Lessons Learned from a Comprehensive Teacher Evaluation System:
An Instrumental Case Study

Heba Abdo
Christian Mathews
Carolyn Ross
Rutgers University
Abstract

Teacher evaluation has become an increasingly important topic since 2009 when policy, such as the Obama administration’s Race to the Top (RTT), and highly funded inquiry, such as the Gates Foundation’s Measures of Effective Teachers Study (MET), began. This led to widespread changes in states’ approaches to teacher evaluation and district-level implementation of new, often highly-punitive, systems. Though the Obama administration’s more recent Every Student Succeeds Act (ESSA) will repeal some requirements related to teacher evaluation, many states continue to implement stringent approaches while new cultures have permeated others (Sawchuk, 2016).

This research is grounded in a view that practitioner experience should inform educational policy. An instrumental case study (Stake, 1995) is used to describe lessons learned from one charter school’s attempt to implement a comprehensive teacher evaluation system. The study includes in-depth interviews with 2 administrators and 12 teachers. Findings reveal positive results for teacher learning and important lessons for sustaining leadership.

Keywords: classroom feedback, formative leadership, teacher assessment
Recent teacher evaluation reform efforts have recognized that school factors inhibit administrators’ ability to reliably evaluate teacher practice. For example, teachers often believe that their practice should not be open to scrutiny and administrators have been unable to differentiate levels of quality between teachers (Weisberg, et al., 2009). Weisberg et al.’s (2009) seminal study of teacher evaluation practices in twelve large districts across the US found that almost all teacher had been rated as “satisfactory” and that teachers receive very little usable feedback about their practice. Their study ushered in a new wave of reform related to teacher evaluation.

The most prominent goals of new research efforts in this area were to create teacher evaluation tools that could validly differentiate teachers along a continuum of practices from least to most effective, to understand ways that evaluators might implement these tools, and to inform how schools might use the data that the tools generate to make reliable decisions about teacher practice (Futernick, 2010; Kane & Staiger, 2012; Kimball & Milanowski, 2009). The reforms came more quickly than was typical of previous educational reforms, perhaps due to the involvement of high profile donors who raised awareness and seeded innovations in this area through grant making. Notably, Bill Gates, a renowned inventor and billionaire philanthropist, joined forces with Charlotte Danielson, a leader in teacher evaluation research, Harvard University, the Educational Testing Service (ETS), and other educational institutions to fund and develop the Measures of Effective Teaching (MET) Project in 2009. The MET Project implemented a comprehensive experimental design study with about 3,000 teachers in which they randomly assigned student rosters to teachers and collected many sources of data on teacher effectiveness, including classroom observation, achievement scores, and student surveys.
Through the course of the study, MET intends to identify the most valid evaluation measures, or combination of measures, and the most reliable processes with which to implement them. During the same year, the Obama administration introduced “Race to the Top,” (RTT) a federal program that aims to eradicate the achievement gap by investing in state-wide educational programs including teacher evaluation (U.S. Department of Education [USDOE], 2009). RTT provides many incentives for states that propose effective plans to implement teacher evaluation and other programs including $46 billion in monetary grants (USDOE, 2009) and waivers for a previous federal program, No Child Left Behind (Popham, 2013).

Not surprisingly, the introduction of RTT led to many teacher evaluation reforms across the United States including increased spending on tools (Chambers, et al., 2013; Garrison-Mogren & Gutmann, 2012) and the adoption of new or improved teacher evaluation systems by forty-three states (National Council on Teacher Quality, 2012; Rotherham & Mitchel, 2014). In addition, preliminary data show that some school districts have implemented more comprehensive teacher evaluation systems and may be more successfully differentiating effective and ineffective teacher practice (Aldeman & Chuong, 2014; Hamilton et al., 2014). Although the Obama administration’s Every Student Succeeds Act (ESSA), set to go into effect in 2017, will repeal some federal teacher evaluation requirements initiated in RTT, many states continue to encourage comprehensive systems.

The thrust of MET and RTT has also created a large research base that envisions teacher evaluation as a strategy to increase the effectiveness of the teaching field as a whole by ensuring that each classroom is staffed with an effective teacher. Yet researchers are divided as to whether teacher evaluation data should be used to make key human resources decisions such as teacher dismissal and differential pay (Gordon et al., 2006; Hanushek & Rivkin, 2006; Jacob,
Lessons Learned from a Comprehensive Teacher Evaluation System

2011; Tucker & Stronge, 2005) or to increase individual teacher competence through meaningful feedback and data-driven professional development (Bambrick-Santoyo, 2012a; Minnici, 2014; Papay, 2012). This study is situated within the latter belief, that teacher evaluation is most useful as a formative tool that supports teacher learning.

While researchers continue to debate the best tools, methods, and uses for teacher evaluation, many schools are still unable to implement basic systems that adequately differentiate between high, medium and low quality teachers. Most schools profess a view of teacher evaluation as central to both personnel decisions and teacher growth. Yet, in practice, research shows that many schools do not enact any approach well (Kraft & Gilmour, 2015).

Studying school implementation of teacher evaluation is an emergent and quickly growing field (e.g. Anast-May, et al., 2011; Bambrick-Santoyo, 2012a; Donaldson 2012; Mitchell & Purchase, 2015; Reinhorn, Johnson, & Simon, 2015; Sartain et al., 2011) and studies that specifically focus on formative approaches are sparse. Many describe what they believe to be best practices in this approach to teacher evaluation, but few have studied their implementation (Riordan et al., 2015). A review of literature on feedback to teachers conducted by Scheeler, Ruhl & McAfee (2004) found 208 articles published on feedback to teachers between 1970 and 2004; however, only eight of those articles focus on in-service teachers, while the rest describe pre-service teachers. Recently, researchers have written about formative approaches to in-service teacher supervision, but to date the area of investigation lacks coherence. More specifically, the literature has not adequately made recommendations about school-level programs and policies that support teacher evaluation, especially with regard to descriptive examples of systems that are performing effectively. As such, the research questions for this study are:
1. What school-level factors support implementation of a comprehensive teacher evaluation system in one highly-effective Washington D.C. Charter School?
   
