

LESSONS LEARNED FROM THE IMPLEMENTATION OF A FORMATIVE
TEACHER EVALUATION SYSTEM: AN INSTRUMENTAL CASE STUDY

BY

HEBA A. ABDO

A dissertation submitted to

The Graduate School of Education

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Education

Graduate Program in Teacher Leadership

written under the direction of

Dr. Alisa Belzer

Approved by

Dr. Alisa Belzer, Chair

Dr. Drew Gitomer, Committee

Dr. Christopher Manno, Committee

New Brunswick, New Jersey

May 2017

IMPLEMENTATION OF FORMATIVE TEACHER EVALUATION

Abstract

Scholarship in the field of teacher evaluation has increased recently following seminal studies on its importance (Kane & Staiger, 2011; Weisberg et al., 2009) that led to federal policy changes that incentivize its use (U.S. Department of Education, 2009). Scholarship began with agreement across the field about the importance of valid tools and reliable methods, but gave way to widespread disagreement about the best uses of teacher evaluation. Many vocal researchers support summative uses (see for example: Gordon, Kane, & Staiger, 2006; Hanushek & Rivkin, 2006), while others call for formative approaches, linked to professional learning opportunities (see for example: Papay, 2012; Sinnema & Robinson, 2007). Engaging in formative evaluation may lead to teacher efficacy (Danielson, 2012), teacher self-reports of professional learning (Organization for Economic Cooperation and Development, 2013), positive teaching practices (Reinhorn, 2013; Sertain et al., 2011), and student achievement (Dee & Wyckoff, 2015).

Yet, the study of formative evaluation systems is an emergent field with many gaps. The literature has not adequately made recommendations about school-level programs that support formative evaluation or provided descriptive examples of systems that perform well. In addition, while many have proposed the importance of understanding how teachers actually use feedback, few detail this process.

This instrumental case study provides a description of the teacher evaluation system at one charter school in Washington D.C. It explores how the formative evaluation system was constructed and what the outcomes are for teacher learning and the school's climate. Findings reveal that the purposeful inclusion of formative processes has led to deep and sustained teacher

IMPLEMENTATION OF FORMATIVE TEACHER EVALUATION

learning in some cases, but has been less impactful in others. Further, formative evaluation practices support positive school climate, such that when the evaluation system was implemented well, teachers reported feeling supported by school climate, and when practices faltered, teachers described negative outcomes regarding the school's climate. Thus, sustaining formative teacher evaluation processes is an important task for school leadership. Yet, data indicate that sustaining these systems is a difficult process that requires multi-year training and follow-up. Findings also reveal that teachers are in need of evaluation that is not just formative, but is also linked to collaborative structures.

IMPLEMENTATION OF FORMATIVE TEACHER EVALUATION

Acknowledgments

All thanks and praise to God, first and always, for all the blessings and lessons in my life.

My choice to become an educator was never really a conscious one at all. Somehow, education showed its path to me and led me through a lushly-lined journey of self-discovery, challenge, and compassion. The Educational Doctorate program has delivered me to the end of that path, a strong and well-versed change-maker, ready to forge ahead on my own trail.

Along the way, I was honored to meet and be supported by the most caring and knowledgeable professors and classmates. To the members of the 2013 Ed.D. Cohort: congratulations and thank you all for the laughs and encouragement. I especially thank the members of my dissertation group for their many (many, many, many) ideas and edits. I am so glad we had each other to lean on, Erin Comollo and Jenny Rich. Thanks also to Juail Goode, Steve Heisler, and Christian Mathews for pushing me through and opening so many doors. And to Carolyn Ross, who has already made it so that I cannot possibly imagine life without her. Thank you for always challenging my thinking and enriching my work, but most importantly, thank you for being my friend.

I also thank the professors at the Graduate School of Education. I have grown so much over these years, and I can remember very specific ways that each of you encouraged that growth. Thank you to Dr. Melinda Mangin for showing me the meaning of leading with wisdom and compassion. Thank you to Dr. Thea Abu El-Haj, Dr. Tanja Sargent, and Dr. Amy Lewis for your guidance when taking on my first professorship. Thank you to my dissertation

IMPLEMENTATION OF FORMATIVE TEACHER EVALUATION

committee. Dr. Chris Manno, I will never forget that you took a lot of time out of your busy schedule to speak to me for over an hour about the importance of “leading up” and share your inspirational and personal account, just because you sensed that I might be feeling overwhelmed. Dr. Drew Gitomer, to say that I was intimidated by your inclusion on my committee would not even begin to describe my level of anxiety. I was blessed just to have a leading expert in my area of study ready to give guidance, but was delighted to find that it was always packaged with so much logic, wisdom, and compassion. Rather than feeling overwhelmed, I felt included and supported. Thank you for not only opening up your professional resources to me and giving of your time, but also opening your home and connecting with me. And, to Dr. Alisa Belzer, a “thank you” could never suffice. But, thank you anyway for your unending counsel. Even before I began the Ed.D. Program, you were advocating for me. In your subtle way, you have broadened my understanding of so many important areas of research and practice while transforming my world-view on so many topics. I could not have picked a better chair to lead me through this process.

Thank you to all the members of “Elmwood School,” and especially to the principal, “Andre,” for opening your school to my research and sharing of your time and experiences.

To my mentors, Dr. Mitalene Fletcher, Dr. Michelle Macchia, Dr. Nora Salem, and Dr. Erol Bailey: your advice and passion has helped me to forge ahead through every challenge of this process. And to Dr. Mecide Gharibo and Eman El-Bedawi for providing a fresh start.

Thank you to my great friends. I am so blessed to have so many wonderful, powerful, compassionate women in my life. To have one great friend is more valuable than gold, and how rich I am to have so many! You have delivered me through some of my darkest days, and for that, I am truly grateful. I love you Amber, Amina, Anna, Beth, Efranz, Eman Y., Gigi, Mona

IMPLEMENTATION OF FORMATIVE TEACHER EVALUATION

A., Mona M., Rania, Walaa, Yasmin, and Youssra. To my Sherine Abolezz, no words are ever needed or worthy.

Most importantly, to momma and baba: without you nothing would be possible.

رَبِّ ارْحَمْهُمَا كَمَا رَبَّيَانِي صَغِيرًا And, unfortunately for you, in my old age as well! I love you both. I am certain that nothing I could say or do could ever come close to showing enough gratitude for all that you have always done for me, especially in the past 5 years, but thank you anyways and always. To my family, near and far, thank you for all of your support throughout the years.

Ronna and Kareem, I love you both so much and hope we can celebrate a lot more successes in the coming years!

IMPLEMENTATION OF FORMATIVE TEACHER EVALUATION

Dedication

I dedicate this dissertation to every woman who endured and persisted so that one day I could succeed.

And to my two daughters: my ambitious and adventurous Ayah and my smart and sweet Hana, who give my life purpose and joy. I hope one day you understand that all of my hard work was for you. More importantly, I hope one day you are *certain* that, as women of color, *everything* is possible.

IMPLEMENTATION OF FORMATIVE TEACHER EVALUATION

Table of Contents

Abstract	ii
Acknowledgments	iv
Dedication	vi
List of Tables	xii
List of Figures	xii
CHAPTER ONE: INTRODUCTION	1
Problem Statement	6
Research Questions	7
Theoretical Framework: Formative Evaluation	7
Chapter Review and Overview of Chapters	9
CHAPTER TWO: REVIEW OF THE LITERATURE	12
The Challenge of Defining and Identifying Teacher Quality	14
Improving Tools	16
Improving Evaluator Practices	17
Summative and Formative Uses of Teacher Evaluation Data	19
Arguments That Support Summative and Formative Views	21
Summative Approaches	22
Formative Approaches	23
Arguing against policy: Teacher evaluation as a limited source of data.	24
Arguing against policy: Unintended consequences of high-stakes uses of teacher evaluation.	31

IMPLEMENTATION OF FORMATIVE TEACHER EVALUATION

Arguing for the benefits of formative approaches for teacher learning and improvement.	34
School-Level Practices that Support Formative Evaluation	37
Evaluator Training	38
Manageable Evaluator Workloads	40
Enacting Formative Feedback	42
Linking Professional Learning	44
Outcomes of Teacher Evaluation Policy: Limited School-Level Changes and Exemplars	46
Conclusion and Chapter Review	49
CHAPTER THREE: RESEARCH DESIGN	51
Site Selection and Research Context	52
Participants and Sampling Strategy	55
Data Collection Procedures	57
Administrator Interviews	58
Teacher Focus Groups	59
Virtual Document and Video Analysis	60
Teacher Interviews	61
Human Subjects Protection	63
Data Analysis Plan	63
Researcher Role.....	66
Trustworthiness	68
Limitations	70
Chapter Review and Conclusion	71

IMPLEMENTATION OF FORMATIVE TEACHER EVALUATION

CHAPTER FOUR: FINDINGS	72
Background: Elmwood School's Teacher Evaluation System	73
Year 1: Initiating the Formative Teacher Evaluation System	74
Year 2: Investing Teachers in the Process	81
Year 3: Sustaining through Turbulence	83
Year 4: Realigning the Teacher Evaluation System	86
Teacher Evaluation Support Structures	87
Distributed Evaluator Caseloads	88
Evaluator Training	89
Building a Growth-Mindset	93
Linked Professional Learning	96
Inductive Findings: Sustaining the Teacher Evaluation System through Programming ..	100
Lessons Learned from Implementation of the Elmwood School Teacher Evaluation System	
.....	109
Teacher Experiences of Formative Evaluation	110
"I want to feel supported, not just evaluated": School Climate and Teacher Evaluation ...	118
Professional Safety.	118
Urgency.	119
Buy-in.	123
Leadership Lessons Learned	128
Chapter Review and Conclusion	136
CHAPTER FIVE: DISCUSSION AND IMPLICATIONS	138
Discussion of Findings	138

IMPLEMENTATION OF FORMATIVE TEACHER EVALUATION

Research Question 1	139
Utilization-focused evaluation.:.....	139
Sustained leadership.:.....	143
Research Question 2	148
Revision to the theoretical framework: Reevaluating the nature of teacher evaluation feedback.:.....	148
Conclusions	156
Missed Opportunities	156
Outcomes of Research at Elmwood School	158
Implications for Practice	159
Implications for Policy	161
Implications for Future Research	162
Personal Reflection	163
References	166
APPENDICES	202
Appendix A: Consent Forms	202
Appendix B- Data Collection Protocols	205

IMPLEMENTATION OF FORMATIVE TEACHER EVALUATION

List of Tables

Table 1. Summative and Formative Teacher Evaluation Methods.....	21
Table 2. Participant Information.....	57
Table 3. Alignment of Protocols with Research Questions.....	63

IMPLEMENTATION OF FORMATIVE TEACHER EVALUATION

List of Figures

Figure 1. The Six Steps of Feedback (adopted from Bambrick-Santoyo, 2012b).....	79
Figure 2. Timeline of Teacher Evaluation System Implementation.....	87
Figure 3. Professional Learning Program Alignment with Teacher Evaluation System.....	99
Figure 4. Google Tracking Spreadsheet.....	105
Figure 5. Teacher Binder.....	106

CHAPTER ONE: INTRODUCTION

Most agree that variances in teacher quality contribute to substantial and lasting differences in students' learning (Bryk et al., 2010; Danielson, 2011; Darling-Hammond, 2000; Goldhaber, 2002; Hanna, 2013; Hanushek, 2010; Marzano, 2003; Rockoff, 2004; Rivkin, Hanushek, and Kain, 2005; Tucker & Stronge, 2005). Yet, researchers do not agree about the specific qualities that effective teachers should possess in order to maximize student outcomes or on how to assess the extent to which these qualities are present in individual teachers (Jacob & Lefgren, 2006; Kennedy, 2008; Morganstein & Wasserstein, 2014). This ambiguity has contributed to differences of opinions about what makes teacher evaluation valid and effective. Additionally, researchers are deeply divided about the uses and outcomes of this process. Many look to the human capital benefits of sound evaluation tools, favoring their use for documenting classroom practices which can support decisions regarding tenure, teacher terminations, sanctions, and rewards (see for example: Gordon, Kane, & Staiger, 2006; Hanushek & Rivkin, 2006; Jacob, 2011; Tucker & Stronge, 2005). This body of work is often accompanied by a view of research and policy as central to the growth of the teaching profession. A separate research body places teacher evaluation within the larger context of positive and sustained professional learning in schools (see for example: Darling-Hammond et al., 2012a; Donaldson, 2009; Papay, 2012; Sinnema & Robinson, 2007; Steinberg & Sartain, 2015). These studies favor valid teacher evaluation tools as a catalyst for professional growth, such that they provide teachers with meaningful feedback that can positively influence practice. Researchers with this orientation towards teacher evaluation often establish teacher growth as central to school-wide reform.

Teacher evaluation has been a developing human resources process in schools since the beginning of the industrial revolution (DuFour & Marzano, 2009; Johnson, 1990). It originated from a perceived need to limit teacher behaviors to a specific list of measurable attributes as a way to garner effective practices (DuFour & Marzano, 2009). However, some have critiqued the validity of this approach because behavior checklists are limited to easily observed teacher actions that leave out the more tacit processes associated with effective teaching. In response, the current teacher evaluation reform movement has incorporated holistic rubric-based tools that researchers believe more clearly depict effective teacher practice than behavioral checklists (DuFour & Marzano, 2009). Additionally, recent teacher evaluation reform efforts have recognized that school factors inhibit administrators' ability to reliably evaluate teacher practice. These factors include teacher beliefs that their practice should not be open to scrutiny and administrator inability 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 teachers had been rated as "satisfactory" and that teachers receive very little usable feedback about their practice. At the same time, many began to question traditional measures of teacher quality including years of experience, degrees, etc. Because these did not correlate to increased student outcomes on standardized tests (Bryk et al., 2010). Together, these studies 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 school leaders on how they might use the data that the tools generate to make reliable

decisions about judging teacher practice (Futernick, 2010; Kane & Staiger, 2012; Kimball & Milanowski, 2009).

The reforms emerged more quickly than had been 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, funded the Measures of Effective Teaching (MET) Project in 2009. The MET Project implemented a comprehensive experimental design study of six districts with about 3,000 teachers in which they collected many sources of data on teacher effectiveness, including classroom observation, achievement scores, and student surveys (Kane, Kerr, & Pianta, 2014b). In addition, during the second year of the study, they randomly assigned student rosters to teachers in order to remove possible population bias. Through the course of the study, MET intended to identify the most valid evaluation measures, or combination of measures and the most reliable processes with which to implement them in order to increase the effectiveness of the teaching force. During the same year, the Obama administration introduced “Race to the Top,” (RTT) a federal program that aimed to eradicate the achievement gap by investing in multiple state-wide educational programs including teacher evaluation (U.S. Department of Education [USDOE], 2009). RTT provides many incentives for states to implement teacher evaluation and other programs including \$46 billion in monetary grants (USDOE, 2009) and waivers related to a previous, highly regulatory, federal program, No Child Left Behind (Popham, 2013).

Not surprisingly, the introduction of RTT led to many reforms across the United States including increased spending on teacher evaluation tools (Chambers, et al., 2013; Darling-Hammond, 2013; Garrison-Mogren & Gutmann, 2012) and the adoption of new or improved

teacher evaluation systems by forty-eight states (Doherty & Jacobs, 2015; National Council on Teacher Quality, 2012; Rotherham & Mitchel, 2014). McGuinn (2012) classifies these reforms as the most substantial and widely adopted in the U.S. In decades. 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 (Goldhaber & Brown, in press; Sawchuk, 2016).

The thrust of MET and RTT has also created a large research base on the use of teacher evaluation 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). Researchers doing this work tend to envision teacher evaluation as a strategy to increase the effectiveness of the teaching field as a whole by utilizing sanctions and rewards to ensure that each classroom is staffed with an effective teacher. Proponents of this view of teacher evaluation use it in *summative* ways for making personnel decisions. In teacher evaluation, summative uses refer to using data to make final decisions about teachers' salaries or positions.

At the same time, another branch of the teacher evaluation debate that favors *formative* uses of teacher evaluation has been growing. Formative uses refer to using data to make to help teachers to learn about their instruction during the course of the school year. This strand also envisions valid and reliable teacher evaluation as central to increasing teacher effectiveness, but looks to ways it can do so through meaningful feedback and data-driven professional development targeted specifically to what teachers need in order to be effective. These

researchers often centralize teacher learning as a means to successful school reform. They envision teacher evaluation as a way for teachers to receive timely and useful feedback about their practice which they can use to help them make appropriate improvements to instruction.

