Teachers’ Perspectives on Evaluation: Exploring the Effects of New Evaluation Policies on Teaching

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

Citation

Citable link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:33797240

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA
Teachers’ Perspectives on Evaluation:
Exploring the Effects of New Evaluation Policies on Teaching

Qualifying Paper
Submitted by
David Braslow
Harvard Graduate School of Education
April 2016

Author Note
This research was conducted in conjunction with the National Center for Teacher Effectiveness, housed at the Center for Education Policy Research at Harvard University.
In 2009, the federal Race to the Top (RTTT) competitive grant program launched with the goal of catalyzing state-level education policy reforms. Among the desired reforms were changes to human resource systems that would develop “great teachers,” including teacher evaluation reform. Spurred on by research that showed that improvements in student performance vary considerably across teachers (Gordon, Kane, & Staiger, 2006; Rivkin et al., 2005), legislators and advocates hoped that new evaluation systems would improve teaching quality – and thus student outcomes – in two ways: by promoting the dismissal of ineffective teachers and by improving effectiveness among extant teachers. By 2013, thirty-six states changed their teacher evaluation requirements (Doherty & Jacobs, 2013), including requirements that student achievement data inform evaluations, that teachers receive more regular observation and feedback, that observations align with new standards of teacher practice, and that rating systems include more than two categories. Thus, in terms of its effect on state evaluation policies, RTTT has largely succeeded in motivating rapid and substantial changes.

What remains unclear is whether these policy shifts translate into improvements in teaching. While evaluation systems may have some impact by pushing out teachers who are found to be ineffective (Dee & Wyckoff, 2013; Rockoff, Staiger, Kane, & Taylor, 2012), preliminary evidence suggests that evaluation reforms are unlikely to lead to widespread teacher dismissals (J. Anderson, 2013). Rather, most teachers who receive evaluations under the new systems will remain in the classroom, and it remains unclear whether new evaluations can help these teachers improve (e.g. Murphy, Hallinger, & Heck, 2013; Peterson & Peterson, 2006).

For evaluation reforms to help teachers improve they must do so through some mechanism, in which a set of conditions and events produce instructional changes. For example, requirements that teachers receive more detailed feedback could improve teaching if
administrators are able to provide detailed feedback to teachers, the feedback has useful content that leads teachers to reflect on their practice, and teachers then devise a new plan of action and implement it successfully in their classrooms. However, if any of these steps fails, new feedback requirements may have limited effects on teaching. Moreover, new evaluation systems could have unintended consequences that harm instruction, such as if they introduce competition among teachers that hinders collaboration (Darling-Hammond, 2013).

In this study, I explore four mechanisms through which new evaluation systems might improve instruction by analyzing teachers’ perceptions of how new evaluation procedures have affected their work. This study focuses on mechanisms of interest to proponents of evaluation reform: feedback, professional development, collaboration, and motivation. By describing teachers’ perceptions of evaluation procedures and relating them to each of these mechanisms, I am able to document some of the reasons why new evaluation policies may or may not lead to desired improvements in instruction.

This study builds on prior research exploring the effects of evaluation reform in five ways. First, this study uses thematic analysis of interviews with teachers to provide detailed descriptions of teachers’ experiences. Second, this study relates teachers’ experiences to specific mechanisms of instructional improvement, informing the theories underlying those mechanisms. Third, this study examines multiple mechanisms of improvement, providing a comprehensive portrait of teachers’ perceptions of the impact of new evaluations. Fourth, this study includes teachers who experienced both the old and new evaluations, providing a useful contrast for examining the impact of the new system. Lastly, this study synthesizes teachers’ experiences across two states, providing findings transferable to other contexts.
Literature Review

In this section, I review existing research on the effects of evaluations on instructional quality, highlighting the importance of teachers’ perceptions for understanding the effects of new evaluations. I further summarize existing findings on the four mechanisms of improvement studied.

The Effects of Evaluation of Instructional Quality

Evaluation designed to support teachers’ growth, known as formative evaluation, has been promoted for decades as a systemic approach to instructional improvement, leading to its inclusion in RTTT (e.g. Brock, 1981; Danielson & McGreal, 2000; Darling-Hammond, 2013; Duke & Stiggins, 1990). Advocates for formative evaluation argue that evaluation processes can serve an important role as one part of a comprehensive teacher development system. Classroom observation and feedback procedures are important in this model because they provide both information about classroom practice and occasions for teachers and administrators to act on that information to improve teaching.

Evidence has been growing that formative evaluation has the potential to improve instructional quality at scale. Two studies of pilot evaluation systems including formative components have found that they have positive effects on student test scores (Steinberg & Sartain, 2015; Taylor & Tyler, 2011), and teachers in numerous other studies have reported instructional improvements resulting from new evaluation procedures (Donaldson et al., 2014; Firestone, Nordin, Shcherbakov, Kirova, & Blitz, 2014; Huckstadt, 2011). However, many teachers in these studies have also reported that there were no effects on their instruction, leaving open the question of why new evaluation procedures might have different effects for different teachers.
The Importance of Teachers’ Perspectives on Evaluation

To understand why new evaluation procedures might be more helpful for some teachers than for others, it is necessary to understand the differences among teachers that might influence their responses to those procedures. Teachers’ personal characteristics, such as their current teaching practices, understandings, and worldviews, could explain the differential effects, as they have been shown to influence how teachers make sense of and respond to instructional reforms (Coburn, 2001, 2004; Hill, 2001). Differences in the implementation of new procedures across districts, schools, or evaluators could also explain teachers’ responses, as there can exist considerable variation at each of these levels (Ehlert et al., 2013; Tennessee Department of Education, 2013).

Teachers can provide unique insight into both the personal and contextual factors that influenced their responses to new evaluation procedures because of their dual roles as participants and potential beneficiaries. As participants, teachers can describe, from their perspective, how evaluations were conducted – a perspective that is sometimes different from observers’ perspectives (Donaldson et al., 2014; Stecher, Garet, Holtzman, & Hamilton, 2012). As potential beneficiaries, teachers can also describe their responses to the new procedures and explain why they responded the way they did, providing a detailed understanding of how new evaluations affected them. With information about both procedural changes and teachers’ responses, it is possible to investigate the causal chain linking new state requirements to school-level evaluation practices, teachers’ evaluation experiences, and, ultimately, teachers’ behavior.

Four Mechanisms of Improvement

New evaluation procedures could affect teachers through many conceivable mechanisms. In this study, I focus on four such mechanisms: providing actionable feedback to teachers,
targeting professional development to teachers’ needs, supporting productive collaboration among teachers, and increasing teachers’ motivation. I chose these mechanisms because advocates commonly use them to promote evaluation reform and because the districts I studied have new systems that rely on each (see Appendix A).

**Feedback.** The clinical supervision model of evaluation argues that post-observation feedback can lead teachers to improve their teaching (R. H. Anderson & Snyder, 1993; Cogan, 1973; Range, Scherz, Holt, & Young, 2011). Evaluation reforms aim to provide teachers high-quality feedback, with increased feedback detail, observation frequency, and conversations between observers and teachers.

If feedback is to yield instructional improvement, three things must happen: teachers must receive high-quality feedback, they must reflect on it to consider potential instructional changes, and then they must actually make changes in practice. The first step involves what is *provided*: observers must give teachers feedback containing a prompt about how to improve, such as a critique, question, recommendation, or low rating. The feedback must be detailed enough to clearly identify specific elements of instruction that teachers could improve and must accurately identify areas for growth (Hill & Grossman, 2013). The teacher must then *reflect* on his or her teaching after receiving the feedback – to examine what happened during the lesson, to consider what might be improved, and to devise a new plan of action. Lastly, the teacher must *take action* to put any desired changes into practice, and the changes must not be superficial adjustments.