   1a. What specific structures support teacher use of evaluation data to improve instruction?
   
   1b. In what ways has the teacher evaluation system supported teacher learning?

**Theoretical Framework: Formative Supervision**

This study explores the uses and outcomes of a comprehensive teacher evaluation system in one Washington D.C. charter school. Therefore the framework will focus primarily on structures and theories that support implementation of teacher evaluation at the school level. Further, the study design is associated with a belief that formative approaches to teacher evaluation are more promising than summative ones for increasing the effectiveness of individual teachers. This theoretical framework will describe the distinction between the two terms “summative” and “formative” in regards to teacher evaluation, then it will discuss “formative supervision” a new term in school leadership literature that describes the intersection between formative approaches and school leadership practices.

**Summative and Formative Approaches to Teacher Evaluation**

The expressions “summative” and “formative” are increasingly used in K-12 schools to refer to the ways that students are assessed, although Scriven first coined these terms in 1967 in reference to curriculum and teacher evaluation. Scriven (1967) notes that an evaluation tool always has the same purpose: to answer questions about the person (or program) that is being evaluated. Moreover, all evaluation tools use data collection and subsequent decision making as primary strategies. However, he also suggests that tools can be utilized summatively or
formatively (Chappuis & Chappuis, 2007; Scriven, 1967). Scriven (1967) defines summative evaluation as using a tool to assess whether results meet the stated goals. As summative uses of evaluation attempt to make a final decision, they are most often conducted when an event or process concludes. Conversely, formative evaluation leads to feedback during the course of the event or program, while “it is fluid” (Scriven, 1967, p. 236). Neither summative nor formative uses are inherently better than the other, and each has an important function dependent on the context and goals of the evaluation (Chappuis & Chappuis, 2007; Scriven, 1967).

Currently, the teacher evaluation debate has split among those who more strongly support summative uses, and others who value incorporating or fully implementing formative uses. Although not all researchers feel that an absolute choice must be made, most lean heavily in one direction or another. Many researchers, particularly those who currently influence national policy, support the use of teacher evaluation tools as summative assessments of teacher practice (for example: Gordon et al., 2006; Hanushek & Rivkin, 2006; Jacob, 2011; Stronge, 2013). These researchers often favor practices that use teacher evaluation data to reward effective teachers, such as differential pay and tenure, and to punish ineffective teachers, through termination or probation. This stance necessitates a belief that teacher performance is stable: that a teacher will continue to perform similarly over years and in varied contexts in order to warrant sanctions and rewards. While summative arguments pervade current teacher evaluation policy dialogue, this literature review only provides a brief overview of them because this study explores school-level outcomes in a school that utilizes a strong formative evaluation system.

Formative approaches are based on a belief that teacher evaluation should be used primarily as a feedback tool that helps teachers learn about and improve their practice. This can be a promising procedure, especially when linked to other teacher learning experiences, such as
coaching and professional development (Darling-Hammond et al., 2012b; Zepeda, 2007).

Researchers who support formative approaches to evaluation believe that teachers are capable of growing their practice with the support of targeted feedback and positive professional environments. Table 1 describes the two approaches in more detail.

Table 1

<table>
<thead>
<tr>
<th>Summative and Formative Teacher Evaluation Methods</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Uses in Teacher Evaluation</th>
<th>Summative</th>
<th>Formative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Final Teacher Effectiveness Determinations</td>
<td>Coaching</td>
</tr>
<tr>
<td></td>
<td>Incentive Pay</td>
<td>Professional Conversation Linked to Teacher Evaluation</td>
</tr>
<tr>
<td></td>
<td>HR Decisions</td>
<td>Professional Development Opportunities Linked to Teacher Evaluation Feedback</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Key Research in Favor</th>
<th>Bill &amp; Melinda Gates Foundation, 2013</th>
<th>Bambrick-Santoyo, 2012</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gordon et al., 2006</td>
<td>Danielson, 2011</td>
</tr>
<tr>
<td></td>
<td>Hanushek &amp; Rivkin, 2006</td>
<td>Minnici, 2014</td>
</tr>
<tr>
<td></td>
<td>Jacob, 2011</td>
<td>Papay, 2012</td>
</tr>
<tr>
<td></td>
<td>Stronge, 2013</td>
<td>Taylor &amp; Tyler, 2012</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Associated Views</th>
<th>Teacher skill-sets do not change over time.</th>
<th>Teachers continuously learn new skills.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Comparing observation results to student achievement yields valid assessments of teacher practice.</td>
<td>Teachers are capable of more than their observed practice may reveal.</td>
</tr>
<tr>
<td></td>
<td>Policy is necessary to change school ineffectiveness.</td>
<td>Teacher agency is central to teacher growth.</td>
</tr>
<tr>
<td></td>
<td>Teachers can be motivated by sanctions and rewards.</td>
<td>Professional knowledge emerges from practice and conversations about practice.</td>
</tr>
</tbody>
</table>

Table 1

**Formative Supervision: The Intersection of Theory and Leadership Practice**

As a new body of research, the uses and processes associated with effectively using teacher evaluation in formative ways are still emerging and lack coherence. Although disjointed, this study utilizes this area of the research to support the research design. It adopts Zepeda’s
(2007) concept of formative supervision to describe these practices, which Zepeda defines as the administrative act of connecting evaluation data to professional development practices in order to heighten teacher learning. This section of the literature review will introduce four elements that increase the success of formative supervision as collected from many sources: distributing evaluator caseloads, evaluator training, giving formative feedback, and linking professional learning to teacher evaluation data.

In order to positively enact formative supervision, evaluators need manageable caseloads (Kelley & Maslow, 2012; Kraft & Gilmour, in press; Reinhorn, 2013). With appropriate caseloads, observers may have more opportunities to observe each teacher. This is important because fewer observational episodes are less likely to capture teacher practice which may decrease validity (Jerald, 2012; Holland, 2005; Sartain et al., 2011); teachers also state that they view decisions about teacher practice based fewer observations as less valid (Donaldson, 2012; Reinhorn et al., 2015). In addition, when caseloads are high, principals report difficulty coordinating observation times, meeting deadlines, and scheduling pre- and post- conferences (Rotherham & Mitchel, 2014; Sartain et al., 2011). Furthermore, unmanageable caseloads lead to generic feedback (Halverson et al., 2004; Rotherham & Mitchel, 2014).

