

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

Heba Abdo  
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 wide-spread 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, 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**

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.

However, 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. Three elements emerge as central to this practice:

distributing evaluator caseloads, evaluator training, 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).

Evaluators also 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 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). Evaluators also need training on how to compose and communicate formative feedback.

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, Amerein-Beardsley, & Haertel, 2012a; Hill & Grossman, 2013; Holland, 2005; Looney, 2013; 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).

### **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; Cogshall et al., 2012; Danielson & McGreal, 2000; Darling-Hammond, 2012; Darling-Hammond et al., 2012a; Hill & Grossman, 2013; Holland, 2005; Marshall, 2013; Zepeda, 2007). The final school site incorporates many of these findings. For example, the principal describes very innovative ways that his school implements a comprehensive teacher evaluation system such as 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.

### **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.

### **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.

**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 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 provide 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

- Aldeman, C., & Chuong, C. (2014). *Teacher evaluations in an era of rapid change: From “unsatisfactory” to “needs improvement.”* Washington, DC: Bellwether Education Partners.
- Almy, S. (2011). *Fair to everyone: Building the balanced teacher evaluations that educators and students deserve. Teacher quality.* Washington, DC: The Education Trust.
- Anast-May, L., Penick, D., Schroyer, R., & Howell, A. (2011). Teacher conferencing and feedback: Necessary but missing! *International Journal of Educational Leadership Preparation*, 6(2), 2-9.
- Archibald, S., Coggshall, J. G., Croft, A., & Goe, L. (2011). *High-quality professional development for all teachers: Effectively allocating resources. Research & policy brief.* Washington, DC: National Comprehensive Center for Teacher Quality.
- Ball, D., & Cohen, D. (1999). *Toward a practice-based theory of professional education. Teaching as the learning profession.* San Francisco, CA: Jossey-Bass.
- Bambrick-Santoyo, P. (2012a). Beyond the scoreboard. *Educational Leadership*, 70(3), 26-30.
- Bambrick-Santoyo, P. (2012b). *Leverage leadership: A practical guide to building exceptional schools.* John Wiley & Sons.
- Beerens, D. R. (2000). *Evaluating Teachers for Professional Growth: Creating a Culture of Motivation and Learning.* Corwin Press, Inc.: Thousand Oaks, CA.
- Behrstock-Sherratt, E., Rizzolo, A., Laine, S. W., & Friedman, W. (2013). *Everyone at the table: Engaging teachers in evaluation reform.* New York, NY: John Wiley & Son.

- Berry, A., & Loughran, J. (2002). Developing an understanding of learning to teach in teacher education. In J. Loughran & T. Russell (Eds.), *Improving teacher education practices through self-study*, (pp. 13-29). New York, NY: Routledge Falmer.
- Chambers, J., Brodziak de los Reyes, I., & O'Neil, C. (2013). *How much are districts spending to implement teacher evaluation systems?* Santa Monica, CA: RAND Corporation.
- Chappuis, S., & Chappuis, J. (2007). The best value in formative assessment. *Educational Leadership*, 65(4), 14-19
- Clayton, C. (2013). *Understanding current reforms to evaluate teachers: A literature review on teacher evaluation across the career span*. New York, NY: Pace University.
- Cogshall, J. G., Rasmussen, C., Colton, A., Milton, J., & Jacques, C. (2012). *Generating teaching effectiveness: The role of job-embedded professional learning in teacher evaluation. Research & policy brief*. Washington, DC: National Comprehensive Center for Teacher Quality.
- Contreras, G. L. H. (1999). Teachers' perceptions of active participation, evaluation effectiveness, and training in evaluation systems. *Journal of Research and Development in Education*, 33(1), 47-59.
- Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five designs*. Los Angeles, CA: Sage Publications.
- Curtis, R. & Wiener, R. (2012). *Means to an end. A guide to developing teacher evaluation systems that support growth and development*. New York, NY: The Aspen Institute.
- Danielson, C. (2011). Evaluations that help teachers learn. *The Effective Educator*, 68(4), 35-39.
- Danielson, C. (2012). Observing classroom practice. *Educational Leadership*, 70(3), 32-37.