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 learning. Yet, in practice, research shows that many schools do not enact either well (Kraft & Gilmour, 2015), failing to identify ineffective teachers or to support targeted teacher growth. This is important because how schools implement evaluation systems changes the success of systems' ability to promote teacher development (Kraft & Gilmour, 2016a). At the same time, schools may struggle to implement formative systems due to a lack of guidance on the implementation of these more positive practices (Riordan et al., 2015). School leaders investigating formative systems find a loosely-connected body of literature. Of the studies that are available on school implementation of formative teacher evaluation, a majority only focus on stakeholder perceptions, including fifteen from the United States, (Anast-May, et al., 2011; Bambrick-Santoyo, 2012; Callahan & Sadeghi, 2014; Dretzke et al., 2015; Donaldson 2012; Ford et al., 2015; LaRocco, Wheeler, & Sennett, 2015; Mette et al., 2015; Mitchell & Purchase, 2015; O'Pry & Schumacher, 2012; Range, Young, & Hvidston, 2013; Reinhorn, Johnson, & Simon, 2015; Sartain et al., 2011; Thomson, 2014), eleven international works (Charalambous et al., 2014; Clipa, 2015; Coimbra, 2013; Collins, 2004; Delvaux et al., 2013; Marcos, Machado, & Abelha, 2015; Milanowski & Heneman, 2001; Peña-López, 2009; Tuytens & Devos, 2010; Vekeman, Devos, & Tuytens, 2014), and nine dissertations (Bonavitacola, 2014; Heyde, 2013; Legg, 2015; Murray, 2014;

Shough, 2010; Smith, 2015; Spina, Buckley, & Puchner, 2014; Thomson, 2014). These studies provide a more detailed view of school-level responses to new policy in this realm and help researchers begin to understand teacher feelings and needs regarding evaluation. Their inclusion in the literature signals an important shift towards an appreciation for teacher voices in reform movements related to teacher evaluation. Further, these studies help the field to more clearly understand what aspects of new policy are well-received in practice. However, these studies do not provide detailed guidance about how schools can implement well-constructed formative systems in order to increase both teacher buy-in and teacher learning. Further, as school implementation varies widely in studied sites, many works can only describe findings from schools that are not implementing formative systems well. This study adds to the literature by examining both perceptions and uses of teacher evaluation at one school that self-reports having a well-constructed evaluation system that is designed to play a formative role in teacher growth and development. Thus, it contributes to a better understanding of the process, products, and lessons learned related to creating a formative evaluation system.

Problem Statement

The study of school implementation of teacher evaluation, and especially formative teacher evaluation systems, is an emergent and quickly growing field. Currently, many gaps exist in this knowledge base. This study seeks to explore two of them more deeply. More specifically, the literature has not adequately made recommendations about school-level programs and policies that support formative teacher evaluation, especially with regard to descriptive examples of systems that are performing effectively. Second, while many researchers propose the importance of understanding how teachers actually use formative feedback, few studies detail what this process looks like. This is an important area in research

that others have highlighted as “pressing” for the field (see for example: Kraft & Gilmour, 2016a). For example, Cohen & Goldhaber (2016) suggest “Little is known about the degree to which, and ways in which, teaching practices evolve in response to observations systems.

Understanding the mechanisms through which teachers learn to improve instruction is a crucial step for research on observational measures as evaluation tools” (p 384). This study will also address this important area of research by seeking to understand how teachers use teacher evaluation feedback to think about instruction in new ways and to implement different classroom practices.

Research Questions:

1. By studying a school with a strong commitment to formative evaluation, what can we learn about school-level features that support teacher use of evaluation feedback?
2. In what ways do teachers use formative teacher evaluation feedback to inform their practice?

Theoretical Framework: Formative Evaluation

This study explores the uses and outcomes of a formatively constructed 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 leads to more effective practices in the teaching workforce, overall, and enhances its impact on student learning. Therefore, it adopts “formative evaluation” as the theoretical framework. In this theoretical framework, I will describe the distinction between the two terms, “summative” and “formative,” in regard 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. 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. In teacher evaluation, summative decisions are most often made at the end of a school year and to decide about sanctions, such as firing, and rewards, such as increased pay. Conversely, formative evaluation leads to feedback during the course of the event or program, while “it is fluid” (Scriven, 1967, p. 236) and is, therefore, a means of evaluating for the purpose of learning and improvement (Goe et al, 2008). 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.

Formative evaluation in teacher education, then, relates to using data generated from teacher evaluation protocols to help teachers with their classroom practice, on the assumption that teacher practice can always be improved. Data could be generated from many sources, including classroom observation, standardized tests, student surveys, professional portfolios, so long as it is used as a tool to inform a teacher’s learning needs rather than to make a

consequential determination about the teacher. For the purpose of this study, I use the term formative evaluation in reference to the most-widely used teacher evaluation practice, classroom observation. Schools that utilize formative teacher evaluation create systemic supports for teacher learning. These include structured feedback about instruction, both written and verbal (Bonavitacola, 2014; Greenberg, 2015; Hill & Grossman, 2013; Kelley & Maslow, 2012; Papay & Kraft, 2015; Sartain et al., 2011), professional dialogue about instruction, especially that which provokes self-reflection (Anast-May et al., 2011; Chappuis & Chappuis, 2007; Costa & Garmston, 2002; Danielson, 2001; Marzano et al., 2012; Ovando, 2005; Range et al., 2013), and formal professional development that is linked to teacher evaluation data (Archibald et al., 2011; Coggsall 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). The literature review will present the work of researchers who support formative teacher evaluation, their arguments in support of this utilization of data, and associated stances. It will also discuss the ways in which schools implement formative teacher evaluation and our limited understandings of the outcomes of these systems.

Chapter Review and Overview of Chapters

This introductory chapter provides an overview of current policies that incentivize the inclusion of teacher evaluation systems in schools and the ways in which researchers and practitioners have responded to these policies. It further describes two possible uses of teacher evaluation data. Summative uses differentiate between teachers for the purpose of making decisions about job status and salary. Formative uses support teacher growth through feedback

and linked professional learning opportunities. The chapter also situates the current study within the latter, formative, approach.

As Chapter Three will describe in more detail, data for this study was collected in a school that the principal described as purposefully implementing a formative teacher evaluation system by providing ongoing, frequent feedback related to observations of teacher practice and supported through other teacher learning opportunities. As such, it provides a rich case site to explore the theoretical framework of formative evaluation.

This dissertation is made up of five chapters: (1) introduction, (2) literature review, (3) research design, (4) findings, and (5) discussion and implications. Chapter One introduces the problem of practice, the purpose of my study, and my theoretical framework. In Chapter Two, I provide a review of the literature on teacher evaluation, formative feedback from both the teacher and student assessment literature, formative supervision, a new area of study that explains the cross-section between instructional and educational leadership, and findings from recent studies that explore the impact of Race to the Top on teacher evaluation practices at schools. Each of these areas helps to more clearly situate my study. Chapter Three describes the steps that were taken in my research to collect and analyze the data. Data for this qualitative, instrumental case study were collected through one focus group, multiple interviews, observations, and a review of documents and videos. In addition, through the course of data collection, I was given access to the results of a commercially-created survey of teacher attitudes that included questions about the teacher evaluation system. The school administered the survey twice a year, beginning the year before I began my study. I was able to access results from the winter administration of the year I collected data. Survey data was collected in January, and I began my data collection in March. Chapter Three also outlines my data analysis procedure of

open coding (Seidel, 1998), through which I continuously combed and recombined my data to find recurring themes. Chapter Four presents these themes, including support structures for teacher evaluation, teacher learning outcomes, and the ways in which the school's climate supported and was supported by the teacher evaluation system. Finally, in Chapter Five, I discuss the findings of my case study in light of my research questions, theoretical framework, and literature review. Specifically, I will describe that few findings related to my second research question and I will expand upon the dichotomy of formative and summative evaluation by looking at a third important strategy in using teacher assessment data. I also provide outcomes of the study that have already impacted the school site, as well as implications for research and practice.

CHAPTER TWO: REVIEW OF THE LITERATURE

The literature on teacher evaluation begins by following a singular line of inquiry which focuses on the need to create valid tools and reliable processes that fairly assess teacher practice. However, “two separate camps” (Clayton, 2013, p. 23) soon emerged; one focuses on dismissing ineffective teachers and the other on improving teacher learning. I add to this analysis by further describing the second “camp,” which I believe to be broader than Clayton describes, encompassing several rationales and approaches for the formative use of evaluation data. This literature review will describe a timeline of research in the field of teacher evaluation that begins with that singular focus on identifying the most valid tools and reliable approaches to teacher evaluation. Then, I will discuss the two main “camps,” both of which draw from the earlier literature on tools and processes. One calls for summative uses of teacher evaluation data, urging schools to use results to make decisions about the teaching workforce including sanctions and rewards. This “camp” is described briefly in my literature review.

I will then summarize the literature which emerges from the second “camp” on formative evaluation. This review will be more detailed as this study was designed to articulate the outcomes of a formative evaluation system. The formative literature emerges from three very nuanced, yet distinct, arguments and this literature review will examine each in depth. All three call for formative uses of teacher evaluation, or more specifically the use of teacher evaluation feedback to help teachers improve their practice. This often accompanies a belief that teachers should be supported in ways which address their individual learning. In the first, researchers assert that teacher evaluation tools cannot fully assess every aspect of a “teacher’s quality,” even

when they are extensively studied. Further, teacher quality itself is revealed as an unclear construct, thus necessitating a more holistic and descriptive view of teaching that recognizes the complexity and context-bound nature of teaching acts. Therefore, this argument proposes that teacher evaluation be used as just one of several approaches that can inform teacher improvement conversations. While the first argument questions the ways in which researchers have constructed teacher evaluation and a theory of action that accepts teacher evaluation as a valid and reliable way to increase professionalism in teaching, the second argument accepts teacher evaluation as a construct, but argues that its use as a summative tool has led to many unanticipated consequences. This argument raises concerns about the fast-paced implementation of teacher evaluation reform and denounces high-stakes uses of teacher evaluation data. In the final argument, teacher evaluation is also accepted as a valid construct. However, rather than condemn high-stakes uses, this argument discusses ways to use teacher evaluation data to heighten and sustain teacher learning. My literature review will then describe “formative supervision,” a new area of practice that describes the intersection between formative approaches and school leadership practices including practices associated with positively implemented formative evaluation. Finally, I will describe the ways that schools still struggle to implement teacher evaluation practices even now, eight years after strong consensus among researchers paved the way for its importance in the federal policy, “Race to the Top” (RTT).

In order to compile this review, I began by searching the Education Resources Information Center (ERIC) database for studies about effective teacher evaluation practices. This search led me to some quantitative studies that linked teacher evaluation practices to student gains. I then used Google Scholar’s search tool to locate additional studies by using the “cited by” function for those works. I widened my search by identifying recurring

articles in the reference sections of the resulting set of papers. Over time, I identified other areas of research that were related to teacher evaluation practices including the role of administrators, teacher perspectives on evaluation practices, and debates about types of teacher evaluation including process-product and standards-based approaches. Recent works also describe new findings on teacher evaluation that resulted from the federal program, Race to the Top, and the ancillary foundation-funded program, the Measures of Effective Practices study (MET). It is important to note that “teacher evaluation” may refer to any policy or program related to the assessment of teachers and subsequent human resources decisions. This includes student surveys of teachers, standardized test scores, larger policy issues related to evaluation, etc. For the purpose of this study, I will concentrate on one aspect of teacher evaluation, namely observations of teacher practice.

The Challenge of Defining and Identifying Teacher Quality

Most agree that teacher quality is the most influential school factor contributing to students’ short and long-term learning (Bryk et al., 2010; Danielson, 2011; Darling-Hammond, 2000; Goldhaber 2002; Hanna, 2013; Hanushek, 2010; Marzano, 2003; Rockoff 2004; Rivkin et al., 2005; Tucker & Stronge, 2005). However, school administrators struggle to identify which teachers are employing effective practices that support student growth (Weisberg et al., 2009). Weisberg et al.’s (2009) seminal description of the current state of teacher evaluation across twelve large districts in four states uses surveys of 15,000 teachers and 1,300 administrators to gauge the types of teacher evaluation tools schools use, the ways that feedback is communicated to teachers, and the ways that teacher evaluation data is used, both summatively and formatively, in the districts. They find that some of their original research questions were unanswerable because teacher evaluation was not being appropriately implemented in most

districts. One indication of this is that, across all of the districts, 98.7-100% of teachers were assigned the same evaluation: “satisfactory.” This was the case in “high” and “low”-performing schools, as measured by high stakes standardized tests. They term this phenomenon “The Widget Effect” in reference to the interchangeability of teachers who are all viewed as similarly effective (2009). Other researchers also observe this phenomenon (Bill & Melinda Gates Foundation, 2013; Danielson & McGreal, 2000; Donaldson, 2009; Glazerman et al., 2011; Marshall, 2005; Tucker & Stronge, 2005) using the term “Lake Wobegon Effect” which refers to a fictional town created by radio personality Garrison Keillor where “all the children are above average”(Danielson & McGreal, 2000, p. 4). The term was first applied to teacher evaluation by Tucker (1997). Weisberg et al. (2009) describe how school culture hinders administrators’ ability to rate teachers accurately, inhibiting them from dismissing ineffective teachers, recognizing teachers that perform well, and differentiating professional learning needs for the “broad plurality” of teachers with specific growth needs (p. 2). They offer two key recommendations for improving the quality of teacher evaluation, both formative and summative, that set the stage for the next phase of teacher evaluation research. First, they propose revising evaluation tools to create fair, credible, and accurate representations of the work that teachers do. This would include clearly defining the core responsibilities of teaching and creating protocols that could accurately assess these core responsibilities. Other researchers recognize that tools should also establish a common professional language because this makes the key elements of effective practice explicit and enables teachers to more purposefully include these elements in their repertoire (Coggshall et al., 2012; Danielson, 2011b; Darling-Hammond, 2012; Goe, 2013; Goe, Biggers, & Croft, 2012; Joe et al., 2013; McGreal, 1983). Second, Weisberg et al. (2009) recommend designing effective training programs that educate

administrators to develop the necessary skills and processes for rating teacher practice and giving appropriate feedback. Progress made in improving tools and administrator practices are discussed in more detail below.

Improving Tools

Throughout American educational history, many have tried to define appropriate teacher practice (DuFour & Marzano, 2009) in an effort to create valid tools that identify effective teaching (Goe et al, 2008). Early attempts to define effective teaching focus on trivial, yet easily measured, classroom attributes such as neatness of the room, colorfulness of displays, and cleanliness of the teacher and/or students (DuFour & Marzano, 2009). But, early in the 20th century, perhaps spurred by nation-wide investment in IQ and other standardized tests, researchers began to recognize a need for more scientific, fact-based evaluations, focusing, instead, on teacher behaviors effect student learning (Barr & Burton, 1926; Charters & Waples, 1929). Process-product orientations to teacher evaluation derive from these views of teacher effectiveness. In this conception, teacher evaluation tools listed specific teaching behaviors, the “processes,” that researchers link to an increase in student achievement, the “products,” most often using standardized test measures (Brophy, 1986; Gage & Needels, 1989; Hunter, 1982; Rosenshine, 1983; Rosenshine & Furst, 1973). These efforts were further refined after *A Nation at Risk* (Gardner, 1983) which encouraged standardization of personnel evaluation (Darling-Hammond et al., 1983; Hazi & Rucinski, 2009; Rothman & Toch, 2008).

Although process-product tools provided a unified way for evaluators to view instruction, researchers began to look for new solutions because results from process-product tools were not always usable, perhaps because they required the evaluator to make a lot of inferences (Kimball & Milanowski, 2009; Tucker & Stronge, 2005) suggesting that they were not very

objective. This often resulted in many varied and, often, opposing interpretations of evaluation tools, which, in turn, led teachers to view evaluation as an unfair practice (Kane, Kerr, & Pianta, 2014a). Further, researchers began to recognize that teaching requires activation of a complex skill-set that cannot be evaluated based on assessments of singular actions (Danielson, 2001; Darling-Hammond, 2013). Currently, many propose rubric-based teacher evaluation tools that attempt to more holistically describe the act of teaching while continuing to link specific teaching behaviors to student learning outcomes (Danielson, 2015; Marshall, 2013; Marzano, 2012; Stronge, 2013). Efforts to capture richer descriptions of teacher practice reduce evaluator inference, which may lead teachers to view assessments of their practice as more fair. In addition to lessening bias in the evaluation process, rubrics concretize a common and clear conception of effective teaching. For administrators, this is beneficial because they can provide more detailed feedback than is possible from a checklist since a clear direction for improvement is indicated within rubrics (Coggshall et al., 2012). For teachers, having clear and common language that spells out descriptions of effective practice provides a helpful point of comparison to an ideal of teaching (Danielson, 2001; Futernick, 2010). This can help them more fully understand their strengths and challenges (Coggshall et al., 2012; Danielson & McGreal, 2000; Joe et al., 2013). Researchers, in support of both summative and formative evaluation, are hopeful that this newest generation of teacher evaluation tools represents valid conceptions of teacher practice.

Improving Evaluator Practices

Despite growing convergence around conceptions of effective teaching reflected in evaluation tools, the reliability of these tools is still contingent upon the skills of the user. The ability to differentiate among teachers with regard to effectiveness is elusive. Observers notice

different things about teachers and teaching based upon their own experiences and motivations (Cook & Richards, 1972; Hamilton, et al., 2014; Kimball & Milanowski, 2009; Sabers, et al., 1991) and are often more able to notice and comment upon classroom management behaviors than instructional behaviors (Bell et al., 2014; Gitomer et al., 2014; Polikoff, 2015; Tucker and Stronge, 2005). In addition, research on observer ratings indicates that feedback data often relates more to the observer's preferences than to the teacher's skill set (see for example: Danielson & McGreal, 2000; Park, Chen, & Holtzman, 2014). When teachers come to recognize this bias, it may contribute to their viewing teacher evaluation as invalid (Johnson, 1990; Maslow & Kelley, 2012; Peterson, 2000; Sinnema & Robinson, 2007; Wise, et al., 1985; Wolf, 1972). Because of this tendency toward observer bias, research on the validity of evaluation tools is not enough and more research on observer skills and processes is warranted.