It remains unclear whether feedback provided in the context of formal evaluations can lead teachers to improve their instruction. Psychological studies suggest that feedback about task performance can yield improvements under certain conditions, but often has negligible or
negative effects (Kluger & DeNisi, 1996). Evaluators can struggle to provide high-quality feedback to their teachers (Hill & Grossman, 2013; Kraft & Gilmour, 2015), and teachers are mixed about whether they find the feedback they receive helpful for improving their teaching (Donaldson, 2012; Jiang, Sporte, & Luppescu, 2015; O’Pry & Schumacher, 2012; Stecher et al., 2012). While some teachers in these studies describe receiving feedback that they thought was helpful, others describe not receiving feedback or receiving feedback without useful content, precluding further reflection or action.

**Professional development.** New evaluation systems often require administrators to ensure that teachers receive professional development (PD) aligned with the areas for growth identified in their evaluations (Doherty & Jacobs, 2013). Administrators may use the information gathered from observations to coordinate learning activities or to direct teachers toward resources that could support their growth. Alternatively, evaluators may expect teachers to seek out PD on their own. Both approaches require that teachers or administrators successfully identify teachers’ needs via the observation, self-assessment, or goal setting process, and then find effective PD that teachers can access to address those needs. In this mechanism, the observation rubric itself may serve as PD, educating teachers about what constitutes high-quality practice so they can enact it in their classrooms (Thomson, 2014).

Putting aside concerns about the effectiveness of PD for improving instruction (Hill, 2009; Wei, Darling-Hammond, & Adamson, 2010), it remains unclear whether administrators have the capacity to use evaluations to inform the PD teachers receive. Districts usually delegate PD supervision to school personnel, who can struggle to connect PD with teachers’ evaluations when they have little time and few resources, or when PD is provided by the district (Donaldson et al., 2014; Tennessee Department of Education, 2013). A number of studies have described
districts where evaluations do seem to inform PD offerings, yielding improvements in instruction (Donaldson & Peske, 2010; Hamilton et al., 2014; White, Education, Cowhy, Stevens, & Sporte, 2012). However, teachers’ sometimes disagree with their administrators regarding whether the PD they receive is actually connected with their evaluations (Huckstadt, 2011; Stecher et al., 2012).

Collaboration. Evaluation systems may promote collaboration, defined here as work toward instructional goals with colleagues, in at least three ways. First, the explicit teaching standards can help teachers establish and reflect on common instructional expectations (Danielson, 2011; White et al., 2012; Wise, Darling-Hammond, McLaughlin, & Bernstein, 1985). The specific elements of practice found on observation rubrics can provide a common language and set of instructional goals, helping teachers and administrators work together toward those goals.

Second, the use of student achievement data in evaluations may support collaboration when teachers are accountable for the performance of shared students. Teachers jointly responsible for a group of students have an incentive to work together to improve the achievement of those students. In addition, evaluation systems often require groups of teachers to choose their own assessments, providing an opportunity for teachers to create shared goals for student learning.

Third, evaluation systems can promote collaboration by directly evaluating teachers’ collaboration. Including collaboration explicitly on the observation rubric could both incentivize teachers to engage in collaboration and help administrators identify teachers who need support for collaboration.
These mechanisms rely on a number of assumptions about the *nature of interactions among teachers*. For example, collaboration must focus on substantial improvements in instruction, rather than on ways to game the system. The evaluation system must also avoid negatively influencing collaboration, such as by introducing competition or decreasing morale (Darling-Hammond, 2013; Johnson, 2015).

Little research exists on the effects of evaluation systems on collaboration, but there are reasons to hope there might be positive effects. Teachers sometimes describe their evaluations as collaborative when there are back-and-forth conversations with administrators after observations (e.g. Donaldson, 2012; O’Pry & Schumacher, 2012); one study found that such collaborative supervision improved teachers’ commitment to collaboration (Ebmeier & Nicklaus, 1999). Administrators in Tennessee reported that school-wide goals for student performance promoted collaboration (Tennessee Department of Education, 2013), and Jiang, Sporte, and Luppescu (Jiang et al., 2015) found that teachers who had positive perceptions of their evaluations had stronger professional communities. However, no studies explore teachers’ perceptions of the effects of new evaluations on their work with colleagues.

**Motivation.** New evaluation systems often rely on incentives to motivate teachers to improve their teaching, including formal consequences such as performance pay or dismissal threats (Firestone, 2014). Expectancy theory outlines how workers are motivated when they value the rewards and believe that their actions can affect the likelihood of receiving those rewards (Vroom, 1964). In the context of evaluations, this theory implies that teachers must *want to receive the benefits tied to high ratings or avoid the sanctions tied to low ratings*, and they must *trust that evaluators will rate them fairly* so that any efforts they make to improve will be rewarded (Vasset, 2010). New evaluation systems try to ensure fairness with features such as a
standardized observation rubric or the ability to present evidence to demonstrate proficiency with a skill not seen during an observation.

In addition to using formal consequences, evaluation systems can capitalize on teachers’ internal motivation by providing clear instructional goals that teachers will want to pursue (Firestone, 2014). Studies have shown that specific, challenging goals can motivate stronger performance under certain conditions (Locke & Latham, 2002). Teachers can internalize instructional goals independently by studying the observation rubric or with help through formal goal-setting processes commonly required by new evaluation systems.

Both approaches to motivating teachers ultimately aim to increase the effort teachers expend or the productivity of their work (Jones & Wright, 1992; Miller, 2005). However, teachers are unlikely to do either if they lack the capacity to do so without experiencing excessive stress (Kelley, Heneman III, & Milanowski, 2002). Moreover, teachers must be motivated to undertake productive work on instructional tasks, as opposed to procedural tasks required by the new evaluation system.

New evaluation systems do seem to influence teachers’ motivation, but the ways they do so are complex. Dee and Wyckoff (2013) found that teachers who received high ratings that came with merit pay responded by improving their performance, but low ratings that came with dismissal threats only led teachers to attrit. Donaldson (2012) found that attaching consequences to student test scores led teachers to focus on raising scores, and that goal setting helped some teachers plan more coherent courses of instruction; however, few teachers changed their regular instructional activities. Teachers often report that new evaluations are stressful, but it is unclear how much of the stress might subside after teachers become more familiar with them (Ehlert et al., 2013; O’Pry & Schumacher, 2012).
Research Questions

Two research questions guide this study: 1) How do teachers describe changes in their work resulting from new evaluation procedures? 2) How do teachers’ responses support or refute theorized mechanisms relating new evaluation requirements to instructional improvement? To answer the first question, I create detailed descriptions of teachers’ perspectives on the ways new evaluation systems have affected them. To answer the second question, I connect those descriptions with the intermediate steps and conditions in the mechanisms outlined above, providing evidence about whether those mechanisms are functioning as expected.

Data

Sample

The National Center for Teacher Effectiveness (NCTE) conducted the interviews for this study in fall 2013 with the intention of recruiting teachers from as many different sites as possible. This variability in the sample ensures that the study’s findings are not proscribed by the features of any one site. Teachers\(^1\) were recruited from two large urban districts (one in Georgia and one in Massachusetts) chosen from among the four districts that had a working relationship with NCTE and that were implementing new evaluation systems. These districts’ evaluation systems shared multiple features related to the four mechanisms under investigation, such as increased observation requirements and revamped observation rubrics (see Appendix A). NCTE recruited teachers from twelve schools that had a working relationship with NCTE (7 in Georgia and 5 in Massachusetts). Thirty teachers ultimately responded and were interviewed (20 in Georgia, 10 in Massachusetts), with no more than three teachers per school. This sample

\(^1\) Other personnel evaluated as teachers, such as teacher coaches, were eligible to participate
provides a range of reported experiences across schools, districts, subject areas, and years of experience (see Table 1).

**Interviews**

Interview questions explored teachers’ experiences with new evaluation procedures in general and related to the theorized mechanisms outlined above (see Appendix B). We developed the interview protocol with input from teachers from the target population. The protocol was semi-structured and took approximately 60 minutes. NCTE researchers conducted the interviews by phone, recorded the conversations, and transcribed them for analysis. Interviewers used pre-determined probes for each question if the respondent did not address the topics in their responses. Interviewers also used follow-up questions at their discretion to explore teachers’ responses further. Interviewers also collected demographic information after each interview.