- Danielson, C. (2015). Framing discussions about teaching. *Educational Leadership*, 72(7), 38-41.
- Danielson, C., & McGreal, T. L. (2000). *Teacher evaluation to enhance professional practice*. Alexandria, VA: Association of Supervision and Curriculum Development.
- Darling-Hammond, L. (2012). The right start: Creating a strong foundation for the teaching career. *Phi Delta Kappan*, 94(3), 8–13.
- Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012a). Evaluating teacher evaluation. *Phi Delta Kappan*, 93(6), 8-15.
- Darling-Hammond, L., Jaquith, A., & Hamilton, M. (2012b). *Creating a comprehensive system for evaluating and supporting effective teaching*. Stanford, CA: Stanford Center for Opportunity Policy in Education (SCOPE).
- Davis, D. R., Ellett, C. D., & Annunziata, J. (2002). Teacher evaluation, leadership and learning organizations. *Journal of Personnel Evaluation in Education*, 16(4), 287-301.
- Davis, D. R., Pool, J. E., & Mits-Cash, M. (2000). Issues in implementing a new teacher assessment system in a large urban school district: Results of a qualitative field study. *Journal of Personnel Evaluation in Education*, 14(4), 285-306.
- Derrington, M. L., & Campbell, J. W. (2015). Implementing new teacher evaluation systems: Principals' concerns and supervisor support. *Journal of Educational Change*, 16(3)305-372.
- Donaldson, M. L. (2012). *Teachers' perspectives on evaluation reform*. Washington, DC: Center for American Progress.
- Dretzke, B. J., Ingram, D., Peterson, K., Sheldon, T., Wahlstrom, K., Baker, J., Crampton, A., Farnsworth, E., Zhi, A., Lim, H., & Yap, S. (2015). *Minnesota state teacher development*,

- evaluation, and peer support model evaluation report*. Roseville, MN: Minnesota Department of Education.
- Ericsson, K. A. (2009). Enhancing the development of professional performance: Implications from the study of deliberate practice. In A. Ericsson (Ed.), *Development of professional expertise: Toward measurement of expert performance and design of optimal learning environments* (pp. 405-431). New York, NY: Cambridge University Press.
- Fisher, S. L., & Ford, J. K. (1998). Differential effects of learner effort and goal orientation on two learning outcomes. *Personnel Psychology, 51*(2), 397.
- Ford, T. G., Van Sickle, M. E., Clark, L. V., Fazio-Brunson, M., & Schween, D. C. (2015). Teacher self-efficacy, professional commitment, and high-stakes teacher evaluation policy in Louisiana. *Educational Policy, 29*(3), 1-47.
- Frase, L. E., & Streshly, W. (1994). Lack of accuracy, feedback, and commitment in teacher evaluation. *Journal of Personnel Evaluation in Education, 8*(1), 47-57.
- Fullan, M. (2007). *The new meaning of educational change*. New York, NY: Teachers College Press.
- Futernick, K. (2010). Incompetent teachers or dysfunctional systems? *Phi Delta Kappan, 92*(2), 59-64.
- Garrison-Mogren, R., & Gutmann, B. (2012). *State and district receipt of recovery act funds. A report from charting the progress of education reform: an evaluation of the recovery act's role*. Washington, DC: National Center for Education Evaluation and Regional Assistance.

- Garth-Young, B. L. (2007). *Teacher evaluation: Is it an effective practice? A survey of all junior high/middle school principals in the state of Illinois*. (Doctoral Dissertation, Loyola University).
- Goe, L. (2013). Can teacher evaluation improve teaching? *Principal Leadership*, 13(7), 24-29.
- Goe, L., Biggers, K., & Croft, A. (2012). *Linking teacher evaluation to professional development: Focusing on improving teaching and learning. Research & policy brief*. Washington, DC: National Comprehensive Center for Teacher Quality.
- Gordon, R. J., Kane, T. J., & Staiger, D. (2006). *Identifying effective teachers using performance on the job*. Washington, DC: Brookings Institution.
- Greenberg, M. (2015). Teachers want better feedback. *Education Week*, 35(1), 47-48.
- Grissom, J. A., & Loeb, S. (2011). Triangulating principal effectiveness how perspectives of parents, teachers, and assistant principals identify the central importance of managerial skills. *American Educational Research Journal*, 48(5), 1091-1123.
- Halverson, R., Kelley, C., & Kimball, S. (2004). Implementing teacher evaluation systems: How principals make sense of complex artifacts to shape local instructional practice. In W. Hoy & C. Miskel (Eds.), *Educational administration, policy, and reform: Research and measurement* (pp. 153-188). Greenwich, CT: Information Age Publishing.
- Hamilton, L. S., Steiner, E. D., Holtzman, D., Fulbeck, E. S., Robyn, A., Poirier, J., & O'Neil, C. (2014). *Using teacher evaluation data to inform professional development in the intensive partnership sites*. Santa Monica, CA: RAND Corporation.
- Hanushek, E. A., & Rivkin, S. G. (2006). Teacher quality. In E. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (pp.1051-1078). Oxford, UK: Elsevier