In order to implement effective formative processes that engage teachers and increase professional skills across schools, evaluators need better training. There has been a considerable amount of research on increasing reliability through training on teacher evaluation tools, yet there is a lack of training on how to use teacher evaluation results to create learning systems for teachers (Goe et al., 2012; Greenberg, 2015; Grissom et al., 2011; Herlihy et al., 2014; The New Teacher Project, 2012). Providing high-quality feedback is “more challenging than simply increasing the frequency of observations” (Aldeman & Chuong, 2014, p. 7). It requires that
Lessons Learned from a Comprehensive Teacher Evaluation System

evaluators understand the need for formative evaluation (Reinhorn, 2013) and know how to give constructive, actionable feedback. This may be why evaluators often request professional development specifically around stimulating feedback conversations (Dretzke et al., 2015; Reinhorn, 2013; Riordan et al., 2015; Rotherham & Mitchel, 2014; Sartain et al., 2011). Good training is also important for school leaders because their skills in evaluating teachers, their ability to create trusting environments within which to give and receive feedback, and their ability to understand new teacher evaluation policy and honestly inform teachers about it correlate with higher teacher commitment to teacher evaluation processes (Davis, Ellet, & Annuziata, 2002; Holland, 2005; Jacob & Lefgren, 2006; Kelley & Maslow, 2012; Lewis, Rice, & Rice, 2011; Milanowski & Heneman, 2001; O’Pry & Schumacher, 2012; Reinhorn, 2013; Riordan et al., 2015; Sartain et al., 2011; Tuytens & Devos, 2010). Evaluator training must also be sustained over multiple years. Many researchers have explored the 3-5 year time gap for teachers to implement new programing (Fullan, 2007; Reeves, 2009); yet, educational leaders also need targeted support during the initial 3-5 years of rollout, not simply at the initiation of new teacher evaluation programing (Derrington & Campbell, 2015; Dretzke et al., 2015; Murphy, Hallinger & Heck, 2013). Finally, evaluators need specific training to address the needs of novice teachers (Clayton, 2013; Darling-Hammond et al., 2012b), accomplished teachers (Donaldson, 2012; Halverson, Kelley, and Kimball, 2004), and teachers outside of their familiar areas of content (Donaldson, 2012; Kelley & Maslow, 2012; Murphy et al., 2013; Reinhorn, 2013; Rotherham & Mitchel, 2014). These needs are heightened the longer the gap between when the evaluator last taught and the time of the evaluation (Murphy et al., 2013).

Evaluators also need training on how to compose and communicate formative feedback. There is a wealth of guidelines that relate to best feedback practices from both the student
Lessons Learned from a Comprehensive Teacher Evaluation System

assessment and teacher evaluation literature. A common element across these is that in order for evaluation feedback to be educative, it must be presented during a time when learning is still fluid (Chappuis & Chappuis, 2007; Coggshall et al., 2012; Goe, 2013; Shute, 2008; Westerberg, 2013; Wiggins, 2012). In terms of engaging the learner, formative feedback should be constructive and supportive (Danielson, 2011; Ilgen, Fisher, & Taylor, 1979; Ford et al., 2015; Range, Young, & Hvidston, 2013; Scheeler et al., 2004; Shute, 2008) and shared during ongoing, responsive, two-way dialogues that specifically engender teacher self-reflection (Beerens, 2000; Danielson & McGreal, 2000; Garth-Young, 2007; Goe, 2013; Pritchett, Sparks, & Taylor-Johnson, 2010; Sinnema & Robinson, 2007; Trinter & Carlson-Jaquez, 2014; Westerberg, 2013; Wood & Pohland, 1979). There are many recommendations about constructing high quality feedback as well. Feedback that is specific is less likely to be viewed as useless or confusing (Williams, 1997). Coupling specific feedback with actionable steps can help learners to view skills as improvable by practice (Shute, 2008). In schools, where attrition rates may be as high as 30% (Ingersoll, Merrill, & Stuckey, 2014; McCreight, 2000), it may be easier to retain teachers if they believe that good teachers can develop with targeted feedback and practice (Ford et al., 2015). Feedback must also be complete and accurate (Behrstock-Sherratt et al., 2013; Coggshall et al., 2012; Goe, 2013; Hill & Grossman, 2013; Jerald, 2012; Westerberg, 2013). In order to provide accurate feedback, feedback should be given based on frequent observations of teacher practice (Hunter, 1988; Jerald, 2012; Marshall, 2013; Ovando & Harris, 1993; Westerberg, 2013; Zepeda & Kruskamp, 2012). Additionally administrators should use many illustrative examples drawn from specific observational evidence of student work and teacher actions and connect their feedback directly to the evaluation tool (Goe, 2013; Westerberg, 2013). Doing so helps teachers connect feedback to a picture of what good teaching looks like
so that they can understand how their present performance lines up with expectations and also
reinforces a common instructional language. Feedback should also be directly related to goals
that give direction on steps a teacher might take to reach those goals over time (Chappuis &
Chappuis, 2007; Coggshall et al., 2012; Danielson & McGreal, 2000; Donaldson, 2012; Fisher &
Ford, 1998; Ford et al., 1998; Frase & Streshly, 1994; Goe, 2013; Hattie & Gan, 2011; Holland,
2005; Malone, 1981; Milanowski & Heneman, 2001; Peña-López, 2009; Shute, 2008; Spillane,
Healey, & Mesler Parise, 2009; Trinter & Carlson-Jaquez, 2014; Westerberg, 2013; Wiggins,
2012). Ericsson (2009) who studies how experts in several fields develop, finds that expertise is
a result of concentrating on carefully selected, specific aspects of performance and refining them
through repetition and response to feedback. Similarly, teacher evaluation researchers
recommend 1-3 goals per teacher feedback session (Bambrick-Santoyo, 2012b; Westerberg;
2013).