Thus, more recent research also focuses on increasing the reliability of observer use of tools. This began as researchers questioned the best ways to train administrators as a next step in creating more valid and reliable evaluation practices. Funded by the Bill & Melinda Gates Foundation, the Measures of Effective Teaching (MET) study plays an influential role in this area (Kane & Staiger, 2011). The MET study follows a large sample of 3,000 teachers in six large districts across the US over seven years. To explore the best ways to identify effective teaching practice, MET tested different combinations of teacher evaluation tools (including several classroom observation tools, student surveys, and student standardized test data) and methods (such as different numbers of observations, and different weightings of tools) to identify which would correlate best with student achievement data. Based on their correlational studies, MET makes many recommendations about how to collect and use teacher evaluation data. In terms of increasing observer reliability, MET recommends that schools require observer training,

certification, and audits. The goal of these recommendations is that administrators to rate all aspects of teaching more effectively. This includes learning how to separate out the observers' bias during an observation, learning about the tools themselves (vocabulary about teaching and learning, parts of the tool), how collected data lines up with the tool's definitions of high quality teaching, and how to provide accurate and helpful feedback to teachers. Moreover, MET researchers recommend that multiple raters evaluate teachers in order to increase teacher evaluation reliability (Kane et al., 2014b; Kane & Staiger, 2012; Kimball & Milanowski, 2009; The New Teacher Project, 2012). Observations should also last for at least 30 minutes (Joe, McClellan, & Holtzman, 2014). Lastly, the MET report recommends combining observation scores with other measures of teacher effectiveness, including teacher value-added on standardized tests and student surveys. MET researchers find that when schools implement the recommendations about rater training, length and schedule of observations, and use of data, they can accurately rate teacher effectiveness.

Summative and Formative Approaches to Teacher Evaluation Data

The MET Study (Kane & Cantrell, 2011) seemed to create some short-lived agreement in the field about the most reliable ways to implement valid teacher evaluation tools; nevertheless, researchers soon split in two distinctive directions about the uses of the data that these “new generation tools” yield. Although not all researchers feel that an absolute choice must be made, most lean heavily in one direction or another. One set of researchers believe that teacher evaluation should primarily be used summatively, to make determinations about teacher practice. These researchers often favor the use of teacher evaluation data to reward effective teachers in terms of differential pay and tenure decisions, 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. The second group of researchers argues that teacher evaluation should be used primarily as a formative tool, to help teachers learn about and improve their practice. These researchers often believe that teacher evaluation holds the most promise when it is linked to other teacher learning experiences, such as coaching and professional development. Further, this stance supports a view of teachers as central to school improvement. In contrast to summative proponents, formative researchers believe that teachers, like other adult learners, are capable of developing and improving their practice with the support of targeted feedback and positive professional environments in which adult learning is valued. Table 1 summarizes the two approaches.

Table 1

Summative and Formative Teacher Evaluation Methods

	Summative	Formative
Uses in Teacher Evaluation	<ul style="list-style-type: none"> Final Teacher Effectiveness Determinations Incentive Pay HR Decisions 	<ul style="list-style-type: none"> Coaching Professional Conversation Linked to Teacher Evaluation Professional Development Opportunities Linked to Teacher Evaluation Feedback
Key Research in Favor	<ul style="list-style-type: none"> Bill & Melinda Gates Foundation, 2013 Gordon et al., 2006 <u>Hanushek & Rivkin, 2006</u> Jacob, 2011 <u>Stronge, 2013</u> 	<ul style="list-style-type: none"> <u>Bambrick-Santoyo, 2012</u> Danielson, 2011 Goe, Bell, & Little, 2008 <u>Minnici, 2014</u> <u>Papay, 2012</u> Taylor & Tyler, 2012 Steinberg & Sartain, 2015
Associated Views	<ul style="list-style-type: none"> Effective teacher practice is definable and measurable Comparing observation results to student achievement yields valid assessments of teacher practice. Policy is necessary to change school ineffectiveness. Teachers can be motivated by sanctions and rewards. The teaching force can be improved by dismissing and replacing ineffective teachers 	<ul style="list-style-type: none"> Teachers are capable of more than their observed practice may reveal. Professional knowledge emerges from practice and conversations about practice. Teachers continuously learn new skills.

Arguments That Support Summative and Formative Views

The next section will describe the major arguments of researchers who mostly support summative or formative approaches to teacher evaluation. Although summative arguments are more prevalent in the teacher evaluation literature and policy dialogue as a result of widespread

support from influential reformers, they will be summarized briefly in this literature review.

Formative arguments will be presented in more detail as this study describes a formatively constructed evaluation system because of the researcher's belief that formative systems are more promising for educational reform.

Summative Approaches

Many researchers, particularly those who currently influence national policy, support the use of teacher evaluation tools as summative assessments of teacher practice (see for example: Gordon et al., 2006; Hanushek & Rivkin, 2006; Jacob, 2011; Stronge, 2013). Similar to proponents of formative uses, supporters of summative evaluation believe that teacher effectiveness is a crucial, if not the most influential, determinant of student achievement, often impacting students for several years. Further, many researchers on both sides of the teacher evaluation debate value current teacher evaluation tools and practices and see them as valid and reliable. Yet, those who argue that teacher evaluation should serve a summative purpose believe that valid tools and reliable approaches can help administrators reach the ultimate goal of differentiating between high quality and low quality teachers. Researchers in this camp cite many potentially positive outcomes of summative approaches. These include societal rate of return arguments like that of Hanushek (2010), a statistician and major contributor to policy formation around teacher evaluation, who proposes that a teacher who is "effective" can make a huge positive impact on student learning and increase students' earning potential as adults. He therefore urges politicians to gear educational policy towards improving student performance by lessening barriers to firing underperforming teachers (Hanushek, 2010) and implementing differential pay scales to reward high performing teachers (Hanushek & Rivkin, 2006). He contends that this can only be done reasonably with valid, summative evaluation data.

Others believe that summative measures can increase teacher effort. This is based on the assumption that many teachers do not strive for excellence unless they are held accountable for doing so through ongoing monitoring. Some point out that summative measures may encourage ineffective teachers to leave the profession. For example, a recent study finds that teachers who are rated more highly on teacher evaluations are more likely to be satisfied with their jobs (Koedel et al., 2014). Thus, it is assumed that teachers who are not performing well may be more likely to leave the profession, opening the possibility for more effective teachers to replace them. Proponents of summative teacher evaluation often also support the use of value-added models (VAMs) to make judgments about teacher quality. In value-added approaches, statistical models determine the impact of individual teachers on their students' standardized test scores. While summative arguments, including those that utilize VAMs, pervade current teacher evaluation dialogue, this literature review only provides a brief overview of them. Instead, because this study explores a formatively implemented evaluation system, I review formative arguments in more detail.

Formative Approaches

As proponents of summative teacher evaluation continue to urge more stringent sanctions to encourage improved instruction, others assert that teacher evaluation is best used formatively. As described previously, this “camp” utilizes three arguments when describing the importance of formative evaluation. The first of these claims that current teacher evaluation tools are limited in what they can capture regarding the complexity of teaching, and, therefore, should only be used as one source of data to inform conversations about effective teaching. At the same time, this argument advises a more comprehensive view of teacher practice that goes beyond the capacity of current teacher evaluation tools to capture teacher practice (see for example: Cohen, 2010;

Kennedy, 2010 a,b,c; Stodolsky, 1984). In contrast, the second and third arguments generally accept the validity of current tools and the reliability of the practices through which they are implemented. However, the second argument in favor of formative evaluation questions the legitimacy of using teacher evaluation summatively because of the risk of error and the unintended consequences (see for example: Baker, Oluwole, & Green, 2013; DeMatthews, 2015; Ford et al., 2015; Gabriel & Allington, 2012; Goe et al., 2008; Hallinger, Heck, & Murphy, 2014; Mielke & Frontier, 2012; Rothman & Toch, 2008). The third asserts simply that formative teacher evaluation data should be central to teacher learning because student learning is contingent upon the skills and abilities of the teachers who serve them (see for example: Cochran-Smith, 2010; Holdheide, & Miller, 2011; Hill & Grossman, 2013; Jerald, 2012; Minnici, 2014; Papay & Kraft, 2015; Sartain et al., 2011; Smylie, 2016; Steinberg & Sartain, 2015; Westerberg, 2013; Zepeda & Kruskamp, 2012). Each argument supports a shift to formative approaches in teacher evaluation for slightly different reasons; this section will describe each in further detail.

Arguing against policy: Teacher evaluation as a limited source of data. As reformers continue to study the ways in which to improve the validity and reliability of teacher evaluation, an unchecked theory-of-action seems to permeate the field (Smylie, 2016). It tries to convince us that evaluation protocols can become flawless representations of practice, given adequate validation, and that educational maladies are curable once teachers can be more accurately classified along a continuum of effectiveness. Yet most adopt summative or formative processes without questioning the theory-of-action that undergirds them. In this first argument in the formative “camp,” researchers urge practitioners to examine this theory-of-action. They propose

that teacher evaluation should be limited to one source of data that can simply be used to enrich formative conversation, but can never be relied upon as a completely valid or reliable tool.

Those who do doubt the teacher evaluation theory-of-action tend to question whether evaluation tools can fully identify what a teacher's skill-set is. To do so, they often differentiate between *teacher* quality and *teaching* quality (Kennedy, 2010 a, b, c). *Teachers* are individuals with multiple sets of abilities and motivations, whose work is irrevocably bound by the contexts in which they perform *teaching* acts. Their work is limited by the resources available to them, the cultures in which they teach, and the subject matters, textbooks, and standards that are dictated. Cohen, Raudenbush, & Ball (2003) explain that "What we casually call teaching is not what teachers do, say, or think ... teaching is what teachers do, say, and think with learners, concerning content, in particular organizations and other environments in time" (p. 124). When observers make decisions about instructional practice, therefore, they can only make determinations about *teaching* within a particular context. It is unwise to use that contextual determination as an assumption of a *teacher's* quality.

Stodolsky (1984) is the first to explore the distinction between teacher quality and teaching as a result of context. She reexamines the definition of reliability in teacher evaluation by questioning the widely-held belief that reliability implies that descriptions of a teacher capture the breadth of her actual practice. Rather, she limits reliability to the ability of multiple raters to notice the same events during a bound teaching period. To support this argument, Stodolsky (1984) used a team of raters to observe 20 teachers during math and social studies lessons over many days. Though observers rated individual lessons similarly, showing reliability to each other, teachers' instructional practices varied widely depending on the content area of instruction. This shows that, although observers were able to match each other's ratings during

individual sessions, the results were not reliable assessments of the teacher, herself. In essence, this means that teaching is at least as much an attribute of a given lesson and context than it is of a given teacher. A more recent study finds the same teachers scored differently when teaching reading as opposed to writing (Grossman et al., 2014), also suggesting that evaluation tools may be unable to capture the quality of teacher practice across content areas. Another finds that while certain teaching behaviors, like classroom management, are usually stable across contexts, the capacity to engage learners in higher level thinking varies (Praetorius et al., 2014). They suggest that raters must observe a minimum of nine teaching instances before they can begin to reliably rate this complex teaching activity. These findings discredit the summative use of teacher evaluation because evaluators can only capture limited and context-bound teaching through classroom observation, not the essence of a teacher herself. Therefore they argue that these limited tools cannot be used to make summative decisions and can only be used as one source of data from which to generate formative conversations (Stodolsky, 1984).

More recent studies also find distinctions between teachers and observed teaching quality over time. By comparing teachers over several years, they find that measurements of teacher performance using teacher evaluation tools (Baker, et al., 2013; Garrett & Steinberg, 2015; Goe, Bill, & Little, 2008; Hallinger et al., 2014; Jacob & Lefgren, 2006; Morgan et al., 2014; Schmoker, 2012; The New Teacher Project, 2012) and value-added models (VAMs) (Baker et al., 2010; Darling-Hammond, 2015a; Harris, Ingle, & Rutledge, 2014; Morgan et al., 2014; Rothman & Toch, 2008) are not stable over multiple years. They conclude, therefore, that results may be subject to measurement error, such that teachers rated as “ineffective” might be prematurely dismissed, while teachers rated “highly effective” may be unfairly rewarded.

Others contribute further to the teacher/teaching distinction by arguing that until teaching is distinguished as a professional field, teachers will not be able to fully showcase their skills and abilities, and, so, the practice of teacher evaluation unfairly disadvantages teachers who cannot control policies and underresourced contexts that undermine their work. Teachers may not have time to plan or access to appropriate instructional materials or professional environments within which to work (Cohen, 2010; Darling-Hammond, Jaquith, & Hamilton, 2012b; Ladson-Billings, 2009). Consequently, the teacher evaluation process holds them accountable for addressing challenges beyond their control (Darling-Hammond, 2013; Glassman & Paulin, 1982; Kennedy, 2010b). Kennedy (2010b) also argues that student characteristics are often more likely to influence observation results than are teacher qualities. Although this argument seems to suggest that teachers can blame low effectiveness on their students, Kennedy makes an important point about evaluator observations in relation to teaching contexts. Large scale findings that teacher observation ratings are higher for teachers serving gifted and talented students, and lower for those serving low-income and minority students (Chaplin et al., 2014; Steinberg & Garret, 2016) suggest further that observers' views of instruction are bound to student characteristics rather than teaching ability. When used as a summative determination of teaching quality, all of these findings indicate that teachers might be classified as ineffective when inadequacies in instruction are a result of systemic challenges and observer bias. More generally, it is not possible to make reliable determinations of a teachers' entire skill-set based on context-bound snapshots of teaching. Yet these determinations are often used to make decisions about teachers. Instead, this argument goes, teacher evaluation feedback should only be used as one source of data with which to generate conversations about instruction.

In addition, other studies question, not only the reliability of observations as they relate to the teacher's skill-set, but also the notion that observer training will eliminate threats to interrater reliability, or the likeliness that observers will similarly rate the same observed lessons (Gitomer, 2008; Jacob & Lefgren, 2006; Peterson, 2000). This is especially apparent when observers are rating complex aspects of instruction (Gitomer et al., 2014; Hill & Grossman, 2013). For example, Gitomer et al. (2014) compare ratings from five well-trained school-level observers with two expert raters from the research team using the mathematics CLASS-S, an extensively validated tool. Four of the five trained raters had previously taught secondary mathematics, while the fifth taught secondary English, indicating that most had a level of comfort with the content area, and all with the student population. Gitomer et al. (2014) find that trained raters are most able to agree upon and match expert ratings of classroom management strategies, but are unable to agree upon or line up with expert raters regarding instructional practices. This may mean that it is easier for observers to recognize student behaviors, which are often highly-visible, than it is for them to understand subtler teacher behaviors, such as pedagogical choices. This finding is especially important because it is those nuanced teacher choices that need to be addressed the most in order for teachers to incorporate more effective practices (Danielson, 2016). Again, this finding supports the assertion that teacher evaluation should be used as just one form of feedback for formative conversation in conjunction with better teacher and observer training about highly effective practices (Gitomer et al., 2014).

Others recognize that teacher evaluation results are highly dependent on several school and personality factors. Some find correlations between student assignment and evaluation scores (Rothstein, 2009 & 2010; Steinberg & Garrett, 2016) and conclude that without random rostering, teacher evaluation scores are subject to bias (Kane, Kerr, & Pianta, 2014b). Still

others find that scores vary based on factors such as content area, time of year, and grade level (Casabianca et al., 2015; Cor, 2011; Hill, Charalambous, & Kraft, 2012; Joe et al, 2014; Pianta & Hamre, 2009). For example, two large-scale studies find significant difference between middle school and elementary school scores (Grossman, Cohen, & Brown, 2014; Mihaly & McCaffrey, 2014). Mihaly & McCaffrey (2014) determine that these differences could not be explained by differences in teacher, student, or rater characteristics, thus indicating that ratings may be skewed in favor of elementary contexts. Others find differences at the individual level, such as vast differences between raters (Casabianca et al., 2013; Curby et al, 2011; Hill et al., 2012; Ho & Kane, 2013; Sartain et al., 2011) or even differences in single rater's scoring over time (Casabianca et al., 2015). On a macro scale, error may seem like an inevitable outcome of any measurement system. However, such high-stakes errors can profoundly impact individuals' lives and create systems in which all teachers feel that their fates are not being determined based on their best efforts. Hill et al., (2011) illustrate some of these problems using case studies of five teachers for whom value-added scores differed greatly from their content knowledge to illustrate the system level errors that could result from high-stakes decisions. All together, these findings call into significant question the core premise upon which summative evaluations are based.

In addition to questioning reliability, some have reservations about the validity of teacher evaluation tools because they may not encompass teacher characteristics that are actually important to learning. Kennedy (2010 a, 2010c) is a major contributor to this argument. In fact, in 2010, on the heels of national teacher evaluation reform, Kennedy edited a volume of research reports that issues major warnings about the state of teacher evaluation. Kennedy points to ever-changing, and often contradictory, definitions of good teaching (Kennedy, 2010a; Kennedy, 2010b). These include education level, value-added to standardized test scores, degree type, and

even individual attributes. Although these teacher characteristics may all seem meaningful, it has not been empirically established which qualities (or combination of qualities) matter most for children's learning (Kennedy, 2010b) and how this may vary across specific contexts of teaching (Robinson, 2015). Further, the conceptions are often combined, creating impossible, and inherently contradictory, standards by which teachers are judged (Kennedy, 2008). Thus it is unsurprising that Kennedy's (2010c) review of several current teacher evaluation tools reveals a wide variety of criteria for judging teacher quality. With so much variance, it may be possible that a teacher evaluated as "effective" using one tool may be rated as "partially effective" on another. Kennedy's work points out that teacher evaluation research has not yet validly defined teacher effectiveness, and, so, teacher evaluation tools cannot accurately determine teacher quality.

In all, these studies issue warnings about the pervasive belief that teacher quality can be reliably determined with available teacher evaluation tools. In addition, they caution that teacher evaluation tools may never be able to validly assess teacher quality because the term "teacher quality" is so multi-layered and context specific and, as of yet, undefined, and perhaps undefinable. As Cohen & Goldhaber (2016) argue "Most observation instruments...tacitly suggest that good teaching looks consistent across classroom, school, and district contexts" (p. 382). Yet it may be that definitions of teacher quality should be "contingent on a teacher's students and the school or district" (p. 382). As such, researchers and practitioners should first address the more real and pressing issue of the contexts in which teachers learn (Darling-Hammond, 2013; Kennedy, 2010b; Kremer-Hayon, 1993) and use teacher evaluation data as one source of information in formative conversations with teachers.