**Analysis**

This study uses cross-case analysis, with teachers as the unit of analysis, to describe teachers’ experiences related to the four theorized mechanisms (Miles, Huberman, & Saldaña, 2013). I used Dedoose (2015) to implement the two-stage approach to thematic analysis described by Miles, Huberman, and Saldaña (2013). In the first cycle, a subset of transcripts were coded using four theory-driven descriptive codes, and descriptive subcodes further categorized teachers’ responses (see Table 2). Analytic memos documented code generation decisions throughout this process.

To ensure that the coding would be replicable, I used inter-rater reliability checks to test my coding against that of another coder. The second coder received training in using the codebook, and then coded a subset of transcript excerpts using the codebook. We discussed and revised codes with low agreement. Codes for three of the mechanisms – feedback, PD, and
collaboration – required two rounds of coding to achieve agreement with Cohen’s Kappa above 0.60, a common benchmark for good interrater agreement (Cicchetti, 1994). For the last mechanism – motivation – three rounds were required. I then coded the remaining transcripts using the revised codebook.

In the second cycle of coding, I used focused coding to look for themes from each theorized mechanism by forming clusters of related descriptive subcodes. These clusters produced a manageable number of themes that related teachers’ experiences to the mechanisms of interest, and structure the findings below.

Once coded, I analyzed excerpts tagged with each subcode to provide rich descriptions of teachers’ experiences and to substantiate the identified themes. Code application counts further describe the prevalence of different experiences. Lastly, I refined the descriptions by looking for teachers with contrasting experiences, providing discrepant evidence that clarified the identified themes.

Results

Feedback

This section presents themes pertaining to the feedback teachers received and their responses. I only include segments pertaining to communications between observers and teachers about instruction seen during observations. I organized codes according to the theorized mechanism underlying actionable feedback: teachers receive substantive, accurate feedback that they must then reflect on in way that leads to changes in practice.

Most teachers (20) did not find that the feedback they received led them to improve their instruction. Teachers often described changes of some kind after receiving feedback, such as thinking differently when planning lessons or providing additional documentation to
administrators. However, only some (6) of these teachers said they made substantial improvements in their instruction, and one teacher said her feedback led her to make changes that were harmful, not beneficial. The themes below show what happened for each step of the theorized mechanism to explain the lack of instructional improvement.

**What is provided.** Teachers generally described receiving feedback that was detailed and accurate, but feedback often lacked a prompt for improvement.

All the teachers we spoke to received feedback of some kind. All teachers got ratings, and all but one got comments, written or otherwise. Comments ranged from simple affirmations (e.g., “Good job”) to observations about features of the lesson (e.g., “Sharing poster information”) to comments about specific strengths or weaknesses (e.g., “The initial transition in those three groups took a little long”). Most teachers also described having formal post-observation conferences to discuss lessons and receive feedback.

Most teachers also thought the lesson feedback accurately described their teaching. More teachers described the feedback as accurate (19) than inaccurate (5). Teachers thought the feedback was accurate when they saw observers being attentive during the observation or when they agreed with the content of the feedback they received. In contrast, teachers found the feedback not reflective of their teaching if they thought that the observed lessons were not a representative snapshot of their general practice, or that their observers did not have the necessary expertise to provide accurate feedback.

Most teachers (19) also thought that the feedback they received went into more detail about their instruction than feedback they had received in the past. Many of these teachers noted the level of detail in the different dimensions of the observations rubrics, such as one teacher who received specific feedback for each dimension:
She actually took time to go in to every single standard, all ten, and to compliment me on all the things that she saw and the things she knows I do outside of the observation…. She had listed everything that I do within our school building.

However, while most teachers received feedback that they thought was accurate and detailed, feedback often lacked strong prompts for improvement. Most teachers received positive feedback or feedback about areas of strength (23), while only half of the teachers (15) mentioned receiving any negative feedback or having observers explicitly identify any areas for growth. Moreover, beyond simply receiving criticisms or ideas about what to improve, few teachers (5) reported receiving any recommendations for how to improve their practice in the areas for growth their observers identified.

Two contrasting examples show how specific recommendations can be more useful for prompting reflection than criticisms alone. One teacher described helpful recommendations she received after an observed lesson went awry:

Last year, I had a student with pretty severe emotional and behavior needs, and one of my students started to have a meltdown during the lesson…. [The observer and I] were able to talk about, “Hey, these are some of the good things that you did within the lesson,”… and then other strategies I might be able to use for that lesson and even classroom management things that I might consider if the situation occurred with that student again. Such instructional recommendations relevant to a teacher’s classroom could be an important resource for teachers looking to improve their practice. However, teachers were more likely to receive simple criticisms lacking any direction for how they might improve:

Where I wasn’t successful was the category of “Teaching All Students”…. She came in when I had one of the most difficult classes in the school…. When they were in my class,
no, they weren’t perfect, but they were actually pretty good. She still used that one particular time… throughout the year and graded me down.

Such critiques, absent identification of possible causes of and solutions to the problems, leave many teachers to their own devices to figure out how best to improve.

**Reflection.** If feedback is to lead teachers to improve their instructional practice, it must do so by prompting teachers to reflect on their teaching and to consider new instructional strategies that would better meet their goals. However, there is no guarantee that feedback will lead to such reflection, even if it does include a prompt.

Given the absence of strong prompts, it is not surprising that only a few teachers (7) reported reflecting on the feedback they received by considering possible instructional improvements. This typically happened when ratings or other comments identified areas for growth. These teachers usually reflected on what it would take to achieve the highest score in as many dimensions as possible, as in the case of one teacher:

- This year it’s, “These are the things that you’re exemplary at, these are the things that you’re proficient at.” So instead of it just being an all-or-nothing… it allows more opportunity for growth. It allows you to say, “Yes, I know what I’m supposed to do, but how can I make this more pervasive so that I can become exemplary?”…. I really want to be an exemplary teacher in all ten areas.

Most teachers did not report engaging in such reflection, but instead had emotional responses. Teachers who received positive feedback often appreciated the affirmation of their current practice (e.g. “It was great. I was blown away”), while teachers who received negative feedback often felt hurt (e.g. “[She] graded me down…. I felt a lack of support from administration”).
These teachers did not seem to engage in reflection, simply accepted or rejected the judgments made about them without considering instructional changes.

**Taking action.** After considering instructional changes, teachers would actually need to enact those changes for instruction to improve. Moreover, changes would need to go beyond temporary or superficial adjustments that do not improve the quality of typical instruction.

While most teachers lacked strong prompts for improvement and did not reflect on their instruction, some teachers (7) did report making direct connections between the feedback they received and specific changes in their teaching. For example, one teacher made changes because she received a recommendation to use a specific instructional strategy from her observer:

My boss pushed me last year to make sure that they were having the visuals to match what I was talking about. So, this year is just anchor charts, and anchor charts, and anchor charts. I’ve got a family tree, so that we can see those words…. a lot of visuals around the room that I wouldn’t have had last year.

Other teachers’ responses, however, were less encouraging: a few (3) made superficial changes to appease evaluators, some (6) made other changes that did not seem to affect instruction, and the rest (14) made no changes at all. For example, one teacher described a behavior management system that she implemented superficially to appease her evaluator without substantially changing her teaching:

I started making more charts to control. I started taking more time making these charts that I could use as artifacts with the help of this mentor I had. Those charts actually – it’s kind of a show…. I see these kids once a week for 45 minutes, and [the charts] look okay. They’re there.
More common were teachers who planned their future lessons to make sure that certain practices would be evident when observers were in the room. As one teacher said, “…if sometimes they won’t observe certain things when they are in the room, then I will make sure that next time he does come by that I do whatever he didn’t observe the first time.” Yet other teachers said they thought differently about instruction after receiving feedback, but did not mention making any specific instructional changes, as in the case of one art teacher who said, “I’ve gotten a lot of really positive feedback, so I can’t say that I’ve so much changed things. I’ve just been more aware of the differentiated instruction in the art room.”