Finally, much of the available research indicates the importance of linking teacher
evaluation to relevant and sustained professional development (Archibald et al., 2011; Coggshall
et al., 2012; Danielson & McGreal, 2000; Darling-Hammond, 2012; Darling-Hammond,
Marshall, 2013; Smylie, 2016; Zepeda, 2007). Danielson, whose evaluation framework is
widely used (Mooney, 2013), states that the framework’s most important use is as a “foundation
of a school or district’s mentoring, coaching, professional development, and teacher evaluation
process” (The Danielson Group, 2013). Others have found that in high performing schools,
creating a cycle of supervision and professional development leads to better instructional
practices (Mette et al., 2015) and increased student achievement (Shaha et al., 2015).
Just as for principals, professional development for teachers should initially focus on the purpose of evaluation tools and procedures (Heyde, 2013; Holland, 2005) and formal instruction on how to have meaningful conversations in relation to teacher evaluation data (Jerald, 2012; Murray, 2014; Sartain et al., 2011; Shaha et al, 2015). Additionally, professional development that is linked to teacher evaluation data should follow the recommendations of best practice from research on professional development. It should be relevant to each teacher’s needs (Contreras, 1999). This is heightened, but not automatic, when utilizing evaluation data and allowing teachers to self-identify learning goals (Almy, 2011; Anast-May, et al., 2011; Curtis & Weiner, 2012). Administrators must recognize that teacher work is characterized by constant and weighty decision making, (Ball & Cohen, 1999; Berry & Loughran, 2002; Danielson, 2012, 2015; Hunter 1988; Kazemi, Lampert, & Ghausseini, 2007; Tripp 2011). Thus, professional development should allow teachers to increase self-reflection by exploring problems of practice (Coggshall et al., 2012; Darling-Hammond, 2012; Darling-Hammond et al, 2012a). Professional development should also provide successful models of new teaching strategies, followed by practice and more feedback (Archibald et al., 2011). Special attention should be given to the needs of distinct groups including novices (Clayton, 2013; Darling-Hammond et al., 2012b) and high-performing teachers (Donaldson, 2012).

**Research Design**

This qualitative research study utilizes an instrumental case study design. As with other case study designs, an instrumental case study explores a particular phenomenon, within its context, over a bounded period of time (Creswell, 1998; Miles & Huberman, 1994; Stake, 1995; Yin, 2003). However, the design further uses a focus on the particular case to facilitate
understanding of a larger issue (Stake, 1995). The proceeding review described how many researchers have begun to call for formative approaches to evaluation, however few examples of such systems are accounted. As the effectiveness of evaluation for this purpose is highly dependent on school leadership and other contextual factors (Aldeman & Chuong, 2014; Davis et al., 2002; Davis et al., 2000; Howard & Gullickson, 2010; Reinhorn, 2013; Russell, 2002), this research will describe one system’s attempt to incorporate a formative evaluation system in order to provide guidance to other practitioners.

Given the research questions, the first task was to search for a case that would illustrate formative supervision practices. In order to be an appropriate selection, the site would need to incorporate administrator training on both tools and practices (Goe et al., 2012; Grissom et al., 2011; Herlihy et al., 2014; The New Teacher Project, 2012), frequent and ongoing teacher evaluations and feedback sessions (Jerald, 2012; Marshall, 2013; Westerberg, 2013; Zepeda & Kruskamp, 2012), and targeted professional development based on teacher evaluation results (Archibald et al., 2011; Coggshall et al., 2012; Danielson & McGreal, 2000; Darling-Hammond, 2012; Darling-Hammond et al., 2012a; Hill & Grossman, 2013; Holland, 2005; Marshall, 2013; Zepeda, 2007). The researcher sought an appropriate case through personal conversations and recommendations from experts in the field, such as the head of professional development at the Danielson Group. The final school site was found through an unlikely meeting with a charismatic school leader in Washington, DC. He spoke at length about very innovative ways that his school implements a comprehensive teacher evaluation system. For example, he described extensive administrator training to conduct formative teacher evaluation, including calibration walk-throughs and feedback calibration sessions wherein administrators watched post-conference videos together and then discussed strengths and weaknesses for further
practice. The principal also described supports for teachers’ use of their evaluation feedback including weekly observations and feedback conferences that were directly linked to professional development opportunities at the school and the use of classroom video to display “benchmark” teaching examples for each section of the teacher evaluation rubric.

“Brady” school is part of a network of charter schools in the Washington, DC area. The campus serves 480 students in grades PreK- 8. All students at the school are Washington, D.C. residents and many come from high poverty backgrounds; 98% of students receive free or reduced lunch. The student body is 96% Black and 4% Hispanic. The school employs 35 full-time teachers; six are first-year teachers and 16 have more than five years of teaching experience. The school employs teachers from alternate teacher preparation programs for underprivileged schools, including two teachers from Teach for America and eight from the Urban Teachers program.

Data Collection Procedures

Data collection included four tasks that occurred over the course of four months and two school visits. On day one of the first visit, the researcher conducted individual interviews with two building administrators (full population) and a focus group with six early childhood teachers. Focus group teachers were chosen based on a convenience sample that could be easily scheduled at the school. On day two, notes from both data sets were used to target interviews with six elementary and middle school teachers. Interviewed teachers were chosen from a maximally variant sample based on grade level, area of practice, and teacher preparation program.

Over the course of the next few months, data was collected virtually via the school’s internet portal. The researcher combed through teacher evaluation support programs like the Six Steps of Feedback tool (Bambrick-Santoya, 2012b), teacher self-reflection support tools,
professional development programs, videos of principal feedback sessions, and two commercially created teacher surveys that asked about teacher evaluation (K12 Insight, 2016).

During the final school visit, the six elementary and middle school teachers were re-interviewed to understand their learning related to teacher evaluation through the course of the year. These interviews included questions based on the initial visit and virtual data collection. The principal was also re-interviewed, which was an opportunity to reflect on emerging themes and to learn about new directions for the school.