Arguing against policy: Unintended consequences of high-stakes uses of teacher evaluation. Whereas the previous studies argue that teacher evaluation tools cannot fully measure teacher quality, the second set of studies does not oppose the ways in which teacher evaluation has been constructed, but rather warns against fast-paced, high-stakes implementation of teacher evaluation while uncertainties about its application still exist. These studies generally accept the constructs of validity and reliability of teacher evaluation tools in theory, but recognize that the schools are so dissimilar from research settings as to introduce many unintended consequences. These studies, which favor formative uses, warn that summative approaches, especially those that lead to high-stakes decisions, should be employed with extreme caution. Some point to the conditions under which teacher evaluation was found to be a valid measure of teacher effectiveness, such as in the Measures for Effective Teaching (MET) Study. They argue that the studies are very dissimilar to real school conditions because they include extensive observer training (Harris et al., 2014) and student rosters are assigned randomly to teachers (Cohen & Goldhaber, 2016; Hallinger et al., 2014; Steinberg & Garret, 2016). As observation scores for teachers with higher-achieving students (Casabianca, Lockwood, & McCaffrey, 2015; Chaplin et al., 2014; Whitehurst et al., 2014) and in higher-achieving schools (Cohen & Brown, 2016; Cohen & Grossman, 2016) are often higher, the argument goes that realistic school conditions may result in skewed evaluation scores. Casabianca et al. (2015) make a strong case against summative approaches to teacher evaluation, stating “We should not tolerate a system that makes it hard for a teacher who doesn’t have top students to get a top rating” (pg. 3), and suggest that teacher observation continue to be used, but that scores be adjusted based on student demographics. However, this may also be a problematic argument because school change initiatives should decrease, not increase, differentiated standards of curricula and

instruction based on student populations. And yet, without first creating such environments in all of our schools, we should not punish or reward teachers.

Some who argue against high-stakes uses worry about unintended consequences of wide-scale sanctions and rewards. In this case, policy may be initiating programs that are unrealistic for real school contexts. Many are concerned that teacher evaluation will lead to huge demands on administrator time (Hallinger et al., 2014; Horng, Klasik, & Loeb, 2010; Kowalski & Dolph, 2015). Relatedly, others worry that because of the huge time demands, administrators will let go of other important practices that may be central to school effectiveness (Harris et al., 2014; Neumerski et al., 2014). Or, because administrators have many competing time demands, they may be making rush decisions based on a small snapshot of instruction (Evans, Wills, & Moretti, 2015). Evaluation may also begin to so narrowly define teacher effectiveness that administrators only look at outcomes (such as student achievement) rather than important teacher behaviors that can lead to heightened effectiveness, such as effort (Harris et al., 2014), professional judgment (Gabriel & Allington, 2012), and responsive teaching (Ford et al., 2015). Finally, when teacher evaluation is implemented as a summative rather than a formative tool, teachers may be more likely to disengage from professional learning and focus only on what is measured (Darling-Hammond, 2015), feeling that they are subject to top-down control rather than that they are a part of a collaborative endeavor to improve practice (Danielson, 2012; DeMatthews, 2015; Mielke & Frontier, 2012).

Others demonstrate that summative arguments about the ways that teacher evaluation will impact teacher instruction are misaligned with actual school conditions. For example, pro-summative teacher evaluation researchers have argued that teacher evaluations can be used to incentivize ineffective teachers to improve or as a basis for firing them, and that this will remedy

low student achievement. But, formative proponents point out that “firing 5% of [the worst performing] teachers (as summative proponents such as Hanushek (2010) suggest) does not guarantee any changes” to teacher effectiveness across school systems (Danielson, personal communication, 2015). Firing teachers does not guarantee that their spots will be filled with better performing teachers or that the teachers left behind will have access to learning that will heighten their effectiveness (Bambrick-Santoyo, 2012). Relatedly, incentivizing good practice does not help struggling teachers improve their practice; presumably, if they could do better without help, they would (Cohen & Goldhaber, 2016; Darling-Hammond, 2013; Moore-Johnson, 2015; Kane et al., 2014a; Whitehurst, Chingos, & Lindquist, 2014). It is therefore likely that simply firing the most ineffective teachers will improve the educational workforce overall (Kane, Kerr, & Pianta, 2014b; Martinez, Taut, & Schaaf, 2016).

Additionally, summative teacher evaluation proponents have claimed that summative approaches are more suited to school-level use. This assumption is based on the long-standing assumption that teachers, who may be more interested in job-security than effective practice and professional learning, will hide the troubling aspects of their instruction from observers (Popham, 1974; 1988; 2013). Popham argues that when teachers hide ineffective parts of their instruction, observers are unable to provide them with formative feedback and concludes that observers should only use summative approaches to evaluation because formative approaches are inconsistent with the nature of personnel practices in schools. However, Hunter (1982; 1988; 1995) argues that a skilled observer would “know for sure” what teachers’ capabilities are (Hunter, 1988, p. 2). Hunter’s assertion is supported by Sartain et al.’s (2011) correlational study of 280 teachers in Chicago. Sartain et al. Find that ratings do not vary largely between scores on pop-in verses pre-planned observations. Teachers could not drastically change their instructional

practices as a result of having more notice of an observation, including their creation of objectives and activities (their planning), and their ability to articulate and assess those objectives (their practice). This finding negates Popham's assertion that administrators can more easily implement summative approaches and the wider assertion that teachers can easily change their practices to comply with evaluation requirements. In all, the second set of studies in support of formative teacher evaluation questions the feasibility, stability, and consequences of summative approaches to teacher evaluation.

Arguing for the benefits of formative approaches for teacher learning and improvement. In addition to offering critiques of summative teacher evaluation, many stakeholders have used research to express the value of formative uses of teacher evaluation. At a macro level, although many vocal decision-makers in public policy favor summative approaches as a way to increase teacher effectiveness, interviews with state policy makers find that more than half view teacher professional development and support as either the main or a major intended outcome of new teacher evaluation systems. This suggests that they see the value of formative evaluation (Herlihy et al., 2014). Benefits for policy makers include a happier teaching force and better trained teachers in schools.

Principals and school administrators also favor formative approaches to teacher evaluation (Dretzke et al., 2015; Reinhorn et al., 2015). Many benefits are associated with it, including teacher buy-in to the evaluation process (Bryk, 2010) and an ability to create and adjust responsive professional development to teacher needs (Chappuis & Chappuis, 2007), both of which can lead to true change at schools (Fullan, 2007; Reeves, 2009). For example, approaching evaluations with a growth mindset influences how observers and teachers reflect on and discuss observations (Danielson Group, 2013). Additionally, tapping into the formative

potential of these tools is a better use of financial and human resources for districts who have already invested large sums of money in teacher evaluation (Chambers, de los Reyes, & O'Neil, 2013) and who are already engaged in the time-consuming processes of observation, data collection, analysis, and interpretation (Danielson, personal communication; Frase & Streshly, 1994; Hunter, 1988).

Teachers most strongly support the formative uses of teacher evaluation (Bonavitacola, 2014; Darling-Hammond, 2014; Donaldson, 2009; Dretzke et al., 2015; Goe et al., 2008; Marzano, 2012; Reinhorn, 2013; Reinhorn et al., 2015; Wise et al., 1985). Teachers want organizational cultures that encourage their retention in the profession, including decreased isolation (Darling-Hammond, 2012; Dretzke et al., 2015; Greenberg, 2015; Johnson, 1984). Some administrators eschew teacher evaluation because they fear teacher reactions to negative feedback. However, teachers report that they are not as dissatisfied with receiving negative feedback as they are displeased by the infrequency of receiving it (Natriello & Dornbusch, 1980) and criticize evaluation processes when they do not receive detailed and useful feedback (Darling-Hammond, 2013; Reinhorn, 2013). Relatedly, Marzano, Frontier, & Livingston (2011) argue that frequent observation and feedback leads to less apprehension among teachers as the teacher evaluation process becomes normalized. Further, while many law-makers believe that they can increase effective teaching through monetary incentives, teachers oppose differential pay (see for example: New Jersey Education Association, 2010), a potential outcome of summative evaluation; instead, they crave environments in which their professional growth is central. Systems that utilize formative evaluation may be able to foster such environments.

Finally, many researchers believe that feedback is the most important role of teacher evaluation because feedback can help teachers at every career stage to improve their classroom practice (Danielson & McGreal, 2000; Greenberg, 2015; Hill & Grossman, 2013; Sartain et al., 2011; The New Teacher Project, 2012). In fact, Fullan (2005) describes formative evaluation as “one of the most high yield strategies for changing teaching and learning” (p. 54). Like any other complex and difficult task, teaching requires practice with specific and immediate feedback in order to develop competence or expertise (Aldeman & Chuong, 2014; Ericsson, 2009). Summative approaches often focus most on the lowest performing teachers, which means that teachers who are performing well, particularly experienced teachers, gain nothing from the evaluation process (Danielson, 2011a; Darling-Hammond, 2012; Donaldson, 2012; Goe et al., 2008). Conversely, many studies have found that even highly-experienced teachers improve through targeted feedback in formative evaluation systems (Bonavitacola, 2014; Kelley & Maslow, 2012; Papay & Kraft, 2015; Reddy & Dudek, 2014; Steinberg & Sartain, 2015). For example, Kelley & Maslow’s (2012) study finds that “non-problematic” teachers, who had previously “plateaued” in their learning (p. 2), developed better classroom practices as a result of formative feedback. Others find that formative teacher evaluation significantly relates to improved student test scores (Dee & Wyckoff, 2015; Gallagher, 2004; Jerald, 2012; Kimball et al., 2004; Milanowski, 2004). The most frequently cited study that demonstrates the relationship between evaluation feedback and student growth is Taylor and Tyler’s (2012) correlation analysis which was conducted with mid-career teachers. In addition to finding that the mid-career teachers are able to improve their instructional strategies, they find that teachers appear to generate higher student test-score gains during the year they receive formative evaluation feedback. These improvements strongly suggest that teachers develop skills or otherwise change

their behavior as a result of receiving formative evaluation feedback, even late in their careers. More recently, Steinberg & Sartain (2015) found that improvements based on formative were most rapid in high-performing, low-poverty schools. They conclude that formative approaches to evaluation should be more frequently used in high-poverty schools, such as this study's site, as changes to teacher practice in those schools is less likely to improve from limited exposure to feedback.

School-Level Practices that Support Formative Evaluation

While researchers in the formative camp make an argument for the importance of formative approaches, few have detailed how schools can create formative systems. This section of the literature review describes the school-level practices associated with formative evaluation. As a new body of research, the uses and processes associated with effective teacher evaluation in formative ways are still emerging and lack coherence. 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). For example, 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 focused on in-service teachers, with the rest focusing on pre-service teachers. Recently, researchers have written about formative approaches for in-service teacher supervision, but to date the area of investigation is loosely-constructed. Although disjointed, I will use this area of the research to support my research design. I adopt 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. In this section of the literature review, I will introduce five elements that increase the success of formative supervision: evaluator training,

distributing evaluator caseloads, giving formative feedback, linking professional learning to teacher evaluation data, and sustaining school level change.

Evaluator Training

In order to implement effective formative processes that engage teachers and increase professional skills across schools, schools must provide specific types of evaluator training. This training begins with shifting observers' perspective regarding evaluation. As discussed previously, there has been a considerable amount of research on increasing evaluator reliability through training on the tools themselves, yet there is a lack of training on how to use teacher evaluation results to give feedback as a way to create learning systems for teachers (Goe et al., 2012; Greenberg, 2015; Grissom et al., 2011; Kraft & Gilmour, 2016a; Herlihy et al., 2014; The New Teacher Project, 2012). Creating systems that provide high-quality feedback is "more challenging than simply increasing the frequency of observations" (Aldeman & Chuong, 2014, p. 7) or increasing the amount of training observers receive on increasing reliability (McKay & Silva, 2015). It requires that evaluators first shift their evaluation orientation toward an understanding that it can be used to help teachers to grow, not just to judge their efficacy, (Reinhorn, 2013) across career stages (Darling-Hammond, 2013; Harvey & Holland, 2013). It requires that evaluators know how to give constructive, actionable feedback (Darling-Hammond, 2013; McKay & Silva, 2015). 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 buy-in is critical to helping teachers to buy-in to teacher evaluation reform (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) and to the success of implementation practices (Sartain et al., 2011).

There are also certain skill-sets that observers must have in order to provide high quality feedback. In addition to being able to collect valid, unbiased data from classroom observation as a result of high quality training, observers must learn how to draw on their knowledge of teaching and learning (Darling-Hammond et al., 2012b) and be able to push teachers beyond their natural abilities (Drucker, 2001; Hunter 1988). This requires that evaluators understand what each teachers can presently do, and are ready to learn next, and how each might best acquire the necessary knowledge and skills to become more effective (Hunter, 1988). Training should be sustained over multiple years. 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 the evaluators’ familiar areas of content (Donaldson, 2012; Kelley & Maslow, 2012; Murphy et al., 2013; Reinhorn, 2013; Rotherham & Mitchel, 2014; Kraft & Gilmour, 2016a). These needs are heightened the longer the gap between when evaluators last taught and the time when they begin conducting evaluations (Murphy et al., 2013).

School districts should be mindful that observation and feedback skills are not easily or quickly acquired, so they should budget appropriate resources, including money, time, and people to both training observers and assessing the impact of those trainings (Derrington et al., 2015; McKay & Silva, 2015). Many researchers have explored the 3-5 year time lag between initiating a change and teachers fully implementing new programming (Fullan, 2007; Reeves, 2009); educational leaders similarly need targeted support during the initial 3-5 years of rollout, not simply at the initiation of new teacher evaluation programming (Derrington & Campbell,

2015; Murphy, Hallinger & Heck, 2013). Evaluators report being consumed with learning how to collect observational data and filling in paperwork during the first year of teacher evaluation reform. Thus they need targeted support in later years (Derrington & Campbell, 2015; Derrington et al., 2015; Dretzke et al., 2015). Administrative training can also be more successful when administrators rely on each other to create knowledge and share information (Hatch, Hill, & Roegman, 2016). Some researchers recommend that schools hire evaluation consultants to address these needs (Darling-Hammond, 2012; Reinhorn, 2013). In addition to including initial training on observation and feedback skills, coaches can provide ongoing support for evaluators (Jerald, 2012) and audit their activities during periodic calibration sessions (Darling-Hammond et al., 2012b; Jerald, 2012; Joe et al., 2013; Kane et al., 2014a; McKay & Silva, 2015; Park et al., 2014; Sartain et al., 2011).

Manageable Evaluator Workloads

In addition to better evaluator training, schools need to structure evaluator work so that they have adequate time to positively enact formative supervision (Kelley & Maslow, 2012; Kraft & Gilmour, 2015; Reinhorn, 2013). Horng et al. (2010) found that due to the multiple demands of administrative work, even highly effective principals could only devote 6% of their time to instructional leadership activities. Within such constraints, principals report difficulty keeping up with observational schedules, including coordinating observation times, meeting deadlines, and scheduling pre- and post-observation conferences (Rotherham & Mitchel, 2014; Sartain et al., 2011). Teachers express frustration with this reality, particularly when observations had to be rescheduled after planning and pre-conferences (Sartain et al., 2011). Furthermore, lack of time can lead to evaluators falling back on generic feedback (Kraft & Gilmour, 2016a; Halverson et al., 2004; Rotherham & Mitchel, 2014). It may also limit

opportunities to observe each teacher. This can decrease evaluation validity as fewer observational episodes are less likely to capture teacher practice (Harvey & Holland, 2013; Holland, 2005; Sartain et al., 2011); teachers also state that they view decisions about teacher practice based on fewer observational sessions as less valid (Donaldson, 2012; Reinhorn et al., 2015).

It is important that school leaders have the time to conduct all elements of the observation effectively by structuring their job descriptions differently (Derrington & Campbell, 2015; Louis et al., 2010), decreasing evaluation caseloads (Darling-Hammond, 2013; Donaldson, 2012; Jerald, 2012; Kraft & Gilmour, 2016a; Riordan et al., 2015), offloading other administrative duties (Jerald, 2012; Murphy et al., 2013; Reinhorn et al., 2015; Rotherham & Mitchel, 2014), or hiring additional personnel specifically to conduct evaluations (Clayton, 2013; Gold, 1996; Louis et al., 2010; Rotherham & Mitchel, 2014). However, if hiring personnel, systems must be created to inform building-level leaders of evaluation findings so that they can engage other important aspects of formative supervision, such as creating professional development programs linked to teacher evaluation results. One problematic suggestion to lessen principal time constraints is to differentiate the number of evaluations or feedback sessions based on teachers' previous evaluation scores, such that teachers who were rated "highly effective" in the previous year would only be formally evaluated one time, while other teachers would be evaluated three times (Danielson, 2015; Darling-Hammond, 2013; Hill & Grossman, 2013; Holland, 2005; Rotherham & Mitchel, 2014). It is likely that differentiating the teacher evaluation process in this way may lead to a system similar to that which Weisberg et al. (2009) describe in which principals feel site-based pressure to apply the least demanding evaluation processes for all teachers.

Enacting Formative Feedback

Evaluators must also understand how to properly construct and communicate formative feedback. There is a wealth of guidelines that relate to best feedback practices from both the student 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 considered to be ongoing (Chappuis & Chappuis, 2007; Coggshall et al., 2012; Goe, 2013; Harvey & Holland, 2013; Shute, 2008; Westerberg, 2013; Wiggins, 2012). Many administrators feel pressure to provide feedback immediately. For example, schools might enforce a 24- hour turnaround for feedback. However, it may be that it is less important to provide immediate feedback than it is to provide well-constructed feedback within a reasonable timeframe. 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). Additionally, it should be shared during ongoing, responsive, two-way dialogues that specifically encourage teacher self-reflection (Anast-May et al., 2011; Chappuis & Chappuis, 2007; Costa & Garmston, 2002; Danielson, 2001; Goldhammer, 1969; Kremer-Hayon, 1993; Marzano et al., 2012; Ovando, 2005; Range et al., 2013).

