to do, especially in literacy, his area of expertise. The fact that he might co-teach on an impromptu basis, even during a formal observation, convinced some teachers that he was a skilled instructor, committed to their improvement.

Teachers spontaneously offered examples of how their teaching improved as a result of being formally evaluated by Gilmore, which happened at no other school in the study. One explained that Gilmore helped her understand how teachers can model learning processes for students. Another said that he taught her ways to improve the quantity and quality of student talk in her classroom. For them, Gilmore was an effective teacher contributing to their learning and growth through a relevant, transparent and respectful process.

Importantly, evaluation at Giovanni was not a stand-alone process. Teachers came to know what Gilmore thought of their work through regular meetings dedicated to activities such as Looking at Student Work, Lesson Study, and team planning. An early-career teacher recalled colleagues helping her understand the evaluation process when she had been new: "When I first had my [evaluation]… I was a little concerned. And other teachers [were] saying 'It's fine. …if he had major concerns about your instruction and like your methods and stuff, you would know.'"

Implementation challenges. Gilmore identified two challenges in implementing the evaluation system effectively—having enough time to carry out the process and knowing both content and pedagogy in all subjects. Although the time commitment was onerous, Gilmore said it was worthwhile because it supported his broader school-wide improvement goals. However, as Giovanni's sole administrator, he could not be the only source of professional expertise for all teachers. Therefore, he relied on others to help teachers improve. A special

education teacher explained that, since Gilmore did not have substantive knowledge in her field, he often suggested development opportunities beyond the school. He also referred individuals to particular colleagues for support. For example, a fifth-year teacher recalled that when Gilmore evaluated her as a new teacher, he recommended that she observe a colleague to support her growth.

> I am sure my first observation was like crash and burn. But he was very helpful …
> instead of . . . taking a negative tone, he said, "You know, I will give you time. You can
> go and watch the other teacher do her guided reading so you can see like how it looks
> and how it should flow.". . . . So that definitely helped me for the first year…."

Although many teachers found the evaluation process challenging and stressful, they also described it as productive and supportive. They suggested that they could trust the process because their principal took responsibility for helping them meet the expectations he set. This contrasted starkly with teachers' descriptions of evaluation as inconsequential or threatening at the other five schools.

Discussion and Conclusions

The teachers interviewed for this study agreed with policymakers that evaluation has the potential to improve the quality of instruction and, ultimately, student learning. Further, they endorsed the use of evaluations to increase accountability. But they also hoped that evaluation would support their professional growth. Across the sample, teachers longed for opportunities to receive detailed, useful feedback, coupled with support for improving their practice. They criticized or dismissed the evaluation process when it lacked those elements.

Despite the intended implementation of a uniform teacher evaluation policy in WCSD, teachers described striking differences among schools in how evaluation was used. In four

schools (Angelou, Morrison, Whitman and Stowe), teachers routinely described evaluation as an exercise in bureaucratic compliance that had no meaningful role either in holding teachers accountable or supporting their professional growth. In the remaining schools (Thoreau and Giovanni), administrators invested significantly in evaluation as a central component of their instructional improvement strategy. However, the two principals' expectations about how this could be accomplished differed markedly. Principal Thomas at Thoreau sought to improve instruction overall by documenting the shortcomings of the weakest teachers and moving to dismiss them if they failed to improve quickly. In contrast, Principal Gilmore at Giovanni approached evaluation as an ongoing process that contributed both to teachers' professional accountability and their development.

The case of Giovanni serves as both a counterpoint to the other cases and as a proof-point of what evaluation can be. Teachers there recognized the dual purposes that Gilmore pursued and found them not only complementary, but also mutually reinforcing. Many described changes in their instructional practice that they made in response to feedback and support from Gilmore during the evaluation process.