- Hattie, J., & Gan, M. (2011). Instruction based on feedback. In R. Mayer & P. Alexander (Eds.), *Handbook of research on learning and instruction*, (pp. 249-271). New York, NY: Routledge.
- Herlihy, C., Karger, E., Pollard, C., Hill, H. C., Kraft, M. A., Williams, M., & Howard, S. (2014). State and local efforts to investigate the validity and reliability of scores from teacher evaluation systems. *Teachers College Record*, *116*(1), 1-28.
- Heyde, C. R. (2013). *Teacher and administrator perceptions of the effectiveness of current teacher evaluation practices and the impact of the new Illinois performance evaluation reform act of 2010*. (Doctoral dissertation, Northeastern University, Illinois).
- Hill, H., & Grossman, P. (2013). Learning from teacher observations: Challenges and opportunities posed by new teacher evaluation systems. *Harvard Educational Review*, *83*(2), 371-384.
- Holland, P. (2005). The case for expanding standards for teacher evaluation to include an instructional supervision perspective. *Journal of Personnel Evaluation in Education*, *18*(1), 67-77.
- Howard, B. B., & Gullickson, A. R. (2010). Setting standards for teacher evaluation. In M. Kennedy (Ed.), *Teacher assessment and the quest for teacher quality: A handbook* (pp. 337-354). San Francisco, CA: Josey-Bass.
- Hunter, M. (1988). Effecting a reconciliation between supervision and evaluation: A reply to Popham. *Journal of Personnel Evaluation in Education*, *1*(3), 275-279.
- Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology*, *64*(4), 349.

- Ingersoll, R., Merrill, L., & Stuckey, D. (2014). Seven trends: the transformation of the teaching force, updated April 2014. *CPRE Report (#RR-80)*. Philadelphia, PA: Consortium for Policy Research in Education.
- Jacob, B. A. (2011). Do principals fire the worst teachers? *Educational Evaluation and Policy Analysis*, 33(4), 403-434.
- Jacob, B., & Lefgren, L. (2006). When principals rate teachers. *Education Next*, 6(2), 59-69.
- Jerald, C. (2012). *Ensuring accurate feedback from observations: Perspectives on practice*. , Seattle, WA: Bill and Melinda Gates Foundation.
- K12 Insight (2016). *K12 Insight: About Us*. Retrieved from <http://www.k12insight.com/>.
- Kane, T. J., & Staiger, D. O. (2012). (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains. MET Project Policy and Practice Brief*. Seattle, WA: Bill and Melinda Gates Foundation.
- Kazemi, E., Lampert, M., & Ghouseini, H. (2007). *Conceptualizing and using routines of practice in mathematics teaching to advance professional education*. Chicago, IL: Spencer Foundation.
- Kelley, C. J. & Maslow, V. (2012). Does evaluation advance teaching practice? The effects of performance evaluation on teaching quality and system change in large diverse high schools. *Journal of School Leadership*, 22 (22), 600.
- Kimball, S. M., & Milanowski, A. (2009). Examining teacher evaluation validity and leadership decision making within a standards-based evaluation system. *Educational Administration Quarterly*, 45(1), 34-70.