Evaluators should encourage teachers to do more talking than they do during feedback sessions (Danielson, 2011; Sartain et al., 2011). In this way, feedback can heighten teachers' ability to understand the outcomes of their decisions about their own practice and to approach new circumstances that constantly arise more thoughtfully, rather than simply to implement administrator recommendations. This can help teachers develop a mindset of thoughtfulness that takes place before, during, and after instruction (Berry & Loughran, 2000; Dewey, 1933; Trip, 2011).

When feedback is presented in these ways, teachers are more likely to view it as positive and constructive. Kraft & Gilmour (2016) observe that when moving away from systems where all teachers are rated as satisfactory, feedback conversations can be especially difficult as teachers have become accustomed to evaluations that do not challenge any aspects of their practice. Kraft & Gilmour (2016) therefore argue that discussions that focus on teacher reflection may allow for these conversations to feel less judgmental as teachers are asked about their practice rather than told about their shortcomings. Formative feedback is also more likely to be educative when the teacher has a positive relationship with the administrator (Shute, 2008).

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 view skills as improvable through practice (Shute, 2008). In schools, where attrition rates are as high as 30% (Ingersoll, Merrill, & Stuckey, 2014; McCreight, 2000), it may be easier to retain teachers if they believe that it is possible to 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, it should be based on frequent observations of teacher practice (Hanna, 2013; 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 practice lines up with expectations and also reinforces a common instructional language that can be used to increase communicative effectiveness.

In addition to being specific and accurate, feedback should be directly related to goals that are created collaboratively by the teacher and the evaluator; goals provide direction on steps a teacher might take to reach them over time (Chappuis & Chappuis, 2007; Coggshall et al., 2012; Danielson & McGreal, 2000; Darling-Hammond, 2013; 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). Setting goals and then practicing with targeted feedback allows teachers to develop expertise. Ericsson (2009), who studies how experts in several fields develop their practice, finds that expertise develops as a result of concentrating on carefully selected, specific aspects of performance and refining it through repetition and response to feedback. However, while providing feedback, evaluators must be aware of teachers' cognitive load (Chappuis & Chappuis, 2007; Shute, 2008; Westerberg, 2013). Thus researchers recommend 1-3 goals per teacher feedback session (Bambrick-Santoyo, 2012b; Westerberg, 2013).

Linking Professional Learning

Much of the available research also 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& 2013; Darling-Hammond, Amerein-Beardsley, & Haertel, 2012a; Hill & Grossman, 2013; Holland, 2005; Looney, 2013; Marshall, 2013; Smylie, 2016; Zepeda, 2007). For example, Danielson, whose evaluation framework is the most-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 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).

Teachers should be central to successful professional learning. First, professional learning should allow teachers to engage in topics that relate directly to their work (Contreras, 1999). This potential for accomplishing 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). Second, as teacher work is characterized by constant and weighty decision making, (Ball & Cohen, 1999; Berry & Loughran, 2002; Danielson, 2012, 2015; Hunter 1988; Kazemi, Lampert, & Ghouseini, 2007; Tripp 2011), teachers must have opportunities to understand those decisions. By talking about through those decisions outside of instructional time, teachers can understand them better in order to respond in better ways when similar problems arise in the future. The act of reflecting on past decisions may also help teachers to develop a reflective stance when approaching new problems as they arise. Thus, professional development should help teachers to increase their capacity for self-reflection by exploring authentic problems of practice (Coggshall et al., 2012; Darling-Hammond, 2012; Darling-Hammond et al, 2012a). To do so, teachers need to view their instruction as constantly evolving and open to scrutiny (Shah et al., 2015), and view others as central to their reflective practice (Behrstock-Sherratt et al., 2013; Holland, 2005; O’Pry & Schumacher, 2012; Sartain et al., 2011).

Additionally, professional development that is linked to teacher evaluation data should follow research recommendations for best practice. Just as for principals, professional

development for teachers should initially focus on the purpose of evaluation tools and procedures (Heyde, 2013; Holland, 2005) and include formal instruction on how to have meaningful conversations in relation to teacher evaluation data (Jerald, 2012; McKay & Silva, 2015; Murray, 2014; Sartain et al., 2011; Shaha et al, 2015). Professional development should also provide opportunities for active learning (Conley et al., 2014; Danielson, 2010; Merriam, & Bierema, 2014) and successful models of new teaching strategies, followed by opportunities for practice and more feedback (Archibald et al., 2011). In this way, professional development can allow teachers to understand their own practice in relation to proposed learning. Special attention should be given to the needs of distinct groups including novices (Clayton, 2013; Darling-Hammond, 2013; Darling-Hammond et al., 2012b) and high-performing teachers (Donaldson, 2012). These goals are most easily incorporated when teachers are engaged in sustained dialogue with other teachers, such as in professional learning communities (PLCs) (Behrstock-Sherratt et al., 2013; Bonavitacola, 2014; Davis et al., 2002; DeMatthews, 2015; Reinhorn, 2013) that honor teachers as experts in their own craft (Cochran-Smith, 2010; Mielke & Frontier, 2012; Minnici, 2014; Sykes & Winchell, 2010). When formative teacher evaluation feedback is presented with professional learning opportunities that meet these recommendations, formative evaluation can become “powerfully connected to instructional improvement” (DeMatthews, 2015, p. 83).

Outcomes of Teacher Evaluation Policy: Limited School-Level Changes and Exemplars

Ideas about teacher evaluation have noticeably changed since Race to the Top was initiated in 2009. Whereas most schools were previously using behavioral checklists about teaching, or rating all teachers as, simply, “satisfactory,” new directions in teacher evaluation have emerged that incorporate better descriptions of effective teaching and more reliable

processes for identifying teachers practice in relation to those descriptions. Many states have implemented new or more comprehensive teacher evaluation systems (Boser, 2012) and researchers have a clearer idea of the positive potential of teacher evaluation (Bill & Melinda Gates Foundation, 2013). Schools have invested considerable money and time to implement new evaluation systems. Stakeholders have become more vocal about the importance of teacher evaluation and there has been increased dialogue about it.

However, although the policy dialogue related to teacher evaluation has shifted, federal policy has not strongly impacted teacher evaluation practices in schools. Federal policy is often implemented with tremendous variety; and this is the case with teacher evaluation as well (Camburn & Han, 2015; Cohen & Mofitt, 2010). While school administrators report that they hope to promote both teacher development (formative) and individual teacher accountability (summative) (Kraft & Gilmour, 2015; Steinberg & Donaldson, 2016), most have not been able to do either well (Darling-Hammond et al., 2012a). For example, observers still rate most teachers similarly (Doherty & Jacobs, 2015; Dynarski, 2016). For example, Kraft & Gilmour (2016b) found that in teacher evaluation ratings across 19 states, evaluators continue to rate all teachers as proficient, even when they have observed ineffective practice. Evaluators give several justifications for skewing scores including not having enough time to follow through on action plans for poor-performers, feeling as though a low rating would be unfair due to the lack of time to provide support, fearing the impact on school culture, of firings and being unsure about the possibilities of finding better candidates to replace fired teachers. This engendered a triage view where only the lowest performing teachers were given appropriate supports.

In a limited number of studies, researchers did find fully implemented summative (Alderman & Chuong, 2014; Dee & Wyckoff, 2013; Kelley & Maslow, 2012; Milanowski &

Heneman, 2001; Steinberg & Sartain, 2015) and formative evaluation systems (Derrington & Campbell, 2015; Donaldson, 2013; Donaldson et al., 2014; Range et al., 2013; Sartain et al., 2011). In one such study, Sartain et al. (2011) report that teachers and administrators report that evaluation conferences are more reflective, less subjective, and more likely to include practice opportunities when an evaluation rubric is used than they were with previously used checklists.

However, conversations about practice are still dominated by principal talk and low level questions. A few small scale studies have even found that teacher practice improves as a result of targeted feedback (Reinhorn, 2013; Sartain et al., 2011; State Collaborative on Reforming Education, 2012; Taylor and Tyler, 2012). In general, however, researchers find limited examples of positive uses of teacher evaluation in schools after the implementation of Race to the Top, and these were often idiosyncratic to particular school-level factors rather than related to the ostensibly powerful thrust of federal legislation. For example, Aldeman & Chuong's (2014) literature review reports that most districts are still unable to identify differences in teacher effectiveness, although some have shown progress in recent years. Large scale studies find that few, if any, schools have been able to implement even basic requirements of high quality summative or formative systems, such as observing teachers multiple times throughout the year (Callahan & Sadeghi, 2014; Cohen & Goldhaber, 2016; Donaldson, 2013; Firestone et al., 2013; Hamilton, et al., 2014; Hill & Grossman, 2013; Mitchell & Purchase, 2015; Reinhorn, 2013; Turnbull et al., 2009). This was even the case under ideal circumstances, such as in MET study partner schools, where administrators and teachers were given many opportunities and resources to help them properly implement teacher evaluation (Hamilton, et al., 2014). Further, in some schools where proper evaluation implementation was originally documented, administrators lost energy and commitment over time, decreasing effective implementation

(LaRocco et al., 2015). Donaldson (2012) even found that some schools have done less teacher evaluation after the implementation of Race to the Top than before. In all, these findings indicate that federally-mandated teacher evaluation has generally not had an impact on schools' implementation of teacher evaluation. This may be because it is simply too early to see change (Fullan, 2007; Reeves, 2009), but it might also be an indication that schools are not being purposeful about changing teacher evaluation approaches (Hill & Grossman, 2013) as a result of poor leadership, or a lack of commitment to the changes. Another factor may be that while states have established expectations for implementing teacher evaluation, few have offered specific guidance on how to do this (Ford et al., 2015).

In addition to mixed, but mostly negative, evidence about implementation, there are many mixed teacher and administrative perspectives regarding teacher evaluation. In some schools, teachers still view teacher evaluation as unfair (Mitchell & Purchase, 2015; Windish, 2012) and see limited utility in the feedback that tools generate (Thomson, 2014), while in others, teachers support new teacher evaluation systems (Donaldson, 2012; Robinson, 2015). Some principals also have negative views about teacher evaluation, including skepticism about its validity and usefulness (Kowalski & Dolph, 2015), while others believe that it holds merit for instructional improvement (Reinhorn et al., 2015).

Conclusion and Chapter Review

This review of literature has argued that while most agree that valid tools and reliable processes are necessary to conduct effective teacher evaluations, there is widespread disagreement, not only about how to create them, but also about the best ways to use the data they generate. Many researchers support summative uses including sanctions and rewards, while others call for formative uses that are linked to meaningful professional learning. Also, although

wide-scale disagreement between researchers and policy-makers has led to many reform initiatives and suggestions for school practice, school implementation has remained largely unchanged, supporting previous findings that policy success is largely dependent on school implementation (Honig, 2006; Lipsky, 2010; Spillane & Kenney, 2012).

The review of literature was used to help situate this study and determine a school site, create the sampling method, and data collection protocols. Initially, as the researcher, I believed that formative approaches to teacher evaluation were an untapped potential for teacher learning and support that could lead to maximized teacher quality and more effective practices within schools. As the discussion of my findings will describe, my view shifted as a result of data collection. However, during the time that I was constructing my study, I pulled largely from the preceding literature and situated my study within the outlined conception of formative supervision. I also used the literature review to create deductive codes used to conduct data analysis. Chapter Five will discuss modifications to this review based on findings from my study.

My study confirmed many findings from previous research about the importance of formative supervision in setting up effective formative evaluation practices. This included providing distributed evaluator caseloads, evaluator training, a focus on formative feedback, and linked professional learning. In addition to the four areas identified in the literature, my study introduces the importance of three structures: a centralized growth-model, integration of other policies and procedures with teacher evaluation practices, and creation of collaborative practices that supports knowledge sharing between teachers.

CHAPTER THREE: 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). In my literature review, I argued that teacher evaluation research and policy seem to be moving towards more formative uses of teacher evaluation, however there are few examples of formative supervision practices that fully utilize teacher evaluation feedback to help teachers grow and develop, and 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), as is the success of any new educational policy or practice (Honig, 2006; Spillane & Kenney, 2012). Therefore, this study provides guidance to schools that seek to implement formative teacher evaluation systems as it presents an example of one such well-implemented system.

Given my research questions, my first task was to search for a case that would illustrate formative supervision practices and might serve as a model for effective formative evaluation uses. In order to be an appropriate site, it needed to incorporate a formative approach to teacher evaluation that includes comprehensive 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 (Hanna, 2013; Hunter, 1988; Jerald, 2012; Marshall, 2013;

Ovando & Harris, 1993; Westerberg, 2013; Zepeda & Kruskamp, 2012), and targeted professional development based on teacher evaluation findings (Archibald et al., 2011; Coggs et al., 2012; Danielson & McGreal, 2000; Darling-Hammond, 2012; Darling-Hammond et al., 2012a; Hill & Grossman, 2013; Holland, 2005; Marshall, 2013; Zepeda, 2007). My case study design allowed me to explore what these supports look like in practice and to try to understand how teachers are using feedback within this setting to better understand, approach, and improve their classroom practice.

Case studies require that the researcher use multiple sources of data to understand several facets of the studied phenomenon through different lenses (Creswell, 1998; Merriam, 2009; Gall, Gall & Borg, 2010). This allows the researcher to understand the case in richer, deeper ways while discovering emergent themes (Creswell, 1998). For this study, I collected and analyzed principal and teacher interviews, a teacher focus group, and school documents and videos to more fully understand how teachers perceive and use evaluation feedback and what school-level factors mediate that use at “Elmwood School” (pseudonym). All interviews and the focus group were semi-structured. Semi-structured interviews were useful for this study because they enabled me to prepare questions about important topics and themes in advance while allowing flexibility during the interview process. This flexibility had two important advantages. First, it allowed me to relate more naturally to the participants, increasing participant comfort. Second, I was able to tailor questions to the interview context and respondents, such that I could probe responses more deeply (Merriam, 2009).

Site Selection and Research Context

To identify an appropriate case, I used personal conversations and recommendations from experts in the field, such as the head of professional development at the Danielson

Group. Nonetheless, after speaking with some of the school leaders of recommended schools, I did not find a site that I felt was fully incorporating research-based practices for formative evaluation. Luckily, in the midst of trying to identify an appropriate research site, I met a principal from a charter school in Washington D.C. while participating in a professional development experience. We struck up a conversation about teacher evaluation during which he spoke at length about very innovative ways they were implementing it at his school. For example, he described extensive administrator training to conduct formative teacher evaluation that included observation and feedback calibration sessions wherein administrators observed classrooms together or watched post-conference videos together and then discussed strengths and weaknesses as a way to continually improve practice. The principal also described supports for teachers use of their evaluation feedback that he believed led to improved practice. These include weekly observations and feedback conferences that were directly linked to professional development opportunities at the school and the use of classroom video to demonstrate “benchmark” teaching examples for each section of the teacher evaluation rubric. I felt that this site might exemplify promising practices identified in my review of the literature and I wanted to understand the degree to which those practices were happening at the school and the ways in which that might impact teacher learning from the evaluation process. So I requested the opportunity to meet with administrators and teachers over the course of four months to conduct my study.

“Student Success” (pseudonym) Charter Schools were founded in 1997 and opened two campuses in Washington D.C. in 1998. The school site for this study, Elmwood Elementary and Middle School, was one of these original campuses. The number of Student Success Charter Schools has grown through the years; now there are many schools that serve students from PreK

to 12, one of which is a former public high school that invited the charter to take it over. The Student Success Schools seek to prepare all of their graduates for higher education and careers by helping them meet high academic standards while developing the whole child (Student Success Schools, 2015). The schools also have many community outreach programs to support their students' growth.

Elmwood Elementary and Middle School serves 480 students in grades PreK- 8. In addition to the academic programs offered at all of the Student Success Schools, Elmwood offers the International Baccalaureate (IB) program to its elementary school students. All students at the school are Washington, D.C. residents and many come from high poverty backgrounds; 86% of students receive free or reduced lunch. The student body is 96% Black and 4% Hispanic. During the 2015-2016 school year, the school employed 35 full-time teachers in grades PreK-8; six were first-year teachers and 16 had more than five years of teaching experience. As an urban school, Elmwood School employed teachers from alternate teacher preparation programs for underprivileged schools, including two teachers from Teach for America and eight from the Urban Teachers program. Elmwood School has received many accolades from local and national government agencies, including a Blue Ribbon Schools nomination. The school has been visited by many dignitaries including former President George W. Bush and former Secretary of Education Margaret Spellings.

The Student Success Schools' teacher evaluation system is based on the District of Columbia Public Schools' (DCPS) model. As a publically funded charter, the district is mandated to include some summative aspects of DCPS's evaluation system, including formally observing and ranking teacher practice three times a year and including student achievement data in final determinations of teacher performance. However, they did have leeway in formative

aspects of the evaluation system. They were allowed to create their own teacher evaluation tool, and chose to modify and expand upon, DCPS's tool, a rubric-based system that includes descriptions of teacher professional roles and instructional practices. The Student Success district began to implement the new rubric in 2010, which is three years earlier than Washington D.C.'s public district mandated use of the teacher evaluation tool (Doherty & Jacobs, 2015). In addition, Elmwood School began to incorporate many formative supports, highlighted earlier, in 2013. Though Washington D.C.'s district has received many funding and professional development opportunities related to its teacher evaluation system (Wexler, 2013), as a charter district, the Student Success Schools were not included. Instead, the district's executive team creates all training opportunities for teachers and administrators.