Therefore, within this small study, teachers reported three distinct approaches to implementing one district's evaluation policy—perfunctory compliance in four schools, aggressive use to pursue dismissals in a fifth, and in the sixth, integration of evaluation with the school's ongoing program of professional improvement. In considering whether or how well this policy worked, it is important to assess its benefits against the resources that were required to implement it. As Papay's (2012) analysis suggests, except at Giovanni, students were not getting the returns they deserved on their school's substantial investment in the teacher evaluation process.

It is clear that implementing this new, ambitious approach to teacher evaluation presents substantial challenges, which are important to understand if policies like this are to successfully enhance teachers' learning and improve their performance school-wide. Most important was the fact that this policy, similar to those of many districts, depended almost exclusively on school administrators to make it work, requiring them to have high levels of expertise and to make large commitments of time. Each of these challenges deserves careful attention.

Expertise

Much has been accomplished over the past decade in developing comprehensive, valid instruments for evaluation to replace dated checklists. Danielson's widely adopted model, for example, addresses central aspects of pedagogy that are widely understood to be essential for effective instruction. Moreover, when evaluators skillfully use such an instrument, as those in Cincinnati did, standards-based evaluation has been shown to produce lasting gains in teachers' effectiveness (Tyler & Talyor, 2011). Many local districts also have invested heavily in training administrators to observe and assess teachers' performance reliably, thus making it more likely that teachers with similar strengths and weaknesses will receive comparable ratings. Achieving reliability among raters is essential if evaluation is to be fairly implemented.

However, districts are only beginning to invest seriously in promoting the knowledge and skills that school leaders need to make evaluation a productive process of learning and development. Evaluators must be able to advise teachers about best practices in a range of subjects. They need to know both the content of the subject being taught and the pedagogical reasoning specific to that content, what Shulman (1987) calls "pedagogical content knowledge." Obviously, school administrators are unlikely to possess the wide range of knowledge called for to supervise a large and diverse cadre of teachers within a school, particularly at the secondary

level. Although Thomas had knowledge and experience in mathematics as a former coach, her supervisory responsibilities included teachers of students with severe special needs as well as those who taught many subjects about which she lacked deep knowledge.

Finally, administrators who want to ensure that evaluation is a learning experience must be skilled coaches, who can help teachers set a trajectory for improvement and provide the scaffolding they will need to develop. A comprehensive evaluation instrument that includes descriptive rubrics provides a teacher with detailed information about her instructional practices relative to a set of standards. However, this feedback, in itself, provides no guidance about how to improve. In this sample, some teachers described the assessment aspect of the evaluation as thorough, but went on to complain that suggestions about how they might improve, if provided at all, were shallow.

If district administrators expect principals to use evaluation as a growth tool, they must help them develop their coaching skills. Sartain and colleagues (2011) found that many evaluators do not know how to ask teachers questions that foster reflection and discussion during post-observation conferences. Skillful coaches engage teachers in explaining their own instructional approach, analyzing student learning and exploring alternative strategies. In this study, only teachers at Giovanni described evaluation as including such support—coaching, modeling, collaborative planning and related professional growth activities. Admittedly, it is a tall order to expect all principals to be skilled in general pedagogy, knowledgeable about best practices in many subjects, and adept at coaching adults to improve their instruction. And yet the responses of teachers in our study suggest that this is, in fact, what they need from their evaluators.

Although historically the principal was the "principal teacher" of a school and states long required candidates to have completed three years of successful teaching before becoming an administrator, that is no longer the case. Over the past decade, alternative paths to administrative roles and the proliferation of charter schools have meant that principals need far less, if any, teaching experience before being appointed as a principal. Thus, there is no assurance that school leaders can model the essentials of effective pedagogy, let alone more nuanced subject-based instructional skills. Principals will very likely need to rely on others—either administrators or master teachers—who bring complementary expertise to the evaluation process. Giovanni teachers explained that, although Gilmore, who was an expert in literacy, evaluated them, he often deferred to his math coach to lead the work of supporting their math instruction.