**Data Analysis Plan**

Qualitative data analysis was done through an inductive process on Dedoose software through which the researcher chunked together small bits of data while looking for larger themes (Merriam, 2009; Creswell, 1998). This was followed by a continuous checking and rechecking of themes against new pieces of data to find a conceptual framework that explains the entire data set (Creswell, 1998; Merriam, 2009; Miles & Huberman, 1994). This framework was verified through the final interview with the school principal. The findings and discussion were constructed based on this final framework (Miles & Huberman, 1994).

**Limitations**

As is true for all case studies, the study is limited by the case by which it is bound, including the particular context, participants, and time. Nonetheless, by purposely sampling a school that is working to apply best practices, findings from the study may be of use in other school contexts. A further limitation of case study research is that the central tool in the research is the researcher herself. While the researcher makes important decisions that allow for in-depth study, she may also bring her own biases that ultimately impact the study (Creswell, 1998; Merriam, 2009). In terms of this study, the researcher’s bias towards formative
supervisory practices might skew participants’ responses during collection and coding during analysis. Also, although there is a purposeful, maximally-variant sample, the sample size is small, representing about one-third of the school population. Finally, the case study site was located at a distance from the researcher’s university. This limited the study in two ways. As a complete outsider of both the school and the social-political context in which the school is located, the researcher was constrained in regards to understanding participants’ shared history with the teacher evaluation system and the inner workings of the school. It also limited the ability to observe key relationships and teacher evaluation practices in real time.

Findings

As the research methods describe, Brady School was chosen as a research site because of its atypical commitment to formative evaluation approaches. However, during the course of data collection, findings changed dramatically as the school lost some key personnel in the middle of the year. This impacted the outcomes of the study significantly and provides important guidance for sustaining fidelity in teacher evaluation systems. This section will describe findings including specific structures that support teacher evaluation practices at Brady School; outcomes, including a brief description of teacher learning; and important lessons learned from leadership’s attempt to sustain change over the course of several years and new staff.

Support Structures

During initial conversations, the school principal, “Andre” described a school in which all building leadership were very invested in the teacher evaluation system. This investment was structured into the areas identified in the theoretical framework as necessary for formative supervision: (a) distributed evaluator caseloads, (b) evaluator training, (c) a focus on formative feedback, and (d) linked professional learning. In addition to the four areas identified in the
Lessons Learned from a Comprehensive Teacher Evaluation System

literature, the school employs two home-grown structures: (e) a continual growth-model, and (f) integration of other policies and procedures with teacher evaluation practices. Interviews with stakeholders confirm that many of these structures are in place and seem to be working well.

Andre is resolute that “every leader instructional leader in the building (should) spend at least 80% of their building schedule inside the classroom” (Principal, Interview 1). To facilitate the effectiveness of this leadership requirement, Andre begins each year by distributing evaluator caseloads and designating a conducive weekly schedule. In terms of distributing caseloads, the core instructional leadership team, “and when I say instructional leader, it’s me, the two Assistant Principals, the SPED (special education) coordinator, my ELA (English, Language Arts) coach, the math coach and the IB (International Baccalaureate) coach. So, seven people,” divides all the teachers between themselves so that each has a “core of six teachers” to work with (Andre, Principal, interview 1). Leo, another administrator confirms a “good (case)load” of seven (Interview). Instructional leaders have many other roles, and may observe teachers outside of their core; for example the math coach works with all math teachers in the building. However, the designated leader is responsible for conveying most of the team’s feedback to his/her assigned teachers (this process is described later in more detail). To further support the 80% requirement, the team creates a weekly schedule so that each leader “observes their core teachers at least 3 times a week and they meet with them at least once a week” (Andre, Principal, interview 1). Other administrative tasks, such as duty schedules, departmental meetings, administrative team meetings, etc. are then worked around that core schedule.

In order to support such a weighty emphasis on observation, the school offers many training opportunities for leaders such that the school often met or surpassed the literature’s recommendations about administrative training. To begin with, the district enrolls all core
administrators in a privately operated leadership institute that introduces leaders to the skills and processes necessary for positive evaluation (Leo, Interview). At the school level, the principal follows up “at the beginning of every year …(with) a professional development with my leadership team around the Six Steps of Feedback” (Andre, Principal, interview 1), which is a structured “20-minute” observation feedback process (Bambrick-Santoyo, 2012b). Most importantly, the administrative team meets each week “for 30 or like 45 minutes” (Leo, Administrator, Interview) and, among other things, spends time norming the observation and feedback process. On some weeks, the whole team will conduct a walk through in a random classroom and talk about the instructional strengths and areas of need. Other weeks, they will watch a video-taped post conference of an administrator giving feedback to a teacher and discuss ways to strengthen the feedback. Norming helps administrators understand “what the expectations were, what we're supposed to do, what it was supposed to look like, and how often we had to do it” (Leo, Administrator, Interview) and “serves as a learning tool to make sure that we're on the same page” (Andre, Principal, Interview 1). Administrators continue to refine the norming process each year. The outcomes have been more alignment across the administrative team and more buy-in as leaders are given continual practice opportunities.

Teacher training, however, has not been as successful. The school “spen(t) a lot of time at the start of the school year just going over the different components (of its teacher evaluation rubric” (Kathy, Middle School, Interview 2) so that teachers “are completely clear, like crystal clear on how they are evaluated as professionals” (Andre, Principal, Interview 2). Teacher surveys reflect that 82% of surveyed teachers “know the criteria that will be used to evaluate (their) performance as a teacher” (K12 Insight, 2016). Yet, the school did not spend as much time training teachers on the components and outcomes of the formative evaluation system.
Teachers were unaware of many components; for instance, no teachers understood references to the Six Steps of Feedback, not even a department head who regularly observes and gives feedback to colleagues.