Participants and Sampling Strategy

In order to understand school-level leadership factors and teacher perceptions and uses of formative teacher evaluation, I included both administrators and teachers in my sample. The sample includes the school principal, one supervisor of instruction and twelve teachers representing different academic content areas and grade levels. Participants were all recruited by the principal to participate in my study. We worked together to choose a group of teachers that varied by grade level, content area, and years of experience. By using a maximally variant sample, I was able to describe diverse cases in depth, leading to a richer understanding of the experiences of an array of participants in relation to the central issues and themes that emerged through analysis (Patton, 2002) and to compare and contrast cases across a variety of participant experiences (Patton, 2002).

As I had never interacted with most of the staff at Elmwood, I asked the principal to facilitate teacher and administrator recruitment. He forwarded to the staff an initial email I

drafted that explained the purpose of the study, why I chose Elmwood School as a study site, and participant responsibilities. I hoped that introducing these aspects of my research would allow me to begin to build trust and encourage participation. I did not receive a lot of responses to my email. So, instead, the principal purposefully selected a group of teachers that he believed ranged in effectiveness, grade level and content area, and years of experience to participate in the study. Unfortunately, as will be described in more detail in Chapter Four, by the time I conducted my interviews, both of the school's assistant principals were no longer working at Elmwood. The principal was able to facilitate an interview with another school administrator. That interview provided a lot of rich detail, but the study would have been enriched by more interactions with other administrators.

The focus group was conducted on the first site visit with early childhood teachers. Interviews with five teachers were to be conducted on both site visits, however one participant, Olivia, was unavailable for the 2nd round of interviews. The principal also facilitated a meeting with a new teacher, Heather. Participants are listed in Table 2.

Table 2:
Participant Information

Pseudonym	Title	Additional Background	Participation Format
Andre	Principal	M.Ed. in Urban Leadership	2 Interviews
Leo	Special Education Coordinator	M.Ed. in Educational Leadership	1 Interview
Lori	Preschool Teacher		Focus Group
Janelle	Preschool Teacher		Focus Group
Patricia	PreK Teacher	Early Childhood Lead	Focus Group
Pamela	PreK Teacher		Focus Group
Allison	PreK Teacher		Focus Group
Lisa	Kindergarten Teacher	Urban Teachers Program	Focus Group
Heather	K-2 nd Grade Reading Specialist	M.Ed. in Special Education	Interview 2 Only
Wendy	3 rd Grade, All Subjects	M.Ed. in Elementary Education	2 Interviews
Olivia	4 th Grade, All Subjects		Interview 1 Only
Denise	5 th Grade Math and Science		2 Interviews
Kathy	6 th Grade Math and Science		2 Interviews
Emily	7/8 th Grade Science	Teach for America Program District Master Teacher	2 Interviews

Data Collection Procedures

Data collection consisted of four tasks that were completed over the course of three months and two school visits. I began data collection during a two-day visit to the school in April. During the first day of the visit, I conducted one-on-one interviews with the principal (about one hour) and the administrator (about 30 minutes), and held a focus group with six early childhood department teachers (about 45 minutes). I used all three interviews to tweak the teacher interview protocol for day two data collection. On day two, I conducted short, introductory interviews (about 20 minutes each) with five teachers. These five teachers became

the focal participants for subsequent data collection. I also toured the school building and used a field journal to collect my reflections and informal observational data about interactions and activities related to teacher evaluation.

In the interim between visits one and two, I was given access to the school's Google Drive evaluation folder (described below). I compiled data about the five target teachers and about school-wide programs relating to teacher evaluation.

In June, I visited the school for my second in-person visit. At that time, I interviewed four of the five targeted teachers using individually-tailored interview protocols based on the first interview and the Google Drive findings. The interviews lasted about one hour each. Unfortunately, Olivia missed her meeting time with me, and my attempts to email her went unanswered. I tried to schedule a phone interview, but we could not coordinate anything. Instead, I was also able to interview one new teacher, Heather, who the principal thought would add some interesting insight to my study. The principal was also out on the day of my visit due to end of the year administrative meetings. I was able to connect with him via a phone interview in July during which we spoke for about 30 minutes.

Administrator Interviews

The purpose of the administrator interviews was to understand administrator perspectives on the teacher evaluation procedures of the school and for principals to describe the components and outcomes of the teacher evaluation system at Elmwood School. The interviews focused on administrator confidence and comfort in evaluating teachers and their perceptions about the teacher evaluation program. Both administrators were interviewed in one-on-one interview sessions. Individual sessions for administrator interviews were appropriate because I wanted each administrator to feel comfortable describing his perspectives without concern for what

others felt or fear of reprisal for dissenting views (Merriam, 2009). This was especially important because the principal was instrumental in creating the teacher evaluation system, so other administrators might not feel comfortable expressing negative opinions in a group session. Most sessions were held in each administrator's office to increase participant comfort. The final principal interview was conducted via telephone. I utilized semi-structured interview protocols (see Appendix B).

Teacher Focus Groups

During the initial visit to the school in April, I held one teacher focus group with six early childhood teachers. Conducting a focus group was an appropriate data collection strategy because it allowed me to learn about many teachers' feelings and opinions in a short period of time. The focus group allowed me to learn about many opinions simultaneously and having a group discussion presented some natural checks and balances in the conversations that provided more insight, at times, than did the individual interviews because they allowed me to gauge the degree to which certain opinions, feelings, or recollections were shared across the group and might be indicative of those experienced by the school community as a whole (Merriam, 2009; Patton, 2002). Because teachers often view teacher evaluation as unfair (Johnson, 1990; Mitchell & Purchase, 2015; Peterson, 2000; Sinnema & Robinson, 2007; Wise, et al, 1985; Wolf, 1972), constantly scanning the "temperature" of the group helped me to understand not only teacher views, but also their feelings about the practices at Elmwood School. Additionally, the focus group afforded participants the opportunity to talk to each other, which helped them elaborate upon important events and ideas that related to the teacher evaluation system at the school (Gall et al., 2010). The focus group was held in the principal's office (he was not present) and audio recorded using a semi-structured protocol (see Appendix B).

The purpose of the focus group was to understand teachers' general knowledge about, opinions of, experiences with, and perceptions of the evaluation system (), including how well the teachers believe the feedback relates to and influences their practice (Jiang et al., 2015; Trinter & Carlson-Jaquez, 2014; Thomson, 2014; Tuytens & Devos, 2010), and how teacher views impact implementation (Fullan, 2007). The focus group interview also provided an opportunity to triangulate data from the teacher and administrator interviews.

Virtual Document and Video Analysis.

In the interim between school visits, I viewed and analyzed data from the school's Google Drive folder for evaluation, to which the principal had granted access. The folder is updated in real time and includes many documents and videos related to classroom observation such as select classroom videos of formal observations, formal and informal feedback forms, videos of administrator-teacher post-conference sessions, teacher goal-setting forms, professional development calendars, and other support documents. One such document that is continually updated is a shared Google spreadsheet. Administrators use it to update weekly formative feedback based on 15-20 minute observations. Administrators input notes related to the Six Steps of Feedback, a structured way to give specific, immediately-useable feedback (Bambrick-Santoyo, 2012) which the school adopted in 2013. I was also given access to the results of a survey that all teachers fill out twice yearly. The Insight Survey (K12 Insight, 2016) provides aggregate data on teacher feelings about many school-wide issues including teacher evaluation and includes open-ended questions, to which many teachers commented on teacher evaluation practices. This provided another source of data for triangulation with interview data.

I used the Google Drive folder two ways. First, to understand school-wide policies and supports, I reviewed all whole-school documents, such as professional development schedules,

and principal calibration videos. Second, I viewed all 2015-2016 teacher-administrator feedback sessions and read all feedback documents for the five teachers selected for interviews. I used this data to ask about specific experiences during teacher interviews in June. I also used the data to think about my findings in that it gave me an idea of the types of feedback that school administrators provide teachers, both in relation to strengths and areas in need of improvement. I reviewed these documents on a weekly basis using a semi-structured checklist (see Appendix B) (Merriam, 2009). The checklist helped me to look for aspects of teacher evaluation that have been identified from my literature review as relevant to formative evaluation, such as the specificity and nature of feedback. The purpose of the document and video analysis was to understand Student Success's teacher evaluation structures and supports in more detail, helping me to answer my first research question and to triangulate data from the teacher and administrator interviews and focus group.

Teacher Interviews

I conducted teacher interviews with four pre-selected teachers during both school visits (April and June). I also interviewed one additional teacher in April and a second in June for a total of 6 teachers. I used the first session to get to know the teachers and get a general idea of their experiences with the teacher evaluation system. The introductory interviews lasted about 20 minutes each (a few pushed closer to 40 minutes) and were held in a conference room at Elmwood School. I used a semi-structured interview protocol (see Appendix B) and audio-recorded each session. I analyzed this data right away so that I could continue to follow emerging themes while looking through the Google Drive in the interim months. Before the final visit, I prepared an individualized protocol for each teacher (see Appendix B). These were more in-depth interviews and targeted specific findings from the Google Drive and initial

interview. Each lasted about 40 minutes and was held in the same conference room. Sessions were one-on-one and audio recorded.

Whereas the focus group was used to gain an understanding of general attitudes at the school and toward the teacher evaluation system broadly, the teacher interviews allowed me to gather more detail about each teacher's feelings and experiences regarding the teacher evaluation process and to understand how, if at all, teachers use the teacher evaluation process to inform their thinking, planning, and instruction (Grossman et al, 2010; Peterson, 2000; Russel, 2002). Interviews allowed me, as the researcher, to understand participants' lived experiences with the evaluation system in more detail than I could through other forms of data collection (Merriam, 2009), including particular experiences with the evaluation process in relation to their own classroom practices and particular findings that emerged through the video and document analysis.

Table 3
Alignment of Protocols with Research Questions

Research Questions		Protocols	Sampling Strategy	Time Frame		
				Initial Visit	Interim	Final Visit
RQ1 By studying a school with a strong commitment to formative evaluation, what can we learn about school-level features that support teacher use of evaluation feedback?		Administrator Interviews	Full Population	✓		
		Document and Video Analysis	Full Sample of Data Related to 5 Cases and Full Sample of School-Wide Data		✓	
		Principal Interview	Full Population			✓
	RQ2 In what ways do teachers use formative teacher evaluation feedback to inform their practice?	Teacher Focus Group	6 Teachers, Volunteer Sample	✓		
		Case Study Teacher Interviews	6 Teachers, Volunteer Sample	✓		✓

Human Subjects Protection

This study was approved for exempt status by the Rutgers University IRB on November 4, 2015. All participant data will remain confidential. All collected data has been stored in a secured location and all participant names and some details about the school setting have been modified to protect confidentiality.

Data Analysis Plan

Because my study is one of only a few that provides examples of recommendations from formative teacher evaluation research, I also used deductive analysis to look for alignment between the schools' practices and those recommendations. To do so, I used my literature

review as a framework for analysis of my transcripts (Creswell, 1998), listing out each recommendation for best practice in formative teacher evaluation. I then compared my data to each area of the literature review to understand if the school met, surpassed, or fell short of each. This provided a really great outline of my data which allowed me to look more deeply at other themes that began to emerge through an inductive process.

In inductive analysis, the researcher starts with small bits of data, and uses coding to put like pieces together while looking for larger themes (Merriam, 2009; Creswell, 1998). As the qualitative researcher discovers connections, she continuously verifies them against what she hears (or has heard) and sees (or has seen) to build further understanding (Merriam, 2009) in a cyclical process (Creswell, 1998). Over time, as the researcher probes the data, she will begin to find emerging themes that can explain the entire data set (Miles & Huberman, 1994). My research design was particularly suited to an inductive data analysis process as well because I used a second set of interviews to verify the themes that emerged initially and to use those themes to probe individuals' perspectives about their experiences of the teacher evaluation system.

In order to understand my data set, I began by transcribing all of my initial interview and focus group sessions. I transcribed the recordings myself because it allowed for greater familiarity with the data than if I had hired a transcriber and let me control the quality of transcription. I used open coding to analyze my data. To do so, I began by reading and rereading the transcripts to get a general sense of my data, continuing to do so until I felt well-acquainted with all of my participants and had an overall sense of my data. After that, I began coding. I used Dedoose, a data storage and analysis software tool to look for recurring concepts and assigned inductive codes to them (Seidel, 1998). I also applied deductive codes from the

literature on formative feedback, including codes about the nature of feedback itself and the ways in which the school applies or does not apply best practices in regards to helping teachers positively change their classroom practices through feedback. Unfortunately, at one point in the data analysis, I got blocked out of my Dedoose account. After many attempts to resolve my access issue with technical support, I elected to use Microsoft Word for coding purposes. I recoded my data set in a single word document. Using Word's navigation bar and headings, I created themes based on what I remembered from Dedoose and copied each excerpt into the appropriate file for each code. Though time consuming, this process actually gave me a much deeper familiarity with my data set. It also made it easier to check that all codes were mutually exclusive and that the codebook truly represented my entire data set. After creating the new word document, analysis of the text was quite simple. The themes and pasted excerpts created a good overview of the outcomes of my study. The down-side, however, was that I did not know how to change my unit of analysis or drill down in my data to look at theme by participant, etc.

By talking about my data set with colleagues, I began to form a preliminary conceptual framework (Miles & Huberman, 1994) that describes relationships between participants and school factors. At this point I was able to determine that the ways that teachers at the Elmwood School use feedback from the teacher evaluation process is extremely limited, which caused me to question the original intent of my research questions. I then began to narrate the details of my case study, piecing together my data in a wide-scale description (Creswell, 1998), using triangulated data (Baxter & Jack, 2008) and direct quotes (Cresswell, 1998) to support my analysis. By doing so, I was able to pick out new themes and ideas that I did not expect at the initiation of this study, which led me to revise my stance towards teacher evaluation. This is discussed in Chapter Five.

Researcher Role

As a researcher, I bring many inherent biases to the data collection and analysis processes. It is important to identify these in order to increase the trustworthiness of my findings (Merriam, 2009). At the time that I proposed this study, my personal experiences with teacher evaluation led me to view it as most productive when it is a formative practice that is linked to targeted professional development opportunities. Though my views changed dramatically over the course of this study, these beliefs framed the ways in which I constructed the literature review, methods, and protocols of this study.

My first experience with teacher evaluation was as a classroom teacher. During this experience, I was rarely given useable feedback about my classroom. Notably, after my first year of teaching, I was never invited to confer with my administrator about my teaching after I had been observed for evaluation purposes. When I returned to graduate school in 2006 for a supervisory certificate, I was impressed by new directions in teacher evaluation which were emerging, specifically creating minute-by-minute scripts of observed lessons and providing rubrics and narrative feedback, rather than checklists. As a former school principal, I found that teacher evaluation can be a powerful tool to increase teacher self-efficacy. In practice, I found that a lot of teachers were able to benefit widely. I recall the transformation of several teachers after I gave them the opportunity to highlight strengths and areas of improvement based on them receiving written scripts of their lessons. Because of these experiences, when I began my studies in the Ed.D. Program, teacher evaluation came to mind immediately as an area of study that I would like to explore further. In a 2014 pilot study conducted in a charter school in NJ, I found that the feedback from the teacher evaluation tool was not integrated into a wider teacher learning system, and, therefore, teacher learning from the tool was limited. Therefore, for this

study, I specifically tried to identify a school site that would meet the guidelines of a positively implemented formative evaluation system as defined by the review of literature.

My beliefs about the importance of formative teacher evaluation impacted my study in many ways. First, I sought to design a study and sought out a school case based on my inherent belief that this approach to teacher evaluation is beneficial to teachers and can contribute to improved education outcomes for students. As I interacted with my study participants, I was aware that I might color the interpretations of my participants by my own enthusiasm for formative teacher evaluation, so it was important to try to maintain a neutral stance in order to minimize bias (Merriam, 2009). This helped to encourage participants to fully share their perspectives and experiences regardless of whether their perspectives were similar to mine.

In this study, my role as a total outsider helped and hindered me in significant ways. First, because I did not know anyone at the school, it was hard to recruit volunteer participants. My attempt to send a recruitment email went unanswered. As a result, the principal chose participants. He assured me that he chose participants that represented the diversity of the teaching staff, but without knowing the whole staff, there I was unable to verify that. It may be that the principal knowingly or subconsciously chose participants who held more favorable views of evaluation or were more proficient at using evaluation data than the entire teacher population, though there is no way for me to know for sure. Conversely, as an outsider, I was able to gain more insight than a member of the school community because I was able to “see” the school in ways that insiders might have taken for granted (Merriam, 2009). And yet as an outsider, I also might have missed key details and subtleties because my eye was not trained to aspects of the school culture that may have been subtle. Further, study participants shared negative opinions freely, which may be an indication that they were not worried that their

thoughts would be shared with members of the school administration or that I would judge them negatively despite my own bias toward formative evaluation.

Trustworthiness

It is of utmost importance that I report my findings in trustworthy ways to both represent my participants' responses in an ethical manner and to draw accurate conclusions. In qualitative research, the researcher explicitly and systematically seeks to maximize the accuracy of her account in a process that begins before data is collected and continues even after she has written her conclusions (Creswell, 1998; Gall et al., 2010; Merriam, 2009; Patton, 2002). I used many strategies to increase trustworthiness including reflecting on my role and biases, using a purposeful sampling approach, utilizing detailed description, employing data triangulation, gathering verbatim accounts, and engaging in peer review.