Most teachers, however, simply reported making no instructional changes. Many of these teachers were the ones who did not receive prompts for reflection in their feedback, as in the case of one teacher who received positive feedback: “[My feedback] just sort of made me feel justified in what I’m doing. It wasn’t inspiring, but it made me say, ‘Okay. I guess I’m doing an okay job. Keep it up. Keep doing what I’m doing.’” Other teachers who did receive prompts for reflection did not think they could address the issues identified, either because of resource constraints or because the issues were inexorable aspects of teaching. For example, one teacher did not make changes in response to critical feedback she received about student engagement because she thought that some disengagement was normal: “I think that's the reality of it. You're dealing with 10 year old kids…. at any given time, the likelihood that there are going to be some kids who are not engaged is high.”

**Summary.** The teachers we spoke to all received feedback of some kind, and usually thought the feedback they received was accurate and detailed. However, only half of the teachers mentioned receiving any kind of feedback that might have prompted reflection, and those prompts were often criticisms that left teachers to their own devices to figure out *how* to improve
their teaching. As a result, only a few teachers ended up reflecting on their feedback to devise a course of action or make any actual instructional improvements.

**Professional Development**

This section presents themes about how new evaluations influenced the PD teachers received. New evaluation systems could improve the effectiveness of PD if administrators use evaluations to align it with the needs identified in teachers’ evaluations. We asked teachers about the PD they received before and after the new evaluation system was implemented, and coded the different ways teachers reported that the new system affected the PD they received.

Overall, most teachers (19) did not describe any changes at all in the PD they received after the implementation of the new evaluation system. Some of these teachers mentioned that they had always attended PD as they saw fit, and the new evaluations did not change that. Others had no explanation for the lack of change. We identified the themes described below from the experiences of the remaining teachers.

**Aligned PD.** Only two of the teachers we spoke to reported receiving PD based on the needs identified in their evaluations. One of these teachers described how her observer assigned a coach to her after she received low ratings. The other teacher described working with her colleagues to ensure that they received PD tailored to the instructional goal they had set during the beginning-of-year goal setting process required by the new evaluation system: “We’re trying to think creatively about how to support our school goal…. So [we’re] thinking about a district-level PD that’s being offered, and supporting teachers and making sure that they attend.” It is not particularly surprising that these were the only teachers who reported receiving PD aligned with the areas for improvement identified in their evaluations, given that many teachers did not even receive feedback identifying potential areas for improvement.
While they did not explicitly mention receiving aligned PD, a few (5) teachers said that the evaluation system helped them understand how they needed to develop, such as one teacher who said the feedback provided “a way to monitor your own growth.” The observation rubric itself seemed to help these teachers better think about how they wanted to develop, as in the case of one teacher who said, “The rubric really clearly delineated, ‘If you doing this, then you are at this level.’ I like having that language to be able to connect with what I’m doing, and to know: ‘How do I improve on my practice?’” For these teachers, the rubric might have served as a resource to help them to improve their own teaching in their areas for growth.

Teachers also described two types of PD that changed with the new evaluation system but that did not specifically align with their identified areas for growth. Some (7) teachers participated in trainings that explained how the new system worked and what was expected of teachers (e.g. “My school did a really good job of making sure we all understood what they were going to look for before anyone put foot in our classroom.”) Teachers also mentioned that they had access to online videos and materials provided by the district that were aligned with the dimensions of the observation rubric, but only one teacher reported actually looking at them. Teachers might have used these either of these resources to learn about topics related to their self-identified areas of need, but none reported doing so.

**Summary.** While administrators could use teachers’ evaluations to direct them to PD that would address their areas of need, this seemed to be the exception rather than the norm. Rather, most teachers bore full responsibility for finding PD to improve in their areas for growth, and there was little evidence that they did so. However, the observation rubric or other training may have helped some teachers learn about – and thus better meet – the instructional standards.
Collaboration

This section presents themes about the effects of evaluations on collaboration. I include segments pertaining to teachers’ interactions with colleagues about work toward improving instruction. I organized codes according to the three ways that new evaluations might support collaboration: creating common instructional expectations for teachers, increasing accountability for shared groups of students, and evaluating teachers on their collaboration directly. In addition, I present evidence regarding the nature of the collaborative interactions promoted by the new system and discuss some unintended consequences described by the teachers we spoke to.

About half of the teachers (17) we spoke to reported that the evaluation system did not affect their collaboration with colleagues. While most (25) teachers did talk about the evaluation system itself with colleagues, the discussions usually focused on sharing information and opinions about it, not instructional improvement. When asked why there had been no change, some teachers had no explanation, but others described how conditions at their schools precluded any positive impact. For example, one teacher described how her school did not make substantial changes in routines to promote collaboration:

I can see what they were going for…. There were supposed to be more points of contact. There was supposed to be more collaboration, more conversation, more discussion about your teaching and about your practice and about the way that you were delivering instruction, and I just feel like that hasn't really happened. To me, it's the way it's currently being implemented, at least in my school.

Despite saying their collaboration had not changed, some of these teachers described ways the new system affected their work that seemed related to collaboration. These experiences are included below, along with those of teachers who did report changes in collaboration.
Establishing and reflecting on common expectations. The new instructional standards, communicated via the observation rubric, could provide common terminology or instructional goals that help teachers and administrators work together more effectively to improve instruction. Some teachers (8) did indicate that the new evaluation system helped to establish common expectations of what instruction should look like in their schools. One teacher described how observers set instructional expectations clearly throughout her school using the observation and feedback process:

[Under the old system] the expectation of what you were supposed to do in your classroom… was a little bit more flexible and not everybody was doing the same thing. Now it’s very clear what we expect to see when we walk into your classroom and if we don’t see this, then were going to sit down and talk with you.

Another teacher described how teachers in her school developed common understandings of instructional expectations by learning what instructional features evaluators would look for:

We are cognizant of the things that they're going to be looking for when we're evaluated. So when we're doing our lessons and we know that they're going to be coming by… we're aware of, “Hey, you need to make sure you have these things in a lesson.”

These examples show how observers’ expectations, as communicated through feedback about observed lessons, can serve to generate shared instructional expectations by providing a common external audit of instructional quality.

Three teachers further described reflecting on their instruction with their colleagues because of the new evaluation system. For example, one teacher described the new conversations: “The evaluation system has got people thinking: ‘Now, how do I need to be teaching? What do they need to see? What does it need to look like in my room’... What is this
job about?" Such collective reflection could help teachers further develop shared understandings, enabling teachers to support each other in working to meet the new expectations.

**Accountability for shared students.** While teachers in both districts were required to set student learning goals at the beginning of the year, evaluators were allowed to push teachers to adopt team-wide or school-wide goals instead of individual ones. Such group goals could encourage collaboration about instruction, since teachers are held accountable for the performance of other teachers’ students. In fact, five teachers described working with other teachers who taught the same students to set student learning goals as part of the evaluation system. One of these teachers described how this process helped to create shared expectations for students, to develop a plan of action, and to coordinate assessments:

> We met as a team and came up with our student learning goal together. Then we just finished creating our action steps for that goal…. Part of our goal… is to do certain assessments. That’s a nice piece of it because it actually gives us measures. We have that kind of language about what our expectations are for students. It also ensures that there’s some consistency in how students are being assessed.

For these teachers, the new evaluation systems provided an opportunity to work together in service of common student learning goals.

**Direct evaluation of collaboration.** In Georgia, where collaboration is an explicit element of the observation rubric, observers rate teachers on their collaboration. For example, one teacher described how her observer sat in on meetings to evaluate the ways in which teachers worked together:

> There are things that she needs to observe… like being a team player and supporting your colleague…. So when we’re together in faculty meetings or committee meetings… she is
watching to see who is displaying those traits, who is bringing things to the table, who is sharing ideas.

Two contrasting examples show how such direct evaluation of collaboration can be used supportively or punitively to improve collaboration. One teacher described how the added accountability got teachers to work together when they otherwise might not have:

I know at this school there is a push to have teachers collaborate. There’s definitely a push on it…. So we’re saying we want you to be on the same page… but unless there’s that accountability… that’s the piece that’s going to hold it together.