The observation schedule and administrative training set a precedence for frequent feedback sessions, but it is the school’s commitment to a growth mindset that allows for a formative approach. Many teachers talked about support to work on goals until they are met, although results were often mixed, as will be discussed below. Teachers confirm frequent formative check-ins. For example, a middle school teacher describes the process as follows: “so there’s formal observations in which you will stay for almost the entire block, and then there’s short quick check-in observations where he may stay for 15 minutes… I would say at least multiple times throughout the week…and after he observed, he will schedule a follow-up interview time in which he will go over his notes compared to how you think your lesson went. And then from there we will discuss next steps, or what work to do if there’s an improvement” (Emily, Interview 1). An elementary teacher confirms weekly check-ins and adds “so say for instance it’s not one of our weekly meetings we can just reach out to them and they can come in and they observe us as well” (Wendy, Interview 1). In addition, many teachers spoke about the feedback itself being very formative in nature. Feedback was usually immediate: “within 5 minutes with just the notes that he was taking ... And then he would either have a meeting later on that day or the following day to discuss everything” (Heather, Elementary, Interview 2). Feedback helped teachers to work towards goals: “it’s a different dynamic. I like just the whole goal setting and everything” (Wendy, Elementary, Interview 1), and develop mastery over time: “the feedback is just bite-sized and the expectation is that they’re going to get to the end point in a gradual way” (Leo, Administrator, Interview). Further, administrators asked teachers to reflect
on their classrooms: “typically (they) ask you first for you to be reflective and talk about how you thought the lesson went and rate yourself on the rubric” (Emily, Middle School, Interview 2). This was supported by innovative programs such as classroom video: “they videotape me practicing that strategy and they gave me feedback practice on it” (Olivia, Elementary, Interview 1). Another teacher explains: “she would ask you over the weekend to watch the video and come with your own glows and grows. She would do the same so when we came to the meeting we would both have something to talk about. And she would have the video available so she could go back to specific spots and show it what you had been doing” (Lisa, Early Childhood, Focus Group). Finally, feedback always included opportunities for guided practice. For example, in one video of a feedback session, the administrator helped the teacher to create a minute-by-minute break down of a 45-minute literacy block in order to meet a larger goal of increasing rigor in the classroom. In all, these findings reflect a school-wide alignment with best practices in formative feedback as described in the proceeding theoretical framework.

The school also has many supports in place to link professional learning opportunities to teacher evaluation feedback. This process begins in August when administrators work with teachers to collaboratively choose three individual teaching goals for the year. Next, administrators use multiple data sources to decide upon PD offerings that are responsive to teacher needs. Data sources included teacher goals (Andre, Principal, Interview 1), teacher surveys (Wendy, Elementary, Interview 1), observation data (Angela, Early Childhood, Focus Group; Andre, Principal, Interview 1 & 2; Leo, Administrator, Interview 1), and teacher requests (Angela, Early Childhood, Focus Group). Usually professional learning opportunities are offered right away. The school offers weekly “professional development Thursdays- our students leave early … and that's a time for our teachers to come together across departments and
grade levels and schools (Andre, Principal, Interview 2). The administrative team uses its weekly Monday meetings to structure those Thursdays: “when we come one of the things we go over is what our plans are for the week and what instructional components we have for each of our caseloads, and then based on that we identify a trend...we then develop PDs for Thursdays on that trend” (Leo, Administrator, Interview). In addition, administrators might offer a PLC to support one department: “being a house leader, if I felt like there was something we needed to work on I would take that information back to her and then she would find a PD to support it or she would conduct a PLC” (Angela, Early Childhood, Focus Group). Unfortunately, though these supports were in place at the beginning of the year, PD and PLCs stopped mid-year “because State testing is coming up” (Wendy, Elementary, Interview 1).

Two unique findings from Brady School support the teacher evaluation program including administrative commitment to an open-dialogue and honest stance of inquiry about their practice and administrator integration of new policies and systems. Teachers appreciate that administrators are open to their ideas and complaints: “that's one thing, like, they're pretty open to if I recommended that that’s something that we should utilize” (Emily, Middle School, Interview 2). Administrators are also constantly seeking ways to improve by looking critically at sources of data. For example, the principal learned that one administrator was not fully supporting teachers through the end of year survey. This “enlighten(ed) (him) to do more frequent surveys.” He hoped that by doing so he could “hold administrators more accountable in the moment, and so that the teachers can hold (him) accountable for holding administrators accountable” (Andre, Principal, Interview 2). This quote is a powerful example of Andre’s commitment to both the importance of formative systems and to the positive experiences of his teachers. Such a stance helps to increase teacher buy-in.
In addition, the school incorporated many programs and structures into the teacher evaluation system, continuously refining each piece. Some examples include teacher binders that helped teachers track progress towards their three yearly goals (Heather, Elementary, Interview 2), aligning Washington DC’s required CLASS evaluation for early learners with the schools tool (Lori, Early Childhood, Focus Group), using classroom data to identify the need for more guided reading instruction, then hiring a guided reading coach, finding programs like the National Academy for Advanced Teaching to support high performing teachers, and hiring teachers from the Urban Teachers Program because the program utilizes a similar evaluative approach. Many teachers also appreciated that administrators provided coverage to visit other teachers’ classrooms as a model for more effective practice (Heather; Kathy; Pamela; Olivia; Wendy). A highlight of Brady’s programming is the school’s Google tracking system that was created a few years ago when administrators noticed that teachers felt overwhelmed by the sometimes conflictual feedback they were receiving from administrators:

“and for them, it was like, ‘I'm getting all these people in my classroom and there's all these different things to do, and now I don't know what your priorities are and what to do first.’ And it was very confusing. And so we wanted to create a system where we could still maintain that frequency of visits but the messaging and what is being asked of you was very fluid across the roles. And this is why we created our Google Doc... For example, because I'm going to go into every classroom at any time … I want to make sure that when I go in, that I'm always going to leave them with something because teachers don't necessarily appreciate you coming in and observing and then never talking to you again about what they saw… But for me so a) I want to be able to walk in a room and know what they're focusing on with their coach and the Google Doc was a way for me to be able to do kind of a spot-check to know, okay when I'm walking in I know they're focusing on right now … so now (it) just give me a lens to focus so I'm not giving them something new on their plate or making your life more stressful. And b) it's also silently messaging to them that were communicating with each other on what your focus is and how we're pushing you to grow on that. And the teachers so much appreciate that because it's like they know we're talking … and so when I'm giving you feedback I'm only giving you feedback around that and if I'm giving you
feedback around something else it's like something that can be fixed immediately... And so we found that was a lot more fruitful ... It's just a lot more planned and structured (Andre, Principal, Interview 1)

Over the years, the FIOT tool has evolved to streamline communication. It also allows the principal to hold all other administrators accountable for keeping up with the observation schedule. In addition to streamlining efforts at the Brady campus, the whole district now mandates the FIOT system, and it has been picked up by a national principal preparatory program (Andre, Principal, Interview 1).