This suggests that principals should be encouraged to assemble a team of expert supervisors with deep knowledge and skill in a broad range of subjects and specialties. Teachers who already have attained leadership roles or been certified as "accomplished teachers" by the National Board for Professional Teaching Standards are prime candidates for such responsibilities. Thus, the principal would no longer be the school's only instructional leader.

Time

School administrators often are expected to evaluate more than 20 teachers each year, which requires an enormous amount of time just to meet the basic, procedural requirements of an evaluation policy. A few states now require that every teacher be evaluated every year, and local policies often call for multiple observations of each teacher, including a conference before and after every visit, accompanied by written feedback. As responsibilities for evaluation have grown, the scope of principals' other responsibilities has not been reduced.

Importantly, the differences across schools in how evaluation was implemented cannot be

explained simply by time constraints. All the principals in this study faced similar demands on their time with respect to evaluation, and the larger schools had more administrators to share the burden. However, Gilmore, who had no assistants, completed the task on his own. Possibly, the other principals did not know how to weave the elements of an evaluation process into the fabric of improvement practices, as Gilmore did. Such integrated practices, although potentially very valuable, are uncommon in schools and WCSD provided no systemic approach for preparing principals to do this work. Therefore, it is not surprising that most of the principals in this study chose to invest minimally in evaluation—just enough to comply with the district's expectations or, in the case of Thoreau, enough to dismiss a very weak teacher.

Implications for policy, practice, and research

The findings from this exploratory study raise questions and possibilities that policy-makers, practitioners, and scholars should consider as they work to improve teacher evaluation as an opportunity for learning.

Policy. As McDonnell and Elmore (1987) explain, policymakers can choose from among various types of policies to improve education—mandates that require compliance, short-term inducements that affect immediate behavior, investments in building capacity, and changes in how the system works. It should not be surprising that some principals simply focus on complying with new laws and regulations about evaluation, since the policies are essentially mandates, coupled with some inducements from the federal government for changing the use of evaluations. For evaluation to play a larger, more important role in teachers' learning, policies will likely need to change how the system works—altering the roles and responsibilities of those who carry out the work and building their capacity. If principals are to take responsibility for

implementing evaluation processes, they need to be relieved of other aspects of their current role so that they can devote the necessary time to make evaluations meaningful and consequential.

Some districts have adopted policies that capitalize on the knowledge of highly skilled teachers from inside and outside the system. For example, in Peer Assistance and Review [PAR] programs highly skilled teachers are trained and supported as consulting teachers, who supervise and then formally evaluate other teachers (Johnson & Fiarman, 2012). Programs such as PAR, which draw upon the varied skills and knowledge of expert teachers, can reduce the untenable demands on school administrators for time and specialized knowledge, while creating opportunities for teacher leaders. They also have been shown to increase retention rates for novice teachers and dismissal rates for underperforming tenured teachers (Humphrey, Koppich, Bland, & Bosetti, 2011; Papay & Johnson, 2012).

Practice. Most school leaders and teachers in this sample of six schools from WCSD described evaluation as a procedural requirement, consistent with the tone of the district's written documents describing the process (WCSD, 2006, 2010). This is not surprising given the public's focus on dismissing ineffective teachers. It is important to consider what messages—both formal and informal—principals would need to receive from district leaders so that they would implement evaluation in the service of teachers' learning and growth.

These findings also suggest that leadership development must go well beyond helping principals learn how to validly and reliably observe and assess teachers. Principals need support in learning how to coach competent teachers to become more effective. They need to learn how to connect teachers' experiences in evaluation with other opportunities for on-the-job learning, such as data inquiry cycles, team planning and professional development sessions. Other researchers have found that teachers' experience with evaluation depends on the degree to which

their principal values the process (O'Pry & Schumacher, 2012; Sartain et al., 2011). If school

systems expect principals to effectively use evaluation as a component of their school's

improvement work, they will need to support them in learning how to do so.