Another teacher, however, described how pressure to collaborate pushed teachers out of her school: Teachers were told that, “[collaboration] was no longer an ‘if’; it was a ‘Yes, we are doing it,’” and when a group of colleagues received low ratings due to insufficient collaboration, they reportedly left the school. Regardless of the approach, the inclusion of collaboration as part of the observation rubric has certainly increased some teachers’ attention to it.

**Nature of interactions.** While the previous sections described ways that the new evaluation system promoted collaboration, this section describes the nature of the new teacher-to-teacher interactions, which logic suggests must focus on improving instruction.

Some teachers (10) said that they shared specific instructional strategies with their colleagues to help them better meet the new instructional expectations. These teachers usually shared strategies that had yielded positive feedback with colleagues who wanted to improve in those areas. For example, one teacher described how the receipt of feedback became an occasion for teachers to discuss instructional strategies and share successful ones:
When we’re getting our evaluations back, we’re sharing with each other and saying, “Oh, I did good at this,” or, “You say I need to work on this.” … People are still saying, “Well, you could do this because this is what I’m doing that’s working.”

These accounts support the idea that teachers do respond to the new evaluation system by learning new strategies or sharing new strategies with others.

However, teachers did not always describe sharing strategies in order to improve instruction. Three teachers described how teachers worked together to make sure that they would receive positive evaluations by sharing information about how evaluators rated different instructional practices. While such collaboration might be helpful, as administrators’ judgments could point teachers towards better practices, these teachers were primarily interested in getting high ratings. For example, one teacher described how discussions of lesson plans focused on how lessons would look to observers (e.g. “There’s a lot more discussion about, ‘Well, if I do it this way, I’m not going to get evaluated as good.’”). Another teacher gave a colleague advice about how to modify her lesson to ensure that that his or her observation went well (e.g. “I was talking with that teacher and sharing some ideas of ways that she could do the things that I knew the principal would be looking for in the lesson. I knew how she had evaluated me.”). While sharing instructional strategies could improve the quality of instruction, such superficial efforts to appease evaluators seem unlikely to do so.

**Unintended consequences.** This section focuses on two unintended consequences of the new evaluation system that may have hindered teachers’ collaboration: increased competition and decreased morale among teachers.

Some teachers (5) described how the new evaluation system increased competition among teachers. These teachers described how they shared ratings and worked to improve their
standing relative to their colleagues. Such competition could potentially create positive pressure to improve, as in the case of one teacher’s school:

Teachers are sharing their results with each other in a competitive kind of way. They want to know what their neighbor’s getting because they want to make sure they’re scoring just as high. So I think it’s pushed teachers to really consider where they are… and to strive to do better.

In contrast, another teacher described how differences in ratings across teachers led to divisions among the faculty, who thought they deserved better ratings than their peers had received: “I heard that our principal gave them a rating of exemplary. Well, I'm a better teacher than they are. How come they got a rating of exemplary and I only got a rating of proficient?” These cases show that competition, while potentially beneficial, is unlikely to improve collaboration.

In addition, some teachers (7) described how the new evaluation system hurt their relationships with colleagues because it engendered negative feelings, decreasing faculty morale. One teacher described how the new system damaged relations at her school because it increased tensions between administrators and teachers: “I think it made it less collegial between administrators and teachers… it feels more punitive… it used to feel like a team or a family.” These teachers perceived low ratings as punitive actions; one teacher described low ratings as, “depressing to some people… you can really see morale go down.” While these teachers did not specifically describe negative effects on their collaboration due to decreased morale, such negative effects on the morale of school personnel may have damaging effects.

**Summary.** While some teachers did not see much impact of the new evaluation system on their collaboration, others identified a variety of effects. Collaboration seemed to improve when shared instructional expectations or shared accountability for students provided teachers
with a framework for working together on instruction, or when teachers responded positively to pressure from observers to collaborate. Alternately, collaboration seemed to be hurt when teachers saw low ratings as punitive or felt the need to improve their standing relative to their peers. While teachers usually worked with colleagues to make substantial improvements in instruction, others seemed to make superficial adjustments to appease evaluators, casting some doubt on the actual instructional impact of any improved collaboration.

**Motivation**

This section focuses on themes about how the new evaluation system motivated teachers to work harder or differently. I organized codes according to the theory for how evaluation systems might motivate teachers: teachers want to receive high ratings and trust evaluators to rate them fairly, or are intrinsically interested in meeting the new instructional standards, so they increase their effort or productivity to achieve their goals.

**Desires for high ratings.** Fourteen of the teachers talked about wanting to receive high ratings or avoid receiving low ones. Teachers who wanted to receive high ratings hoped to earn the exemplary label in recognition of their good work (e.g. “How can I make myself exemplary…?”) Some of these teachers seemed particularly motivated to get high ratings on every dimension (e.g. “This has actually challenged me to do better, to want to score exemplary in all areas that I could.”) Others talked specifically about their fears of receiving low ratings, such as one teacher described a low rating she received as, “like a threat.” In contrast, three teachers said they did not fear receiving low ratings, saying things like, “I’ll think they’ll be doing me a favor if they get rid of me so that I can just be an artist.”

**Trust.** When asked whether they thought the new evaluation system was fair, fourteen teachers described it as fair, eight described it as unfair, and eight described fair and unfair
elements. One feature of the new systems that teachers thought promoted fairness was that they provided a more comprehensive or representative evaluation of their work because it included multiple observations (e.g. “It’s fair because it’s not a snapshot. You have a lot of chances.”)

Another feature some teachers appreciated was the consistency in rating they perceived among evaluators (e.g. “They were working on reliability. They were looking to make sure that they were all seeing the same, and how they were scoring, and the feedback they were giving.”) Others appreciated the ability to provide evidence to ensure that ratings were fair (e.g. “It's easy to prove if you're doing what you're supposed to do.”) While teachers had different reasons for trusting the evaluation process, most were able to identify some feature of the new system that they thought ensured some level of fairness.

Teachers described their evaluations as unfair when they saw observers rating subjectively or in a biased manner. Teachers thought ratings were subjective when the observation rubric left room for interpretation, observers lacked the necessary expertise to evaluate them (e.g. “One of the evaluators has never been a classroom teacher.”), or some observers appeared generally harsher than others did. Teachers thought ratings were biased when certain categories of teachers received relatively low ratings, such as teachers of lower-performing students, minority teachers, and strong teachers (some teachers said principals did not give out the highest rating). While some amount of judgment is required in any observation-based system, these teachers did not think evaluators were exercising it properly.

**Instructional goal setting.** The new evaluation system may motivate teachers differently by helping them set clear and challenging instructional goals. One requirement of the new systems was that teachers had to set official goals at the beginning of the year. These goals
included student learning goals and goals for professional practice aligned with the new teaching standards, and could be either group or individual goals.

Eight teachers talked about the goal-setting process during the interviews; while some simply mentioned that it happened, others described how it affected their work. For example, one teacher described how goal setting helped build coherence among teachers at her school: “It gave us a chance to discuss the goals of our school as a whole and to try to align our personal goals with the goals of the school.” Another teacher liked how student learning goals helped her reflect, saying, “you can definitely see growth, and you can see if there are any lulls.” These comments suggest that some teachers were able to use the goal setting process to focus on different elements of teaching that they wanted to work to improve.

Aside from the formal goal setting process, nine teachers spoke about holding themselves to higher standards because of the new instructional expectations and the increased scrutiny of their work. For example, one teacher described how the new evaluation system, “lets teachers know, ‘Hey, I can’t just teach the lessons. They’re looking for more than this.’ So I think it also helps guide teachers to say, ‘Here’s what my boss is looking for. Here’s what’s expected of me.’” Another teacher explained how she appreciated the new, clear expectations, so that she could adjust her teaching to meet those expectations:

I didn’t really like the old way because it was just, “Hey, you are doing a good job.”… You didn’t know the expectation, and I like to know what is expected of me…. If there is anything that I need to be doing better or that they could help me with, then I definitely want that.