Outcomes

Many of the study’s findings can inform the work of practitioners and policy makers nationwide. The outcomes section will briefly describe findings about teacher learning, school culture, and sustainability that are of interest to wider formative supervisory practices. This is a rich data set, and many more findings could be covered in other works.

In all, teacher outcomes are mixed. Some sampled teachers benefitted from feedback, while others felt that the process did not impact their instruction. This indicates that there is much more to teacher evaluation than the tools and processes. Each stakeholder brings a unique set of skills, abilities, and dispositions. When thinking about systems to support teacher learning, it is important to think through how these factors mix with procedural factors. The following vignette will illustrate a very positive case example and will be followed by a discussion of how it relates to some of the themes that seem to be true for many teachers.

“Heather” is a seasoned teacher. She has worked with kindergarten students in 3 different Washington D.C. schools over the past ten years. She knows curriculum, she knows kids, she knows pedagogy, and she loves teaching. Here, she describes her ninth year of teaching, which was her first year as a kindergarten teacher at Brady School:
“So (the KG students) came in being able to read. Most of them could add. Some of them knew how to subtract. And I'm like, okay, so I have to scrap everything that I thought kindergarten was about. … And so I went to administration here and I said I really love my job. I really want to stay here, however I don't know what this looks like any more… I'm not sure what to do. And so, I had to learn that it wasn't a negative thing. I had to learn that I could open up and ask questions and it wouldn't reflect on me as being a bad teacher, but it actually looked better that I was asking those questions. And so they immediately paired me up with another teacher that was here. I did some observations here. I went to another school for observations. I went to several off-site PD's. And I really felt like they equipped me with what I needed to become this rigorous teacher that they were expecting me to be... I went to Ms. Angelo (Elementary AP) at the time. And Ms. Angelo actually came in on a Saturday. And she said, “Give me your lesson plan. What are you planning on teaching on Monday? What questions are you going to ask?” I'm like, “this is what I'm going to ask.” And she literally helped me dissect the entire thing. And then she said, “if it makes you feel any better, I will come in and watch you on Monday. I will present a similar lesson on Tuesday. And then I will come back and watch you on Wednesday.” So it wasn't like okay let's go- if you fail you fail- she actually, like- and I was so impressed that she came in Saturday morning and sat and literally cleared the table. We made charts, we did everything. And she walked me through how the lesson should look and how to become more rigorous. And that spoke measures about Brady to me. (Heather, Elementary School Teacher, Interview 2)

Clearly, this vignette describes a high level of commitment to students from both the administrator and the teacher. This was not everyone’s experience of Brady School, however Heather’s experiences were not uncommon. Teachers that had similar characteristics were also able to benefit from Brady’s teacher evaluation system. Teachers who were novice teachers, or novice in a particular element, like Heather, reported positive effects. For example, Kathy describes her best feedback occurring during her “first year teaching. I had lower rankings in certain areas and a lot of it had to do with the rigor or the timing of questioning… so I can remember them giving me specific help and modeling what that looks like for me if they would work well in certain areas” (Middle School, Interview 2). Teachers who were self-reflective or able to specify an area of concern, like Heather, were also well-served. For example, Wendy, an
elementary teacher, needed specific guidance in question stems for reading and was immediately supported. Along the same lines, teachers who were proactive in asking for help, like Heather, were more successful: “If you're proactive they'll support it as long as it benefits your students” (Emily, Middle School, Interview 1). While these are very positive outcomes, not all teachers are able to self-identify needs or be so transparent about their shortcomings, so the school needs to look at ways to support all teachers. Another area of concern was that high-performing teachers felt that the evaluation system did not adequately support them. They were more likely to get generic feedback: “he said, “oh you're doing great.” Well I don't feel like I'm doing great. I feel like I'm drowning in here” (Denise, Middle School, Interview 1). Further, even when they were given areas to improve upon, they were not given resources or support tools: “there hasn't been that much push to get you great. It's kind of been like you have it keep going and it will come” (Kathy, Middle School, Interview 2). While the school did provide high-performers with opportunities to attend leadership trainings to “develop (other) adults in the building” (Kathy, Middle School, Interview 2), administrators did not help grow these teachers’ practice. As Emily succinctly concludes: “of course administrators give their time and resources …to those who are most in need, but I think it is still extremely vital and necessary that whether you been teaching for 15 years or 20 years you're still constantly being … observed because that’s how you grow and that’s how you become great in your craft (Emily, Middle School, Interview 2).

Heather’s vignette also reveals a lot of positive aspects of the school culture around evaluation. First, Heather felt quite safe to open her classroom up to such scrutiny, even as a new teacher in the building. Other teachers also communicated professional safety: “so when she's coming in, you're not coming to find wrong. You're actually coming to support my instruction and to kind of like give me tips and feedback, or even just to see or to encourage me
Lessons Learned from a Comprehensive Teacher Evaluation System

and let me know that I'm on the right track. So it doesn't feel like somebody standing over you judging your work. It's almost like a welcome presence” (Pamela, Early Childhood, Focus Group). Second, Heather’s vignette demonstrates that the school’s focus on high professional standards pushes teachers to have a sense of urgency. The principal also felt that the focus on growth held teachers accountable to high standards (Interview 1). Finally, Heather’s vignette demonstrates buy-in from the teacher and the administrator. Other teachers describe that because the system felt fair, they felt that engaging with it would be good for their practice: “for me, my goal was appropriate. That’s how I knew she was actually in my class because she knew what I needed to improve on” (Angela, Early Childhood, Focus Group). Others discussed how the school rallied their buy-in towards teacher evaluation: “She made us all stakeholders and we all felt like we were part of the change that she wanted to see for the school. So we could all get behind and support it” (Pamela, Teacher, Focus Group). Administrators also talked about buying-in to the evaluation system, especially after preliminary data was showing a shift in teacher culture and student achievement.