- Kraft, M. A., & Gilmour, A. (2015). Can evaluation promote teacher development? Principals' views and experiences implementing observation and feedback cycles. *Harvard Education Review*.
- Lewis, T., Rice, M., & Rice, R. (2011). Superintendents' beliefs and behaviors regarding instructional leadership standards reform. *The International Journal of Educational Leadership Preparation*, 6(1), 1-13
- Looney, J. (2011). Developing high-quality teachers: Teacher evaluation for improvement. *European Journal of Education*, 46(4), 440-455.
- Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science*, 5(4), 333-369.
- Marshall, K. (2013). *Rethinking teacher supervision and evaluation: How to work smart, build collaboration, and close the achievement gap*. New York, NY: John Wiley & Sons.
- McCreight, C. (2000). *Teacher attrition, shortage, and strategies for teacher retention*. College Station, TX: Texas A & M University
- Merriam, S. (2009). *Qualitative research: A guide to design & implementation*. San Francisco, CA: Jossey-Bass.
- Mette, I. M., Range, B. G., Anderson, J., Hvidston, D. J., & Nieuwenhuizen, L. (2015). Teachers' perceptions of teacher supervision and evaluation: A reflection of school improvement practices in the age of reform. *Education Leadership Review*, 16(1), 16-30.
- Milanowski, A. T. & Heneman III, H. G. (2001). Assessment of teacher reactions to a standards-based teacher evaluation system: A pilot study. *Journal of Personnel Evaluation in Education*, 15(3), 193-212.

- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.
- Minnici, A. (2014). The mind shift in teacher evaluation: Where we stand--and where we need to go. *American Educator*, 38(1), 22-26.
- Mitchell, K., & Purchase, N. Y. (2015). An empirical analysis of New York's teacher evaluation system: Flaws in the design. *AASA Journal of Scholarship and Practice*, 12(1), 4-51.
- Mooney, J. (2013). Majority of NJ schools opt for widely used teacher-evaluation method. *NJ Spotlight*, pp. B7.
- Murphy, J., Hallinger, P., & Heck, R. H. (2013). Leading via teacher evaluation: The case of the missing clothes? *Educational Researcher*, 42(6), 349-354.
- Murray, P. K. (2014). *An investigation of teacher and administrator perceptions of Pennsylvania's new teacher evaluation system, based upon the Danielson framework for teaching, and its impact on teachers' instructional strategies in an urban school district*. (Doctoral dissertation, Indiana University of Pennsylvania).
- National Council on Teacher Quality. (2012). *State of the states 2012: Teacher effectiveness policies*. Washington, DC.
- O'Pry, S. C. & Schumacher, G. (2012). New teachers' perceptions of a standards-based performance appraisal system. *Educational Assessment, Evaluation and Accountability*, 24(4), 325-350.
- Ovando, M. N., & Harris, B. M. (1993). Teachers' perceptions of the post-observation conference: Implications for formative evaluation. *Journal of Personnel Evaluation in Education*, 7(4), 301-310.

- Papay, J. P. (2012). Refocusing the debate: Assessing the purposes and tools of teacher evaluation. *Harvard Educational Review*, 82(1), 123-141.
- Peña-López, I. (2009). *Creating effective teaching and learning environments: First results from Teaching and Learning International Survey (TALIS)*. Paris, France: Organisation for Economic Co-Operation and Development.
- Popham, W. J. (2013). On serving two masters: Formative and summative teacher evaluation. *Principal Leadership*, 13(7), 18-22.
- Pritchett, J., Sparks, T. J., & Taylor-Johnson, T. (2010). *An analysis of teacher quality, evaluation, professional development and tenure as it relates to student achievement*. (Dissertation, Saint Louis University).
- Range, B. G., Young, S., & Hvidston, D. (2013). Teacher perceptions about observation conferences: What do teachers think about their formative supervision in one US school district? *School Leadership & Management*, 33(1), 61-77.
- Reeves, D. B. (2009) *Leading change in your school: How to conquer myths, build commitment, and get results*. Alexandria, VA: Association of Supervision and Curriculum Development
- Reinhorn, S. K. (2013). *Seeking balance between assessment and support: teachers' experiences of teacher evaluation in six high-poverty urban schools*. Boston, MA: The Project on the Next Generation of Teachers.
- Reinhorn, S. K. Johnson, S. M., & Simon, N. S. (2015). *Supporting continuous development for teachers: Teacher evaluation in six high-performing, high-poverty, urban schools*. Boston, MA: The Project on the Next Generation of Teachers.