These descriptions suggest that, for some teachers, the feedback they received allowed them to understand and work to meet their administrators’ instructional expectations.
**Effort and productivity.** Evaluations might improve teaching by inducing teachers to exert more effort, such as by working more hours or completing tasks they would otherwise avoid. For this to work, teachers would need to have the capacity to exert more effort without overlooking other important tasks or experiencing excessive stress, and the new tasks teachers complete must lead to instructional improvement.

The new evaluation system did not seem to increase teachers’ level of effort, except perhaps for evaluation-related tasks. Twelve teachers reported that the new system did not affect the level of effort they put into their work, while five said that it did.¹ Fourteen other teachers mentioned that the new system itself took a large amount of time, but it was unclear whether those teachers exerted more effort to complete evaluation tasks or completed them at the expense of other work. Teachers who did not exert more effort usually said that they were already working as hard as they were able. Teachers who did exert more effort usually described spending extra time documenting their instruction to meet evaluation requirements. Only one teacher reported exerting additional effort to improve her instruction because of the new system, finding herself “[taking] extra time to type up rubrics on a project that you didn’t think needed a rubric, or preparing different lessons for different groups of students.”

The extra work required of teachers often seemed to have a negative effect on teachers’ well-being. Thirteen teachers described how the new evaluation system increased their levels of stress or anxiety. Some teachers described stressing about completing required evaluation tasks, while others found the challenge of scoring highly on all of the observed dimensions stressful (e.g. “Good teachers feel, ‘Oh, I should be doing all of this,’ and so we get very paranoid about, ‘Oh, well, let me see if I can do more.’”) In one extreme case, a teacher described having heart

¹ We added this question after completing some interviews, so we did not ask all teachers.
palpitations because of anxiety about the new system. Such stress is an unfortunate side effect of efforts to motivate teachers to work harder or more productively.

**Summary.** Teachers are often motivated to get high ratings because they want to receive the recognition conveyed by a high rating or to avoid the sanctions that come with low ratings. However, teachers differ in the amount that they trust their evaluators to rate them fairly, potentially undercutting their motivation to make substantial changes in their work. The establishment of clear instructional goals did support some teachers’ motivation to meet their observers’ instructional expectations. This motivation, however, rarely resulted in teachers expending additional effort improve instructional tasks, as teachers were often working at capacity or burdened with administrative tasks.

**Discussion**

The accounts of the teachers we interviewed suggest that the four theorized mechanisms of instructional improvement may not work for many teachers. Descriptions of the observation and feedback process suggest that there are multiple points of failure in the chain that links feedback through reflection to instructional improvement, such that few teachers ultimately make substantial changes. Few teachers reported receiving any PD that would help them to improve on the areas for growth identified in their evaluations. Many teachers saw little impact of the new system on their collaboration, and positive effects, such as common instructional expectations, seemed as prevalent as negative effects, such as competition and low morale. Motivational effects were mixed, with teachers interested in scoring well and stressing themselves out to do so, but only sometimes in the service of actually improving instruction. These findings do not bode well for the prospect of using evaluation as a tool to drive system-wide instructional improvement.
Some teachers’ accounts, however, do show ways that new evaluations can support instructional improvement. For each mechanism, there were examples of teachers who did find that the system worked as intended. Most promising were the feedback and collaboration mechanisms, for which a substantial minority of teachers reported positive effects, and the goal-setting component of the motivation mechanism.

This study reinforces the need to attend to teachers’ perspectives when trying to understand and refine new evaluation systems. Teachers shared many ways new systems supported their efforts to improve instruction, while also surfacing unintended consequences that hindered their work. Teachers varied in the ways they made sense of and responded to the same evaluation procedures, suggesting that administrators may need to tailor evaluations to serve the needs of individual teachers. Understanding why some teachers find new evaluations more useful than others is essential for determining how to improve new systems. This study provides some explanations for differences among teachers, but a number of important questions remain. For example, teachers seemed to find their feedback more useful when it included prompts for reflection or specific recommendations for how to improve, but it remains unclear why only some teachers reported receiving such feedback. What factors explain whether observers provided feedback with prompts? Did some teachers receive feedback including prompts that they did not notice? Future research should further explore the reasons why teachers differ in their responses to new systems. Teacher or contextual characteristics could explain some of this variation. Alternatively, there may be important conditions or steps missing from the mechanisms studied here, or other mechanisms entirely may help better understand how new systems’ impact.
This study’s findings suggest some concrete recommendations for policymakers to make new evaluation systems more useful for instructional improvement. Observers need support in providing feedback that includes specific recommendations for how to improve, especially for teachers who are already strong performers. This could include training for school administrators or having trained experts provide teachers with feedback. Administrators and teachers also need support finding PD to support their areas for growth – districts could provide detailed guidance about what PD offerings address which instructional standards and ensure that teachers access those resources. Encouraging groups of teachers to set shared student learning or instructional goals could help push teachers to work together more closely. Teachers also should be relieved of as many administrative tasks related to evaluation as possible to ensure that their efforts to earn high scores focus on instructional improvement.

Decades of reforms aimed at helping teachers improve have failed to produce large-scale change, and contemporary evaluation reform may soon suffer the same fate. New evaluation systems may be an improvement over the ones they replaced, but that does not mean that they will succeed at driving instructional improvement where the previous systems failed. While some observers, schools or districts may succeed at implementing evaluations in ways that support teachers’ growth, these cases seem likely to be the exceptions. In addition, teachers’ interpretations of and responses to evaluation reforms seem likely to vary considerably, such that a limited number may ultimately find the process helpful. Policymakers will need to work to understand why evaluations only benefit some teachers and address the identified problems if they hope to use evaluations to improve instruction at scale. Some states and districts have already begun to make changes to their newly minted teacher evaluations since the reauthorization of NCLB, providing some hope that ongoing refinements are possible and could
make new evaluation systems more useful for teachers (e.g. Brown, 2016). Alternatively, policymakers may choose to focus on just using the new systems to dismiss ineffective teachers, but they should consider the negative effects these systems can have on extant teachers’ instruction and ensure that the benefits outweigh the harms.
References


www.americanprogress.org


Table 1

*Sample Characteristics*

<table>
<thead>
<tr>
<th></th>
<th>Range</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years of Teaching Experience</td>
<td>4 - 31</td>
<td>13</td>
</tr>
<tr>
<td>% Special Education Students</td>
<td>0% - 100%</td>
<td>22%</td>
</tr>
<tr>
<td>% English Language Learners</td>
<td>0% - 100%</td>
<td>25%</td>
</tr>
<tr>
<td>Grades Taught</td>
<td>K – 8&lt;sup&gt;a&lt;/sup&gt;</td>
<td>—</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>83%</td>
</tr>
<tr>
<td>White</td>
<td>87%</td>
</tr>
<tr>
<td>Taught All Subjects</td>
<td>57%</td>
</tr>
</tbody>
</table>