Qualitative researchers often begin by reflecting upon their role in the research, its possible outcomes, and the ways in which their role might positively or adversely impact participants in the study. By thinking deeply about these issues, the researcher can develop a more neutral stance that encourages participants to disclose their feelings and opinions fully (Merriam, 2009), thereby leading to more representative descriptions of the case. Further, reflecting on her stance allows the researcher to become more aware of her biases so that she can make purposeful decisions about how to conduct and analyze data in as unbiased a way as possible (Miles & Huberman, 1994). Similarly, choosing an appropriate sample allows the researcher to more accurately represent the case. For my study, I tried to have maximum variation across my sample in order to understand a wide range of teacher experiences at Elmwood. I believe that this helped me to judge which opinions and experiences were shared

across the school's population and which were more related to particular participants' own experiences and dispositions.

During data collection, a qualitative researcher should detail what she observes without judgment by developing a disciplined subjectivity (McMillan & Schumacher, 1997). In order to develop a disciplined subjectivity, I tried to constantly self-monitor at each stage of the research process. This included checking my interpretations against my own biases and constantly reevaluating my thoughts and conclusions. Additionally, in order to triangulate my analysis, I collected data from multiple sources. Triangulation is the process of drawing on multiple sources as a way to verify accounts and strengthen the trustworthiness of findings (Bogdan & Biklen, 2006; Gall et al., 2010). For this study, participant views were triangulated across interview types and with findings from data and video analysis. Additionally, I used two forms of audio recording to document all interactions so that I could be confident that my final transcription represented the data accurately. Finally, I personally transcribed my data to have a deeper connection to the data. I also coded my data several times to eliminate possible bias and mistakes.

Even after collecting and analyzing data, qualitative researchers must continue to strive for trustworthiness as they move into reporting on their findings. In order to do so, I included verbatim accounts from my participants whenever possible to build my analysis. Using verbatim accounts allows for a rich description of the case and builds validity by accurately representing the participants' views (Creswell & Miller, 2000). In addition, as I reported my study's findings, I constantly checked my own beliefs against the data that I collected. I also used peer and participant review extensively to verify that my interpretation of the data was an accurate representation of what was recorded. I used member checking in the final principal interview to

verify that my initial findings were an accurate representation of what he believed to be true at Elmwood School. On three occasions in the interview, he affirmed that my description was accurate, even referring to its accurateness as “refreshing.” These instances came without prompting. For example, towards the middle of the interview he said “Once again, a lot of the questions that you're asking are all things that we saw in our Teacher Insight Survey. And those are things that we, as a leadership team, already sat down and discussed with the new incoming principal present, so she got the feel of where we ended off the year and could put levers in place to respond to those [issues.]” Peers also helped at many stages of my analysis and reporting, including reviewing my coding scheme, my assignment of codes to chunks of data, and my final draft. Peer feedback helped me to check my own biases and to determine if my analysis and conclusions were a good representation of the data. By keeping the trustworthiness of my findings at the forefront of my research design and procedures, I hope to have validly described and analyzed my case in order to accurately understand and report upon the implications for practice that the case illustrates.

Limitations

As is true for all case studies, my 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 and to researchers. 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 research, I am particularly concerned that my

predisposition towards formative supervisory practices might have skewed my participants' responses. I tried to minimize these biases by being cognizant of them throughout data collection and analysis and used this awareness to modify interview questions and reexamine my analysis between data collection sessions. Also, although I purposefully sampled a maximally varied participant pool, my sample size was small, representing about one-third of the school population and so it is possible that my data is not representative of all of the perspectives of the entire school community. Finally, my case study was based a far distance from my home which limited me in two ways. As a complete outsider of both the school and the social-political context in which the school is located, I was constrained in my understandings of my participants' shared history with the teacher evaluation system. I could not be immersed in the inner workings of the school, which limited my ability to observe teacher evaluation practices in real time and key relationships that impact teacher evaluation.

Chapter Review and Conclusion

The methods section describes how I used multiple data collection methods from many school stakeholders to conduct an instrumental case study (Stake, 1995). Through in-depth interviews, one focus group, and perusal of the school's actual evaluation feedback, I was able to deeply understand and describe the teacher evaluation system from the perspectives of the administrators and teachers at Elmwood Elementary and Middle School in Washington D.C.

The upcoming findings chapter explores how school-wide teacher evaluation practices are related to teacher learning at the school. It further describes how the practices around teacher evaluation and aspects of the school's culture are so linked that they form a symbiotic relationship. As an instrumental study, this case study contributes to the literature base on promising practices in using teacher evaluation as a formative supervisory practice.

CHAPTER FOUR: FINDINGS

As the research methods describe, Elmwood was chosen as a research site because of its atypical commitment to formative evaluation approaches. However, during the course of data collection, the context changed dramatically as the school lost some key personnel in the middle of the year. This change impacted the findings of the study significantly. Most importantly, there were a very limited number of findings that informed Research Question #2 which sought to uncover the ways in which teachers used evaluation data to inform classroom practice. Instead, most stakeholders talked about how much evaluation changed between years and the impacts of these changes on the evaluation system and the wider climate of the school. Thus, the findings led to important implications for sustaining fidelity in teacher evaluation systems. This chapter will describe these findings, while Chapter Five will provide an analysis of how the limited findings in relation to the second research question impacts this study's theoretical framework.

In this chapter, I begin with a description of the teacher evaluation system at Elmwood School including the original design and ways in which that design was modified in response to changes in personnel. Next, I will detail how Elmwood's evaluation practices aligned with recommendations from the literature for specific structures that support effective formative evaluation and define one new support structure that emerged from Elmwood's data. Then, I will describe lessons that emerge from Elmwood School's attempt to implement a formative teacher evaluation system. This section will include findings in regards to teacher experiences of the teacher evaluation system, impacts of implementation of the teacher evaluation system on the

school's climate, and sustainability of the teacher evaluation system over the course of several years and changes in staff.

Background: Elmwood School's Teacher Evaluation System

As described in Chapter Three, Elmwood Charter School is a Preschool- 8th Grade charter school in Washington D.C. that is part of a larger charter district, Student Success Schools. Elmwood School was one of the two original campuses; it opened in 1998, followed by twelve more campuses so far. Thus, at the time of data collection, the school had been operating for eighteen years. This description begins in the sixteenth year of its operation, the 2013-2014 school year, hereafter referred to as Year 1, when district leaders decided to adopt a formative teacher evaluation system. A few years previously, D.C. Public Schools had begun to implement teacher evaluation procedures encouraged by Race to the Top policy (Doherty & Jacobs, 2015). As a public charter, Student Success Schools were required to adopt a similar evaluation policy, but were allowed to choose some elements, including teacher evaluation tools and practices. The charter district was mandated to conduct three formal classroom observations for each teacher, each year, and to use a combination of scores from these observations and student achievement data to make final, summative, determinations about teacher practice. A few years previously, in 2010-2011, the district elected to create its own rubric for teacher evaluation, which central office administrators loosely based on Washington D.C.'s "Impact" Rubric. These elements became part of Elmwood School's "formal," or summative, teacher evaluation system. In Year 1, the district also elected to incorporate a comprehensive formative, or "informal," evaluation system that would be used as a supplement to provide weekly feedback to teachers that they could utilize to improve their classroom practices. After consulting building principals, the

district decided to utilize the Leverage Leadership framework for formative evaluation (Bambrick-Santoyo, 2012b).

Year 1: Initiating the Formative Teacher Evaluation System

As described, the first step in creating a teacher evaluation system at the school was to adopt a rating tool. The school chose to modify Washington D.C.'s "Impact" tool, which is itself a modification of the Danielson Framework for Teaching. The tool is primarily used to rate teachers during their three formal observations throughout the year. Ratings from the tool are then used to make summative decisions about teachers. One dimension of the tool is also used for formative feedback, as will be described below. The tool is set up as a rubric with four designations of teacher effectiveness for each area of practice, including exemplary (rating of 4), proficient (3), in progress (2), and unsatisfactory (1). There are also four dimensions, or "value drivers," of practice that the tool describes. Each of these is broken down into several standards. The district created a descriptive sentences about each standard that clearly communicates key values that the district holds for its teachers. For example, one standard is that "Student Success teachers maintain high standards of ethics and use good judgement in and beyond school."

The first value driver, "Excellent Teaching and Learning," often referred to as "ETL," describes practices that teachers should employ during daily instruction. Andre referred to this section of the rubric as central and reported that the administrative team uses it for the three formal observations as well as as a key resource for informal observations. Thus it is described in more depth than the other three drivers. This area communicates that teachers should create very structured, yet student-centered learning activities, and reinforces a student-centered nature by listing teacher descriptions of practice alongside evidence that students might exhibit. For example, the tool specifies that the "teacher [should] organize information that shows clear

knowledge of content and how to convey [content] to students” and that students will exhibit an ability to “explain what point in the lesson they are working on.” Overall, this section holds teachers to high-standards for instruction and is, at times, very prescriptive as to the ways in which teachers must enact instruction. For example, in addition to widely-understood descriptions of good teaching, like differentiating tasks, the tool includes district-specific definitions of good teaching that teachers are expected to adhere to, such as “regularly us[ing] a variety of “stretch it” questions at just the right points to maintain pacing.” This standard is unclear to me, as an outsider, but teachers reported having been trained extensively on the expectations of the tool.

The second value driver, “Outstanding Leadership,” describes teachers as active contributors to professional learning. This area of the tool relays a powerful investment in teacher learning through the generation of shared professional knowledge. For example, the tool calls on teachers to attend and lead professional learning communities (PLCs). However, as will be described below, teachers are not provided with time or access to undertake many of the standards in this section and, so this may be an area where teachers are held accountable for policies and practices beyond their control (Darling-Hammond, 2013; Glassman & Paulin, 1982; Kennedy, 2010b). The third, “Environment Conducive to Learning,” describes ways to keep students engaged in the classroom including things teachers can do within class time, such as to correct behavior, and outside of class time, such as to engage parents. Similar to the first value driver, some standards dictate a very specific approach to teaching. For example, exemplary teachers must “use appropriate words and tone, and narrate and correct 90% or more of the time.” Finally, “Organizational Strength and Viability” describes professional behaviors, such as timeliness, reliability, and ethical behavior. As is often the case with all rubrics, teachers

reported finding this area of the evaluation tool to be troublesome because the described behaviors are not easily observed. For example, Kathy described being unsure “how administrators know” about the nature of communication with stakeholders.

After deciding upon the evaluation tool and formative structure they would adopt, the district invested heavily in training for administrators so that they would understand how to observe teachers and give constructive feedback. To begin with, the district enrolled all administrators who would be primarily responsible for providing feedback to staff in Relay Graduate School of Education’s Leverage Leadership Institute in New York City that introduced them to the skills and processes necessary for conducting valid evaluation and giving formative feedback. The institute took place in the summer before Year 1. Andre, the principal, Victoria and Sean, the elementary and middle school assistant principals (respectively), and the math, English Language Arts (ELA), International Baccalaureate (IB), and Special Education (SPED) coaches from Elmwood School all attended (all names are pseudonyms). The district paid for all travel and course fees for every participant. This was the first of many examples that Elmwood administrators and teachers relayed of being financially supported to attend relevant training, including covering their travel expenses.

Interviewed administrators felt that these learning opportunities were instrumental to their practice. Andre, the principal, seemed particularly motivated by his experience at the institute, perhaps because he understood his learning to be well-aligned with Elmwood’s mission and needs. Andre gave many examples of changes and innovations he immediately instituted at Elmwood School upon his return. He also often described how he adapted suggestions he heard at the institute to fit the context of his school. One such program, administrative “learning walks,” is an opportunity for the administrative team to observe classrooms together to get a

“pulse” on teaching and learning practices at the school. As will be described in more detail below, Elmwood’s administrative team transformed these into a powerful administrative learning tool. To support further understanding of the formative evaluation process, the district purchased copies of the accompanying handbook for all school leaders (Bambrick-Santoyo, 2012b). Leo, one of the coaches at Elmwood School, described the book as more essential for his learning than other forms of professional development, and was grateful that his district had invested in it. He explained that the book was important to him because it is “where I really learned the skill [of] thinking about what would take [a] class to the next level [and] what would be the catalyst [for bettering instruction].”

After returning from the institute, the Elmwood School administrative team collaborated to develop an implementation plan. It is unclear from the data if all of the following were implemented in Year 1 or developed over Years 1 (2013-14) and 2 (2014-15), but many interview participants described them as fully implemented in Year 2. One important structure was a weekly observation calendar. The administrative team divided teachers so that each administrator would have six or seven teachers to whom they would be responsible for providing feedback. They further committed to an 80% administrative time allotment to formative teacher evaluation which included three observations (of at least 15 minutes each) and one feedback session (of at least 20 minutes) each week for each teacher. Andre described implementation of this schedule as uneven in Year 1, but reported that when administrators began to see the impact on teacher practice, they began to adhere more closely to the schedule. When preliminary data

showed a shift in teacher willingness to participate and student achievement outcomes, administrators bought-in more fully.

The feedback sessions were meant to be very structured. The team adopted the Six Steps of Feedback model which they had learned about at the institute (Bambrick-Santoyo, 2012b). The model leads the administrator through a step- by- step procedure for providing feedback in which the administrator helps the teacher choose an area of need, gives a practice opportunity to improve it, and helps the teacher put a plan for implementation in place, to be observed at the next classroom observation session (See Figure 1). The model is further structured so that teachers can ultimately take the lead in self-identifying areas of need. One benefit of the model is that it provides clear guidelines to leaders about how to provide feedback. However, one limitation is that it forces areas of need to be limited to “quick fixes” that can be easily practiced within a few minutes and observed immediately in a subsequent observation, as will be described in more detail below. Both interviewed administrators relayed initial difficulty implementing the model, especially completing so much in the twenty allotted minutes. However, as Leo described, when administrators became comfortable thinking through the Six Steps of Feedback as a way to look for catalysts to change, they could set up a pathway for teachers to incorporate complex instructional outcomes by focusing on small guideposts over time.

The school already had a process for teacher goal setting (sometimes called professional growth plans) in place. Around the middle of Year 1, administrators decided to align these with the teacher evaluation system. Thereafter, teachers were supposed to choose two goals from the formal evaluation rubric and their assigned administrator was to choose the third. These goals “were supposed to be goals that you worked on throughout the course of the year” (Angela). To align with the formative evaluation system (sometimes referred to as the informal evaluation

**SIX STEPS FOR EFFECTIVE FEEDBACK:
Leading Post-Observation Face-to-Face Meetings**

1 Praise	Praise—Narrative the positive:
	What to say: We set a goal last week of _____ and I noticed this week how [you met the goal] by [state concrete positive actions teacher took.]. What made you successful? How did it feel?
2 Probe	Probe—Start with a targeted question:
	What to say: What is the purpose of _____ [certain area of instruction]? What was your objective/goal for _____ [the activity, the lesson]?
3 ID Problem & Action Step	Progress to Concrete Action Step—Add scaffolding as needed:
	What to say: Level 1 (Teacher-driven)—Teacher self-identifies the problem: Yes. What, then, would be the best action step to address that problem? Level 2 (More support)—Ask scaffolded questions: How did your lesson try to meet this goal/objective? Level 3 (More leader guidance)—Present classroom data: Do you remember what happened in class when ____? [Teacher then IDs what happened] What did that do to the class/learning? Level 4—(Leader-driven; only when other levels fail) State the problem directly: [State what you observed and what action step will be needed to solve the problem.]
4 Practice	Practice—Role play/simulate how they could have improved current lesson:
	What to say: Let's try that. [Immediately jump into role play.] Let's re-play your lesson and try to apply this. I'm your student. I say/do _____. How do you respond? [Level 4: Model for the teacher, and then have them practice it.]
5 Plan Ahead	Plan Ahead—Design/revise upcoming lesson plans to implement this action:
	What to Say: Where would be a good place to implement this in your upcoming lessons? [Teacher works by self or follow these next steps]: Let's write out the steps into your [lesson plan, worksheet/activity, signage, etc.]
6 Set Timeline for Follow-up	Set Timeline for Follow-up:
	What to Say: When would be best to observe your implementation of this? Levels 3-4: I'll come in tomorrow and look for this technique. What to Do—Set Timeline for: Completed Materials: when teacher will complete revised lesson plan/materials. Leader Observation: when you'll observe the teacher (When valuable) Teacher Observes Master Teacher: when they'll observe master teacher implementing the action step (When valuable) Video: when you'll tape teacher or master teacher to debrief in upcoming meeting

Figure 1: The Six Steps of Feedback (adopted from Bambrick-Santoyo, 2012b)

system by participants), administrators were tasked with giving immediate “bite-size” targets or “levers” (Andre) that could “leverage” (Leo) the teacher’s growth, helping the teacher move toward mastery of the larger goal over time. The administrative team developed teacher binders,

which will be described later in more detail, to help teachers track their goals in relation to the feedback they received over time.

The school also already had many professional learning opportunities in place for teachers which the administrative team started to align with the formative evaluation process. Each week teacher teams met in “houses” that spanned grade level groupings including Preschool- PreK, K-2, 3-5, and 6-8. As the teacher evaluation system began to take hold at Elmwood School, teachers were encouraged to use those house meetings to discuss areas of concern the administrators had identified through observation, or to watch videos of teacher practice together and provide feedback to each other. Another important feature of the school’s schedule that was already in place was “professional development Thursdays.” Every Thursday, students leave early and teachers come together to discuss instruction or attend workshops. The administrative team began to debrief things they had observed each week on Mondays, and used those debriefing sessions to structure the Thursday PDs. As Leo described, during these meetings, administrators “go over...what our plans are for the week and what instructional components we have for each of our caseloads. Then, based on that, we identify a trend... We then develop PDs for Thursdays on that trend.”