Research. Although research has been conducted about current evaluation policies and

what is required to achieve validity and reliability, we need to know much more about how and

how well standards-based evaluation works. Carefully designed quantitative studies can yield

causal findings about the effect of introducing specific policies and practices on student

achievement. They can also provide information about the average effects of new evaluation

policies. However, policymakers and practitioners need to understand the implementation of

these policies—how and why they work (or don't work)—in order to expand the effective use of

these models. This calls for longitudinal, qualitative studies that document teachers' experiences

and views or track adaptations that districts or principals make as they adopt state or local

policies. For example, do different state policies—especially those with components designed to

build capacity, rather than simply mandate compliance—influence how principals and teachers

implement them? Does it matter whether local policies are developed collaboratively with

teacher unions or adopted unilaterally by school boards? Do recent efforts to create career

ladders for teachers in states such as Iowa or districts such as Baltimore, lead them to rely on

expanded teams of evaluators including expert teachers? What can we learn from natural

experiments in districts that adopt PAR programs or rely on outside experts to confirm or

disconfirm principals' ratings, as in Washington D.C.'s IMPACT program or New Haven's

TEVAL? Overall, we need to know more about how different approaches to implementing

evaluation affect instructional practice and students' learning.

In redesigning policies, improving practices and conducting research, it is important to realize that, although evaluation focuses on assessing and improving the performance of individuals, the challenges that principals and teachers face are largely organizational ones that call for organizational responses. Strategies for swapping out a few weak teachers for adequate teachers will not substantially improve a school's instructional capacity or enhance the quality of learning for students as they move from class to class or grade to grade. If schools are to truly improve so that all students have skilled and committed teachers, evaluation must simultaneously support both accountability and development. But, as the experiences of these six schools suggest, effectively implementing such a system is no small matter.

References

Almy, S., & Education, T. (2011). Fair to everyone: Building the balanced teacher evaluations

    that educators and students deserve. Teacher Quality. *Education Trust*.

Berman, P., & McLaughlin, M. W. (1978). Federal programs supporting educational change.

    Retrieved October 2, 2015, from http://www.rand.org/pubs/reports/R1589z8.html

Berman, P., McLaughlin, M. W., Pincus, J. A., Weiler, D., & Williams, R. C. (1979). An

    exploratory study of school district adaptation. Retrieved October 2, 2015, from

    http://www.rand.org/pubs/reports/R2010.html

Bill and Melinda Gates Foundation (2013a). Ensuring fair and reliable measures of effective

    teaching: Culminating findings from the MET project's three-year study.

Bill and Melinda Gates Foundation (2013b). Feedback for better teaching: Nine principles for

    using measures of effective teaching.

Borman, G. D., & Dowling, N. M. (2008). Teacher attrition and retention: A meta-analytic and

    narrative review of the research. *Review of Educational Research*, *78*(3), 367-409.

Cubberley, E.P. (1915). The Portland Survey: A textbook on city school administration.

    Yonkers-on-Hudson: World Book.

Curtis, R., Wiener, R., & Aspen, I. (2012). Means to an end: A guide to developing teacher

    evaluation systems that support growth and development. Aspen Institute.

Danielson, C. (2013) The Framework for teaching evaluation instrument, 2013 edition: The

    newest rubric enhancing the links to the Common Core State Standards, with clarity of

    language for ease of use and scoring. The Danielson Group.

Donaldson, M. (2009). So long Lake Wobegon? Using teacher evaluation to improve teacher

    quality. Washington, DC: Center for American Progress.

Donaldson, M.L. (2011). Principals' approaches to developing teacher quality: Constraints and

opportunities in hiring, assigning, evaluating and developing teachers. Washington, DC:

Center for American Progress.

Elmore, R., & McLaughlin, M. (1983). The federal role in education: Learning from experience.

Education and Urban Society, *15*(3), 309–330.

Feinman-Nemser, S. (2001). From preparation to practice: Designing a continuum to strengthen

and sustain teaching. *Teachers College Record. 103(6*), 1013-1055.