N = 30 teachers

<sup>a</sup> Two teachers taught middle school students in addition to elementary students.
<table>
<thead>
<tr>
<th>Theme</th>
<th>Subcode</th>
<th>Description</th>
<th>Number of teachers</th>
<th>Example Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feedback</td>
<td>Amount of feedback</td>
<td>Teacher discusses amount of feedback received</td>
<td>10</td>
<td>&quot;She gave a tremendous amount of feedback.&quot;</td>
</tr>
<tr>
<td></td>
<td>Accurate/Inaccurate description</td>
<td>Feedback did (not) accurately reflect what happened in the lesson</td>
<td>19/5</td>
<td>&quot;I feel that they did a very good judgment what I do each day.&quot;</td>
</tr>
<tr>
<td></td>
<td>Detailed feedback</td>
<td>Teacher thinks the feedback was detailed</td>
<td>19</td>
<td>&quot;The write-up was much more detailed on the brief than what I would have gotten [in the past].&quot;</td>
</tr>
<tr>
<td></td>
<td>Positive/Negative feedback</td>
<td>Observer provided positive/negative feedback about instruction</td>
<td>21/7</td>
<td>&quot;He gave me a really good report.&quot;</td>
</tr>
<tr>
<td></td>
<td>Identified strengths/weaknesses</td>
<td>Evaluator pointed out specific areas of strength/weakness for the teacher</td>
<td>10/9</td>
<td>&quot;It wasn’t just, 'Oh, yeah. You did a great job.' It was like very specific: 'It was good when you did this and you mentioned this.'&quot;</td>
</tr>
<tr>
<td></td>
<td>Recommendations for improvement</td>
<td>Observer recommended specific strategies for teacher to use</td>
<td>5</td>
<td>&quot;He will write… any ideas that he has to make it even better.&quot;</td>
</tr>
<tr>
<td>Reflection</td>
<td>Reflection on practice</td>
<td>Feedback led teacher to reflect on his or her practice</td>
<td>5</td>
<td>&quot;How can I make this more pervasive so that I can become exemplary?&quot;</td>
</tr>
<tr>
<td></td>
<td>Positive/negative emotion</td>
<td>Teacher has positive/negative response to feedback</td>
<td>19/12</td>
<td>&quot;I was very appreciative to her. She actually took time to go in to every single standard… and to compliment me.&quot;</td>
</tr>
<tr>
<td>Theme</td>
<td>Subcode</td>
<td>Description</td>
<td>Number of teachers</td>
<td>Example Text</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>--------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Taking action</td>
<td>Direct/No response</td>
<td>Teacher did (not) make changes in instruction directly related to the feedback received</td>
<td>7/14</td>
<td>&quot;Now I have that focus and I keep that at the top of my head and I try to include it now weekly in the lessons.&quot;</td>
</tr>
<tr>
<td></td>
<td>Superficial response</td>
<td>Teacher made superficial changes to instruction in response to feedback</td>
<td>3</td>
<td>&quot;As soon as an administrator walks in they switch into the dog and pony show…. We all do it.&quot;</td>
</tr>
<tr>
<td></td>
<td>Ambiguous response</td>
<td>It is unclear whether teacher made changes in instruction related to feedback</td>
<td>6</td>
<td>&quot;I can’t say that I’ve so much changed things, I’ve just been more aware of the differentiated instruction in the art room.&quot;</td>
</tr>
</tbody>
</table>

**Professional Development**

<table>
<thead>
<tr>
<th>Theme</th>
<th>Subcode</th>
<th>Description</th>
<th>Number of teachers</th>
<th>Example Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aligned PD</td>
<td>Assigned collaborator</td>
<td>Teacher was assigned a collaborator as a result of his or her evaluation</td>
<td>1</td>
<td>&quot;They blocked me down and… I was demoted and now I have a mentor, a peer mentor.&quot;</td>
</tr>
<tr>
<td>Collectively seeking PD</td>
<td>Monitor growth</td>
<td>Teachers work together to access professional development aligned with evaluations</td>
<td>1</td>
<td>See p. 19</td>
</tr>
<tr>
<td></td>
<td>PD on evaluation system</td>
<td>Evaluations allow teacher to monitor his or her own growth</td>
<td>1</td>
<td>&quot;It’s a way to monitor your own growth&quot;</td>
</tr>
<tr>
<td></td>
<td>Explicit rubric</td>
<td>Teacher uses the detailed guidance in the observation instrument to conceptualize and/or improve his or her teaching</td>
<td>4</td>
<td>&quot;It was more specific… if we wanted to improve in certain areas, we knew how to go about that improvement.&quot;</td>
</tr>
<tr>
<td></td>
<td>PD on evaluation system</td>
<td>Teacher received PD about the evaluation system</td>
<td>10</td>
<td>&quot;Everybody is required to attend a meeting where we look at the rubric that they're going to be assessing us on&quot;</td>
</tr>
<tr>
<td>Theme</td>
<td>Subcode</td>
<td>Description</td>
<td>Number of teachers</td>
<td>Example Text</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-----------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>--------------------</td>
<td>-------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Online resources</td>
<td>Online resources</td>
<td>Online resources are provided that are aligned with the observation rubric</td>
<td>5</td>
<td>&quot;They have this… 'interactive rubric,' where you can click on each of the standards… and you can see what an exemplar teacher looks like.&quot;</td>
</tr>
<tr>
<td>Collaboration</td>
<td>Collaboration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Establishing and reflecting on common expectations</td>
<td>Shared understandings</td>
<td>Teachers have shared understanding of what instruction should look like due to new evaluations</td>
<td>3</td>
<td>&quot;Everybody is more on the same page as far as what is expected of all of us.&quot;</td>
</tr>
<tr>
<td></td>
<td>Collective reflection</td>
<td>Teachers collectively reflect on their evaluation results</td>
<td>3</td>
<td>&quot;The evaluation system has got people thinking, 'Now, how do I need to be teaching?'&quot;</td>
</tr>
<tr>
<td>Accountability for shared students</td>
<td>Common goal setting</td>
<td>Teacher worked with colleagues to coordinate goal setting for evaluations</td>
<td>5</td>
<td>&quot;We formed team goals together.&quot;</td>
</tr>
<tr>
<td>Direct evaluation of collaboration</td>
<td>Accountability</td>
<td>Teachers worked together because of accountability from evaluation system</td>
<td>2</td>
<td>&quot;At this school there is a push to have teachers collaborate… the accountability piece… is going to hold it together.&quot;</td>
</tr>
<tr>
<td>Nature of interactions</td>
<td>Sharing instructional strategies</td>
<td>Evaluation led teachers to share instructional strategies</td>
<td>10</td>
<td>&quot;These are my goals for this year and, if we both do technology, we can work on some of the same PowerPoints together.&quot;</td>
</tr>
<tr>
<td></td>
<td>Gaming observations</td>
<td>Teachers worked together to find ways to improve evaluation scores aside from genuine interest in improving practice</td>
<td>3</td>
<td>&quot;There’s a lot more discussion about, 'Well, if I do it this way, I’m not going to get evaluated as good.'&quot;</td>
</tr>
<tr>
<td>Theme</td>
<td>Subcode</td>
<td>Description</td>
<td>Number of teachers</td>
<td>Example Text</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>---------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>--------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Unintended consequences</strong></td>
<td>Competition</td>
<td>Teacher feels competitive with other teachers at the school because of the new evaluation system</td>
<td>5</td>
<td>&quot;We have had a couple situations where teachers are sharing their results with each other in a competitive kind of way.&quot;</td>
</tr>
<tr>
<td></td>
<td>Negative impact on collaboration</td>
<td>Evaluation had a negative effect on collaboration</td>
<td>12</td>
<td>&quot;People feel really threatened by being observed… and want to just be left alone.&quot;</td>
</tr>
<tr>
<td><strong>Motivation</strong></td>
<td>Encouragement</td>
<td>Teacher was encouraged by evaluator</td>
<td>2</td>
<td>&quot;I got a good evaluation so it’s nice to be reinforced for the work I was doing.&quot;</td>
</tr>
<tr>
<td><strong>Desires for high ratings</strong></td>
<td>Recognition</td>
<td>Teacher received or sought to receive recognition for strong performance through evaluation system</td>
<td>6</td>
<td>&quot;I want other people to be able to see the really great stuff that I try to do.&quot;</td>
</tr>
<tr>
<td></td>
<td>Fear of negative consequences</td>
<td>Teacher is afraid of potential negative consequences</td>
<td>4</td>
<td>&quot;You have to get an either proficient or exemplary. Your job depends on that. If you get two or three [below], they can let you go.&quot;</td>
</tr>
<tr>
<td></td>
<td>Protection</td>
<td>Evaluation allows teacher to protect self from negative consequences</td>
<td>2</td>
<td>&quot;It’s good because… it’s a way to protect yourself.&quot;</td>
</tr>
<tr>
<td></td>
<td>Greater accountability</td>
<td>Teacher feels motivated because of accountability from new evaluations</td>
<td>5</td>
<td>&quot;It lets teachers know, 'Hey, I can’t just teach the lessons. They’re looking for more than this.'&quot;</td>
</tr>
<tr>
<td>Trust</td>
<td>Evaluation was fair</td>
<td>Teacher thinks his or her evaluation was fair</td>
<td>22</td>
<td>&quot;But overall, it was a fair – I think it was a fair observation.&quot;</td>
</tr>
<tr>
<td>Theme</td>
<td>Subcode</td>
<td>Description</td>
<td>Number of teachers&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Example Text</td>
</tr>
<tr>
<td>------------------------------</td>
<td>--------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>---------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Evaluation was subjective</td>
<td>Teacher thinks evaluation ratings are based heavily on rater's personal opinions rather than objective criteria</td>
<td>10</td>
<td>&quot;Even though they try to make it as objective as possible, there is always a subjectivity to it.&quot;</td>
<td></td>
</tr>
<tr>
<td>Evaluation was biased</td>
<td>Teacher thinks evaluation ratings are biased toward certain kinds of teachers</td>
<td>8</td>
<td>&quot;The new evaluation system is a way to get all the older teachers... to quit or give them bad reviews, then you get laid off.&quot;</td>
<td></td>
</tr>
<tr>
<td>Assessments invalid for evaluations</td>
<td>Teacher thinks student assessments are inappropriate for use in evaluation system</td>
<td>5</td>
<td>&quot;I'm not sure that the test that I'm required give here would be considered to be valid compared to other standardized tests.&quot;</td>
<td></td>
</tr>
<tr>
<td>Instructional goal setting</td>
<td>Teacher formally set a goal as part of the evaluation process</td>
<td>8</td>
<td>&quot;What I had to do was I had to come up with goals.&quot;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teacher was motivated by the clear expectations created by the evaluation system</td>
<td>5</td>
<td>&quot;I like to know what is expected of me... it keeps us all more involved in doing what’s best for the kids.&quot;</td>
<td></td>
</tr>
<tr>
<td>Effort and productivity</td>
<td>Teacher puts same/greater amount of time or energy into work because of new evaluations</td>
<td>12/5</td>
<td>&quot;We get very paranoid about, 'Oh, well, let me see if I can do more.'&quot;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tasks associated with evaluation take time away from other work activities</td>
<td>14</td>
<td>&quot;We had to work to figure out what the data was that we were going to use and so on.... that was time-consuming.&quot;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teacher reports increased anxiety or stress because of new evaluations</td>
<td>13</td>
<td>&quot;It's a lot of stress with a lot us with this new evaluation system.&quot;</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Teachers may have had multiple subcodes within the same theme.