However, there were also many findings about supports that were still needed for school culture. The first, which is also found throughout teacher evaluation literature, is that the tools are not enough to increase teacher practice. When administrators could not give in-person feedback, evaluation was not a successful practice. This is a continued area of concern for the school. Second, some teachers felt that the feedback was unfair, either because they were not given the tools necessary to address the change: “one (area I was marked down for) was more like professional opportunities, which I've asked for and I haven't received, so I kinda feel like you can't really mark me down for that because you haven't given me any (Denise, Middle School, Interview 2) or because they felt it wasn’t reflective of their teaching: “quite honestly the
formal feedback like that rubric that they do here is not helpful at all for me...I'll score well in something and I'm like, “no, that's an area that I know that I'm not doing as well as I could be in” and sometimes it's a surprise cause I'll get knocked on something- not knocked – but I’ll get a lower score on something (Kathy, Middle School, Interview 1).

Most importantly, because there was a shift in administration mid-year, the school site offers many lessons about sustaining teacher evaluation culture. This includes two important factors: training new leaders and maintaining conditions through turbulence. This is the third year of the school’s teacher evaluation initiative, and preliminary results were showing some really positive changes. At the beginning of the year, the school hired two new assistant principals (APs), and the old APs were promoted to principal positions throughout the district. In October, the middle school AP was pulled to cover a principal position in another school, and in March the elementary AP was also pulled. Neither of the open positions were filled. Teachers reported huge changes between the first two and the third years.

The first contributing factor was that new administrators were not trained with the same level of fidelity as first cohort of leaders. For example, they did not attend the leadership institute. As a result, the AP who was new to the district, and had never seen the evaluation process in action, was trying to implement support, but had a different “perspective of a check-in or feedback conversation with a professional is. It was not what they expected, and not what they got before” (Andre, Principal, Interview 2). The second factor was that when administrators were not replaced, the workload became overwhelming. Over time, many programs were lessened or forgotten. For example, many teachers did not receive the minimum of three observations, and, with the exception of the lowest-functioning teachers, no one was receiving consistent feedback. This led to a lot of teacher frustration: “this past year is not a good model
for formal feedback” (Kathy, Middle School, Interview 2) and, in some cases, teacher complacency: “so there started to be some complacency- nothing drastic, but still people being a little more laid-back than they should be” (Leo, Administrator, Interview).

These factors led to some really important lessons learned. The principal now feels that it is very important to sustain “systems and procedures and everything still go like it needs to go even though the people aren't there” (Andre, Principal, Interview 2). To ensure this work, many actions are necessary. First, on the district level, there is a need for better planning so that there can be consistency of roles and a more efficient mechanism for hiring new staff. Second, the district is looking at ways to standardize principal induction in relation to teacher evaluation procedures. Third, the principal has recognized his own need for accounting systems and personnel by surveying stakeholders and checking processes more frequently (previously quoted). Finally, the principal recognized a need for distributed leadership, so that if personnel is lost, systems are “safeguarded.” Part of that work would include a more realistic look at what is possible when systems experience turbulence: “I didn't take anything off of my plate that would get me into the classroom more, which is absolutely what I should have done, and would have done, and will do in the future. Just really redistributing some of my responsibilities that I can give to others, so that I can spend more time in the classroom” (Andre, Principal, Interview 2).

Discussion

The results reveal several important findings regarding setting up and sustaining formative teacher evaluation systems. These lead to many important implications. At the school site, many positive outcomes have already arisen because of this action research. For example, during summer teacher training, all staff were introduced to the feedback process including the
Six Steps of Feedback. Department heads were further trained on coaching using this framework (Andre, Principal, Interview 2). Also, given feedback from this study, the school has incorporated more opportunities for teacher networking. Last year, teachers “only got together like four times throughout the year, and this year it's like fourteen times. (Andre, Principal, Interview 2). As previously mentioned, the principal has also committed to more frequent surveys of staff. The school will also set up a more tiered approach to observation that is more supportive of high-performing teachers, including new dedicated staff that will be “responsible for pushing the thinking and learning and the practice of those mid-tier teachers” (Andre, Principal, Interview 2).

There are also wider implications for all schools. First, there are many “quick fixes” to improving formative systems, such as developing a professional library to immediately support goals. However, to sustain change, higher-order programs will be necessary. For teachers, schools will need to target teachers across the learning continuum, increase teacher collaboration, and invest in mentorship programs. For leaders, schools need to carefully distribute leadership tasks to make room for the difficult and time-consuming work of formative evaluation. Perhaps most importantly, principal induction program are a necessary component when implementing demanding formative systems. Principals need training in procedures, but they also need to understand the more nuanced work of supporting teacher thinking. Generally, schools who are moving towards more formative systems must recognize that such systems require many moving pieces. It is not a matter of just training, or just providing systems, or just hiring good leaders, but really of creating synergy and consistency, and constantly reevaluating progress.

There are also many implications for policy. First, states should consider adding the study of teacher evaluation processes to requirements for principal licensure. As this study’s
findings show, creating learning systems requires a huge shift in the ways that leaders approach evaluation, which could be enhanced through extended study. Second, we need communal solutions that empower leaders to do this work more thoroughly. This may include restructuring the role of principals, or reevaluating time-consuming state requirements, or establishing another way to support principals in implementing time-consuming formative systems. Finally, principals in high-needs areas need different types of support to more effectively lead in more dynamic systems and to support teachers to excel with more troubled school populations.

Finally, for research, more longitudinal studies of positive case examples will be needed to provided more guidance to schools and researchers seeking to improve this area of leadership. The goals of this research might also be to develop tools and practices that support teachers across the growth continuum. Additionally, more research is needed to understand how principals hone the craft of giving feedback and supporting teachers so that these qualities can be replicated in principal preparation programs.

References


Lessons Learned from a Comprehensive Teacher Evaluation System


Lessons Learned from a Comprehensive Teacher Evaluation System


Lessons Learned from a Comprehensive Teacher Evaluation System


Lessons Learned from a Comprehensive Teacher Evaluation System


Lessons Learned from a Comprehensive Teacher Evaluation System