Griffin, L (2013). Charlotte Danielson on teacher evaluation and quality: A School

Administrator interview with the creator of the Framework for Teaching. *School

Administrator (1)*70. 27-31.

Hannaway, J. (2009). Testimony before the District of Columbia City Council on DCPS Human

Capital Initiatives. Washington, DC: The Urban Institute.

Hanushek, E. A. (2009). Teacher deselection. In D. Goldhaber & J. Hannaway, (Eds.), *Creating

a new teaching profession* (pp. 109-118). Washington D.C.: Urban Institute Press.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2004). Why public schools lose teachers. *Journal

of Human Resources*, *39*(2), 326-354. University of Wisconsin Press.

Hill, H. C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating

teacher value-added scores. *American Educational Research Journal*, *48*(3), 794–831.

Honawar, V. (2007). New York City taps lawyers to weed out bad teachers. *Education Week,

27*(14),13.

Humphrey, Koppich, Bland & Bosetti (2011). *Peer review: Getting serious about teachers'

support and evaluation.* Menlo Park, CA: SRI International.

Johnson, S. M., & Fiarman, S. E. (2012). The potential of peer review. *Educational*

   Leadership, *70*(3), 20-25.

Johnson, S. M., Kraft, M. A., & Papay, J. P. (2012). How context matters in high-need schools:

   The effects of teachers' working conditions on their professional satisfaction and their

   students' achievement. *Teachers College Record, 114*(10), 1-39.

Kraft, M.A. & Papay, J.P. (2014). Can professional environments in schools promote teacher

   development? Explaining heterogeneity returns to teacher experience. *Educational*

   Evaluation and Policy Analysis, 36(4), 476-500.

Ladd, H.F. & Sorensen, L.C. (2014). Returns to teacher experience: Student achievement and

   motivation in middle school. Working paper 112, CALDER.

Lipsky, M. (2010). *Street-level bureaucracy, 30th Ann. Ed.: Dilemmas of the individual in public*

   service. Russell Sage Foundation.

Maxwell, J.A. (2005). *Qualitative research design: An interactive approach* (2nd ed.). Thousand

   Oaks: CA. Sage Publications.

McDonnell, L.M. & Elmore, R.F. (1987). Getting the job done: Alternative policy instruments.

   Educational Evaluation and Policy Analysis, 9(2), 133-152.

McLaughlin, M. W. (1991). The RAND change agent study: Ten years later. *Education policy*

   implementation, 143-155.

Miles, M. & Huberman, M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.).

   Thousand Oaks: CA. Sage Publications.

O'Pry, S.C., & Schumacher, G. (2012). New teachers' perceptions of standards-based

   performance appraisal. *The Journal of Educational Assessment, Evaluation and*

   Accountability, 24(4), 325-350.

Papay, J. P. (2012). Refocusing the debate: Assessing the purposes and tools of teacher evaluation. *Harvard Educational Review*, *82*(1), 123-141.

Papay, J.P., & Johnson, S.M. (2012). Is PAR a good investment? Understanding the costs and benefits of teacher peer assistance and review programs. *Educational Policy, 26*(5), 696-729.

Pressman, J. L., & Wildavsky, A. B. (1984). *Implementation: how great expectations in Washington are dashed in Oakland: or, why it's amazing that federal programs work at all, this being a saga of the Economic Development Administration as told by two sympathetic observers who seek to build morals on a foundation of ruined hopes*. Univ of California Press.

Rivkin, S., Hanushek, R., & Kain, J. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417–458.

Sartain, L., Stoelinga, S., Brown, E. R., & Consortium on Chicago School, R. (2011). Rethinking Teacher Evaluation in Chicago: Lessons Learned from Classroom Observations, Principal-Teacher Conferences, and District Implementation. Research Report. CCSR.

Shulman, L.S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, (*57*)*, 1–22.