<sup>b</sup> These subcodes consist of two similar subcodes codes with different valences.
### Appendix A: Overview of District Contextual Factors and Evaluation System Features

<table>
<thead>
<tr>
<th>Feature</th>
<th>District 1</th>
<th>District 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>State</td>
<td>Georgia</td>
<td>Massachusetts</td>
</tr>
<tr>
<td>Size/Urbanicity</td>
<td>Large suburb</td>
<td>Large city</td>
</tr>
<tr>
<td>Observations</td>
<td>6 observations per year, conducted by trained evaluators</td>
<td>Between 1 and 5 observations per year, depending on tenure and rating.</td>
</tr>
<tr>
<td>Feedback</td>
<td>Three conferences throughout the school year. Ratings (and optional comments or conferences) with each observation.</td>
<td>Required conferences following announced observations, other conferences optional. Conference required if summative rating is below Proficient.</td>
</tr>
<tr>
<td>Professional development</td>
<td>Teachers rated below Proficient placed on Professional Development Plan. State provides online “mini-modules” aligned with observation rubric. State recommends that professional development be linked to areas for growth identified during conferences.</td>
<td>All teachers on “Educator Plans” with varying levels of prescribed professional development. Teachers identify needed professional development and support through initial self-assessment.</td>
</tr>
<tr>
<td>Collaborative structures</td>
<td>Conferences can be group conferences. Professional development options include collaborative options, such as lesson study and co-teaching. Collaboration is a standard on the teacher evaluation rubric. Allows teachers to adopt school or team goals.</td>
<td>District exploring peer assistance and review. Educator Plans can include collaborative action plans. Collaboration is a standard on the teacher evaluation rubric. Allows teachers to adopt school or team goals.</td>
</tr>
<tr>
<td>Rewards and sanctions</td>
<td>Teachers with low ratings placed on Professional Development Plans. State plans to base merit pay on evaluation ratings.</td>
<td>Low ratings can result in dismissal. Proficient ratings qualify teachers for certain positions or additional responsibilities. High ratings are supposed to be accorded some sort of recognition.</td>
</tr>
</tbody>
</table>
Appendix B: Interview Protocol

1. Before I get into specific questions about teacher evaluations, I would like to get a general sense of you and your school. How would you describe your school — the people and programs — to someone who doesn’t know it?
   a. What is it like to teach there?
   b. How long have you been teaching?
   c. What grades and subjects do you teach?
   d. Do you have tenure? NO What is your professional status?

2. Prior to this past school year, when was the last time that you were formally evaluated under the old evaluation system? Was it at the same school? NO Was it in the same district?

3. Describe your most recent formal evaluation under the old evaluation system.
   a. Who formally observed and evaluated you?
   b. How often did your evaluator(s) observe you and give you feedback?
   c. Aside from being observed and attending pre- and post-conferences, did you participate in the evaluation process in any other way?

4. Describe how you were evaluated this past year.
   a. Who formally observed and evaluated you?
   b. How often did your evaluator(s) observe you and give you feedback?
   c. Aside from being observed and attending pre- and post-conferences, did you have participate in the evaluation process in any other way?

5. Which changes in the evaluation process would you say are most important to you? Why?

6. Describe a specific time this past year that you were observed and received feedback.
   a. Were there pre- or post-observation conferences? YES Describe how they went.
   b. Was this typical of the observations and feedback you’ve had before?
   c. Do you think the new evaluation system has changed the way that you were observed or the feedback you received? YES How so? NO Why not?

7. What was your response to the feedback you received?
   a. To what extent did the feedback accurately reflect your teaching?
   b. To what extent did you adjust your teaching in response to that feedback?
   c. Do you think that you responded to feedback you received differently from how you would have responded to feedback under the old evaluation system? YES How so? NO Why not?

8. Do you collaborate with other teachers at your school? YES
   a. Describe who you collaborate with and how.
   b. What sorts of things do you work on together?
   c. This year, have you changed who you collaborate with or how you collaborate with them?

9. Do you think that you work with your colleagues differently as a result of the new evaluation system? YES How so? NO Why not?

10. Do you think that the new evaluation system has changed the professional culture in your school? YES How so? NO Why not?
11. Do you think that the new evaluation system has changed how much time or effort you spend on your work? **YES** How so? **NO** Why not?

12. What resources does your school offer you to support your teaching?
   a. Does your school offer you or send you to professional development workshops?
   b. Does your school give you curricular materials or technology?
   c. Do you have common planning time with other teachers?
   d. Do you observe other teachers, or do other teachers observe you?
   e. Do you have anyone who you work with to discuss your teaching or to design lessons, such as a mentor teacher, specialist, department chair, AP or coach?

13. Are the resources you have access to this year the same as or different from the resources you had last year?
   a. Have you changed what resources you use? Why?
   b. Has your evaluator directed you to use any new resources or to collaborate with anyone new? **YES** What did they recommend, and what did you do?
   c. Do you think these changes in resources and your use of them are related to the new evaluation system?

14. Have you had any conversations about the new evaluation system with your colleagues? **YES**
   a. What did you talk about with them?
   **NO**
   b. Is there any particular reason why you haven’t discussed the new system with your colleagues?

15. What do you see as the purpose of the new evaluation system?

16. If it were entirely up to you, how would teachers be evaluated? **If different** Why do you think that would be better?

17. What do you see as the strengths and weaknesses of the new evaluation system?

18. Do you think that the new evaluation system is fair? Why?

19. Before we wrap up the interview, is there anything else that you would like to add?

Thank you – I have a few more quick questions, and then we will be finished.

20. Did you go through a traditional or alternative certification program?

21. Do you know the overall evaluation rating you received this past year? **YES** What rating did you receive?

22. How many students do you teach?

23. What is the approximate percent of English language learners in your classes?

24. What is the approximate percent of special education students in your classes?

25. What is the approximate percent of students in your classes who are on grade level?

26. Do you mind if I ask your age, ethnicity, and gender?