- Riordan, J., Lacireno-Paquet, N., Shakman, K., Bocala, C., & Chang, Q. (2015). *Redesigning teacher evaluation: Lessons from a pilot implementation*. Washington, DC: Regional Educational Laboratory.
- Rotherham, A. J., & Mitchel, A. L. (2014). *Genuine progress, greater challenges: A decade of teacher effectiveness reforms*. Washington, DC: Bellwether Education Partners.
- Russell, T. (2002). Can self-study improve teacher education? In J. Loughran & T. Russell (Eds.), *Improving teacher education practices through self-study*, (pp. 3-9). New York, NY: RoutledgeFalmer.
- Sartain, L., Stoelinga, S. R., Brown, E. R., Luppescu, S., Matsko, K. K., Miller, F. K., Durwood, C. E., Jiang, J. & Glazer, D. (2011). *Rethinking teacher evaluation in Chicago*. Chicago, IL: Consortium on Chicago School Research.
- Sawchuk, S. (2016). ESSA loosens reins on teacher evaluations, qualifications. *Education Week*, 14-15.
- Scheeler, M. C., Ruhl, K. L., & McAfee, J. K. (2004). Providing performance feedback to teachers: A review. *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children*, 27(4), 396-407.
- Scriven, M. (1967). The methodology of evaluation. In R. Tyler, R. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 151-161). Chicago, IL: Rand McNally.
- Shaha, S. H., Glassett, K. F., & Copas, A. (2015). The impact of teacher observations with coordinated professional development on student performance: A 27-state program evaluation. *Journal of College Teaching & Learning*, 12(1), 55-64.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153-189.

- Sinnema, C. E., & Robinson, V. M. (2007). The leadership of teaching and learning: Implications for teacher evaluation. *Leadership and Policy in Schools*, 6(4), 319-343.
- Smylie, M. (2014). Teacher evaluation and the problem of professional development. *Mid-Western Educational Researcher*, 26(2), 97-111.
- Spillane, J. P., Healey, K., & Mesler Parise, L. (2009). School leaders' opportunities to learn: A descriptive analysis from a distributed perspective. *Educational Review*, 61(4), 407-432.
- Stake, R. E. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.
- Stronge, J. (2013). *Effective teachers= student achievement: What the research says*. New York, NY: Routledge.
- The Danielson Group. (2013). The Framework. Retrieved from <https://www.danielsongroup.org/framework>
- The New Teacher Project (2012). 'MET' made simple: Building research-based teacher evaluation. New York, NY.
- Trinter, C. P., & Carlson-Jaquez, H. (2014) *Teacher evaluation: The role of subject matter in supervisor feedback*. Richmond, VA: Metropolitan Education Research Consortium.
- Tripp, D. (2011). *Critical incidents in teaching: Developing professional judgement*. Abington, Oxon, UK: Routledge.
- Tucker, P. D., & Stronge, J. H. (2005). *Linking Teacher Evaluation and Student Learning*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Tuytens, M., & Devos, G. (2010). The influence of school leadership on teachers' perception of teacher evaluation policy. *Educational Studies*, 36(5), 521-536.
- U.S. Department of Education. (2009). Race to the Top program: Executive summary. Washington, DC.

- Westerberg, T. R. (2013). Feedback for Teachers. *Principal Leadership*, 13(7), 30-33.
- Wiggins, G. (2012). Seven keys to effective feedback. *Educational Leadership*, 70(1), 10-16.
- Williams, S.E. (1997, March). *Teachers' written comments and students' responses: A socially constructed interaction*. Paper presented at the annual meeting of the Conference on College Composition and Communication, Phoenix, AZ.
- Wood, C. J., & Pohland, P. A. (1979). Teacher evaluation: The myth and realities. In W. Duckett (Ed.), *Planning for the evaluation of teaching* (pp. 73-82). Bloomington, IN: Phi Delta Kappa.
- Yin, R. (2003).K. (2003). *Case study research: Design and methods*. Thousand Oaks, CA: Sage.
- Zepeda, S. J. & Kruskamp, B. (2012). *Teacher evaluation and teacher effectiveness*. Larchmont, NY: Eye on Education.
- Zepeda, S. J. (2007). *The principal as instructional leader: A handbook for supervisors*. New York, NY: Routledge.