Strauss, A. C. & Corbin, J. M. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed)*. Thousand Oaks, CA: Sage Publications.

Taylor, E. S., & Tyler, J. H. (2011). The effect of evaluation on performance: Evidence from longitudinal student achievement data of mid-career teachers (Working Paper 16877). Cambridge, MA: NBER.

Thomas, E., Wingert, P., Conant, E., & Register, S. (2010). Why we can't get rid of failing

teachers. *Newsweek, 155*(11), 24-27.

TNTP. (2010). Teacher evaluation 2.0. New Teacher Project.

Toch, T., & Rothman, R. (2008). Rush to judgment: Teacher evaluation in public education.

Washington, DC: Education Sector.

Tucker, P. (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.

Walker City School District (2006). *Teachers contract.*

Walker City School District (2010). *Superintendent's Circular: Performance evaluation of

teachers.*

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:

TNTP.

Figure 1: Conceptual Framework—Purposes of Evaluation as Experienced by Teachers

High	*Quadrant A: Teachers are held accountable for their professional practice, but **not** supported for learning and development*	*Quadrant B: Teachers are held accountable for their professional practice **and** supported for learning and development*
Low	*Quadrant C: Teachers are neither held accountable for their professional practice, **nor** supported for learning and development*	*Quadrant D: Teachers are supported for learning and development, **but not** held accountable for their professional practice.*

Accountability (vertical axis label)

Low — High (horizontal axis)
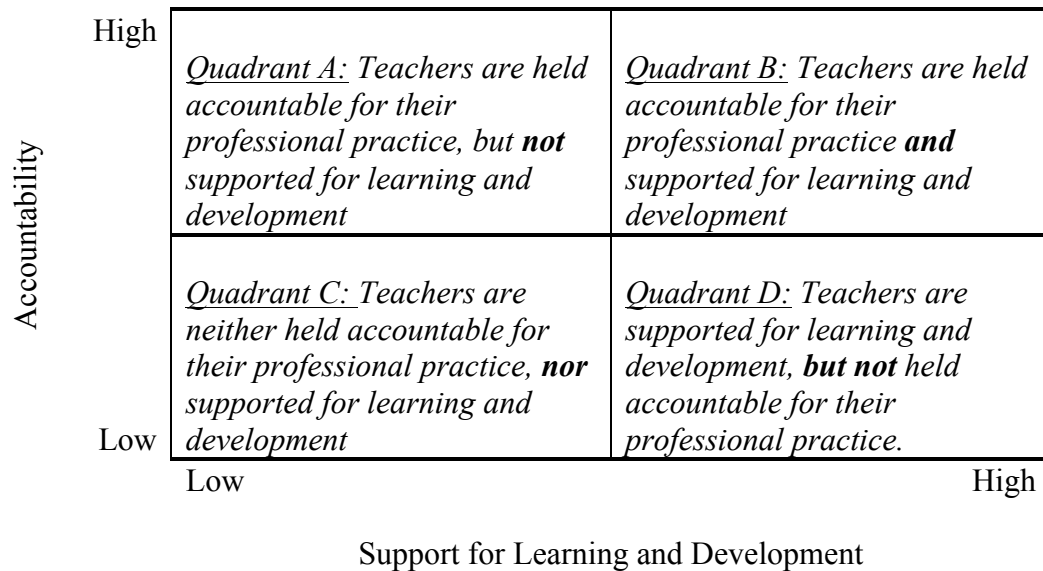
Support for Learning and Development

Figure 2: Conceptual Framework—Predominant Purposes of Evaluation by School

High	*Quadrant A:* • Thoreau High School	*Quadrant B:* • Giovanni Elementary School
Low	*Quadrant C:* • Morrison Elementary • Whitman Academy • Stowe Middle School • Angelou Elementary School	*Quadrant D:*

Accountability (vertical axis label)

Low — High (horizontal axis)

Support for Learning and Development