The district had also already begun to create a career ladders for teachers. Data reveal that this career ladder was continually being refined, but one important structure that was already in place was a “Master Teacher” program. Master Teachers are chosen through a rigorous vetting process after at least 3 years of successful practice in the district. Master Teachers are given a lightened teaching load in exchange for presenting professional learning workshops, heading departments, and providing mentoring to other teachers. They are also rewarded with a substantial salary increase. In Year 2, the administrative team began to link the formative

teacher evaluation system to the master teacher program. They began to ask master teachers to record videos of their classroom instruction so that administrators could then use those videos as “models” that linked to areas that were observed as lacking in non-master teacher classrooms.

As the team was implementing the teacher evaluation system, they noticed different aspects of the system that could be improved to better support teachers, and began to think through ways to modify those practices. One big example of this was that the team was noticing some discrepancies between themselves in terms of the ways that they gave feedback to the same teacher or were rating the same observed lessons when the need arose, such as when hiring new teachers. In response, the team (or possibly just the principal) developed two structures. The first was an online spreadsheet to track the feedback that teachers received over time which was sometimes referred to as the FIOT or the Google Doc. The term “FIOT” was not used often and was never explained in writing or verbally, so I am unsure of what the letters denote. The tool helps all the administrators keep a similar focus when observing the same teacher. The tool and its uses will be described in more detail below. The second, which will also be discussed more thoroughly below, was that administrators began to use their weekly meetings to engage in classroom walk-throughs. They used this time to create consensus about how they were viewing and rating classroom instruction and to calibrate, or “norm” those ratings. Overall, during Year 1, administrators were learning how to use the evaluation system effectively and were becoming more adept at providing feedback. Throughout this time, they also developed procedures that could streamline the process.

Year 2: Investing Teachers in the Process

As administrators became more comfortable observing teachers, teachers began to recognize that the feedback that administrators provided had the potential to positively impact

their practice. For example, Pamela described how her assistant principal's constant presence was "supporting us and growing us in our craft." She described how, by Year 2, the assistant principal began to allow teachers to determine which areas of practice they needed feedback on, which she found helpful. In addition, administrators began to find other ways to help teachers take ownership in the process. For example, Pamela stated that Victoria, the Elementary AP, made everyone feel like they "were part of the change that [the elementary AP] wanted to see." Perceptions like these of the evaluation process and the help that administrators were providing to improve practice led teachers to invest more fully in the process. Teachers described teacher evaluation as comfortable and welcome in Year 2. They also confirmed that many of the structures that were set up in Year 1 were being implemented well.

The administrative team continued to incorporate responsive revisions to the evaluation process during Year 2 based on their own reflections on what needed improvement and feedback that they were getting from teacher surveys. For example, the team had been video recording many classroom observations to share with individual teachers during the Six Steps of Feedback post-observation sessions. The team realized that the videos could be used in other ways to support teacher learning. The principal described how the team is now in the process of cataloging these videos. He foresees that videos will be used as benchmarks of positive practice, linked directly to the teacher evaluation tool. The principal envisions that teachers will be able to immediately watch a model video that aligns with an area of need identified during evaluation, "so it should be easier to coach a teacher through a skill when they have something that they can just easily watch and [can] model [effective practice]." The school also began using a commercially created teacher survey, the "K12 Insight Survey," in Year 2. The survey is administered twice a year to measure teacher perceptions of the school's climate and culture

(K12 Insight, 2016). Because the tool asks specific questions about summative and formative aspects of the evaluation and feedback system, it has become a helpful way for administrators to assess their own practice. In all, the timeline of implementation in Year 2 shows that teacher evaluation practices were beginning to take hold in Elmwood School and that stakeholders were beginning to view teacher evaluation as a worthwhile system that could lead to improved practice. There were many positive outcomes for teachers during Year 2, which are discussed in more detail below.

Year 3: Sustaining through Turbulence

Data for this study was collected at the middle and end of Year 3. Teachers talked very positively about the teacher evaluation system in Year 2, but were less enthusiastic about implementation in Year 3. One reason for variations between the years was that there were many changes to the administrative staff throughout Year 3. At the beginning of Year 3, the school needed to hire two new assistant principals (APs) and a new math coach, as both the previous elementary and middle school APs were promoted to principal positions elsewhere in the district and the math coach moved out of state. The new middle school AP, Michael, had worked at Elmwood School for many years as an English teacher and department lead before being promoted to AP. The new elementary AP, Dana, and math coach, Joan, were new hires from outside of the district. The principal began training all three on the evaluation system only to find that in October, Michael, the middle school AP, was pulled out of the building to cover a principal position in another school. Andre described feeling frustrated because Michael left just as he was starting to understand the evaluation process and developing relationships with the teachers that would facilitate his ability to be effective. He observed that Michael was beginning to gain confidence with the system.

In October, [Michael was just] getting his feet wet with the normal observation cycle. Recognizing that in September you're building relationships with your teachers [and] building that foundation so that you can give them good feedback. And October is [when you start to] put it into action. And [that's when district administration] took him... [to replace a principal who had been] let go. (Principal)

As an interim solution, the team redistributed caseloads so that each took on a few extra teachers. In January, the new elementary AP, Dana, took on a principal position in another district. Neither of the two open positions was filled. Andre took on responsibility for conducting formal evaluations, which had previously been divided between himself and the two assistant principals, for his entire staff and left the informal coaching to the math, literacy, IB, and special education supervisors. However, the team was unable to provide the same frequency of observations and feedback as that to which teachers had become accustomed. In addition, Andre did not properly train Joan, the new math coach, on the elements of the teacher evaluation system.

As a result, teachers reported huge differences between Years 2 and 3. With the exception of the lowest-functioning teachers, no one was receiving consistent formative feedback. And, because only the principal could not complete formal observations for all 45 teachers, many teachers did not even receive the minimum of three formal observations. In April, some reported they had not had any formal observations (Lori, Early Childhood); some reported one (Allison, Angela, & Lisa, Early Childhood; Wendy, Elementary; Denise & Kathy, Middle School); only one teacher reported that she had had two formal evaluations (Olivia, Elementary). Of these teachers, four additionally reported that they did not receive any written

feedback after the observation (Allison, Lisa, Olivia). Extensions of the evaluation system, like tracking teacher progress towards goals, were also abandoned. The review of Google documents revealed that although all teachers created goals in the beginning of Year 3, only the middle school administrator had signed off on the goal sheets in September, and no teachers had revisited their goals during the check-in point in February. Teachers confirmed that they “never got the feedback for what the goal [would be] or any follow-up about what that was supposed to look like throughout the year” (Lisa).

In all, teachers recounted a shift in the climate around teacher evaluation in Year 3. For example, some teachers spoke about an overall feeling of the evaluation system being rushed, which contributed to feeling judged rather than supported. Teachers reported that administrators relied on very short observation sessions to assess their classroom practice and felt that this could lead to inaccurate conclusions about their practice. They could not understand “how [administrators] can tell” (Denise) anything because the “length of the observation didn't really allow [them] to make that type of critique” (Kathy). Teachers expressed frustration, including feeling as though “what people saw in a short span of time is not necessarily everything that was happening” (Heather). Teachers reported feeling like they had to “prove” (Kathy) or “explain” (Heather) themselves or reported disregarding suggestions from what they felt to be incomplete assessments of their practice.

One positive impact of the turbulence between years may be that it helped teachers appreciate the process that had been created in previous years more. For example, Heather describes how when the elementary AP, Victoria, was “pushing in to class times and planning meetings, teachers didn’t really respect that. They felt that she was micro-managing the situation.” This year, when the new elementary AP, Dana, would “just ask ‘are you planning? I

want to give feedback I want to see what you're doing,' but not necessarily be there for every step of the way” teachers felt like they were “not getting the support” they needed. As Heather’s comments demonstrate, it may be that future attempts to reinstitute a comprehensive feedback and evaluation systems may be better received by the staff because they had come to see the initial formative feedback program as supportive rather than intrusive in contrast to Year 3.

Year 4: Realigning the Teacher Evaluation System

Although data collection for this study was limited to Year 3, during the final principal interview, Andre spoke about many programs that the school planned to put in place during Year 4 in order to get formative elements of the teacher evaluation system back on track. These will be described in greater detail below and include differentiating support for high-performing teachers, implementing peer observation through the Master Teacher program, and incorporating more frequent teacher surveys. Figure 2 outlines a timeline of many of the initiatives introduced in Years 1-4. Some elements will be described in more detail in later sections.

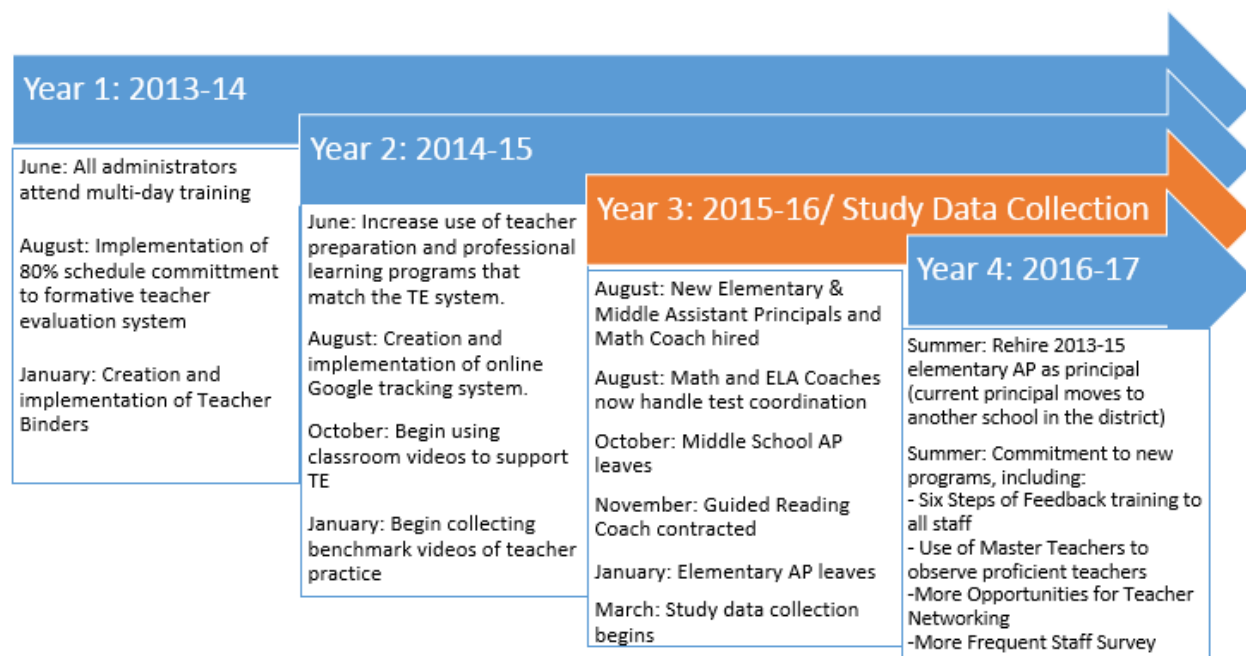


Figure 2: Timeline of Teacher Evaluation System Implementation

Teacher Evaluation Support Structures

Administrators at Elmwood School thoughtfully utilized existing and constructed many support structures for the teacher evaluation system. As described in the literature review (see Chapter Two), many researchers describe best practices that support formative approaches to teacher evaluation. Zepeda (2007) establishes that leaders that enact these practices engage in “formative supervision.” Best practices in formative supervision inform the analysis for the formative feedback process at Elmwood School because faithful implementation of these practices likely leads to better teacher learning outcomes. These include: (a) distributed evaluator caseloads, (b) evaluator training, (c) a focus on formative feedback, and (d) linked professional learning. In addition to practices that are aligned with research based practices on formative evaluation, data analysis also indicated that the school implemented additional practices that seemed to provide important supports to the teacher evaluation system including

(e) a continual growth-model and (f) integration of other programs or procedures that support the formative teacher evaluation system, such as streamlining teacher communication through the use of teacher binders. These supports developed and became more institutionalized especially in the first two years of implementation.

Distributed Evaluator Caseloads

Andre believes that leaders can make the biggest impact on instruction by observing teacher practice and providing immediate feedback. He is resolute that “Every instructional leader in the building [should] spend at least 80% of their building schedule inside the classroom” (Principal), and that administrative tasks, like meetings and paperwork, should be attended to during after-school hours. Though his vision of 80% time on observation does not seem to have been fully adopted by all leaders, Andre tries to institute it by structuring leaders’ responsibilities to fit this expectation. He begins each year by working with his team to create evaluator caseloads for administrators’ weekly observation schedule. To do this, the core instructional leadership team, made up of the principal, two assistant principals, the special education (SPED) coordinator, and the English, Language Arts (ELA), math, and International Baccalaureate (IB) coaches divides all the teachers among themselves so that each is responsible for providing consistent feedback to about six teachers. This was reported in Years 1, 2, and 3, however in Year 3, administrators had to redistribute caseloads when the middle school assistant principal left, then again when the elementary assistant principal left. This resulted in very large caseloads for the remaining administrators. In addition, instructional leaders have many other roles that involve working directly with the whole staff, and may observe teachers outside of their caseload; for example, the math coach, Joan, works with all math teachers in the building and provides them all with feedback about their mathematics instructional practices. Yet, while

all administrators might potentially give feedback to any teacher in the building at any time, only the designated administrator is expected to convey feedback on general instructional practices and to set instructional goals with her assigned teachers. In this way, each teacher only communicates about core practices with one key administrative point person.

To support the continual implementation of the 80% expectation, Andre helps leaders set up a weekly schedule so that each leader observes their core teachers at least 3 times a week (for 15-20 minutes each) and they meet with them at least once a week (for at least 20 minutes). Other administrative tasks, including duty schedules, departmental meetings, and administrative team meetings are then worked in around that core schedule. Andre continually communicates that work with teachers should be an untouchable aspect of the schedule. He stated, “My message to my team [is that] instruction, observation, and feedback is priority, so if there are lots of meetings taking place... being in classrooms takes priority.”

Evaluator Training

In addition to a weighty time requirement, Andre holds his instructional leaders to a high standard of quality for observation and feedback. He helps his team meet these requirements through extensive training and support. This includes providing initial training inside and outside of the school, professional resources, and sustained learning opportunities throughout the year. These practices are so robust that the school surpassed what is recommended for administrative training in the literature, especially in the initial years of implementation.

As described previously, observer training began with a summer intensive for all administrators and the provision of an accompanying professional resource guide in Year 1. In the beginning of each year, the principal also reviews and reinforces a key component of the training, the Six Steps of Feedback (Bambrick-Santoyo, 2012b). This yearly refresher

professional development is also a time for the leadership team to learn about or review practical skills like using video, keeping useable observation notes, deciding upon the best feedback within the Six Steps framework, and documenting teacher progress. According to the principal, this review was extensive in the beginning of Years 1 and 2, but was not developed enough in Year 3 as the principal incorrectly assumed that weekly meetings (described in the following paragraph) would be enough of a professional learning opportunity for administrators to fully understand the school's observation system.

Another key piece of observer training is that learning is sustained throughout the year as the administrative team meets each week for observation calibration, which the team refers to as "norming." During this time, the team discusses evaluation data and feedback data to make sure that they all can agree on scores on the rubric and can build consensus around ways to develop teacher practice. On some weeks, the whole team conducts a walk through in a randomly selected classroom and talks about the instructional strengths and areas of need. In these instances, Andre explains:

We all go in together. We all observe together, and then we come back [to my office] and we debrief. So we talk about what we saw and what we think would be the next priority for that teacher. And then we discuss if there are differences in opinion, we look at those and then the person who's responsible for giving them their feedback normally goes back and delivers that feedback and continues the work with them.

As Andre described, the weekly walk-throughs, often called "learning walks" in other schools, enable administrators to get a pulse on instruction and to possibly generate richer, feedback for teachers. However, Elmwood's use of learning-walks goes beyond simply seeing and commenting upon instruction. Administrators also use this time to develop shared

interpretations of instruction and gain consensus on the ways in which they should give feedback. This process is further enhanced as the team watches video-taped post-observation conference sessions together on other weeks. In each video, an administrator is giving feedback to a teacher. Together, the team reflects on the video and discusses ways to make feedback more effective. Much like teacher video groups that help teachers notice student reasoning (Sherin & van Es, 2005), these sessions allow administrators to focus on the cognitive process of the learners (in this case, teachers). This in turn can support better understanding the process of helping teachers develop self-reflective skills and mindsets and improve practice. With these procedures in place, training goes beyond teaching observers to collect unbiased data and pushes them to develop a formative stance toward “instructional” leadership. In addition, norming helps administrators understand “what the expectations are, what they're supposed to do, what it is supposed to look like, and how often we have to do it” (Leo). The principal believes that Elmwood School’s model of observation and feedback is not hard to master. Instead, he proposes that “learning and getting better at it comes with the practice, and the coming back to the table as a team and norming and pushing your practice.” As a result, he makes the norming process a priority and works with administrators to continually refine it each year. However, as described, he relied heavily on this norming process in Year 3 to bring new staff up to speed without providing extensive initial training. This led to some gaps in implementation, as will be described below.

Finally, Andre constantly communicates that “observation and feedback is [a] priority” and offers administrators his support in blocking out time for observations and talking through the process of evaluation. According to both interviewed administrators, the outcomes of detailed and sustained instructional leadership training have been more confidence and more

consistency across the administrative team as leaders are given continual practice opportunities and more buy-in as leaders repeatedly debrief observed outcomes of the process. Additionally, the norming process allows for a level of consistency and objectivity in that the potential for teacher ratings to be based on the personality, style, or priorities of one individual is decreased. Because of extensive time devoted to learning-walks throughout the year, all administrators have a shared understanding of the target teaching and learning practices they are trying to encourage and this shared understanding can also be shared with teachers during feedback conversations.

As the literature review described, teachers should also have opportunities to learn about the evaluation process in order to utilize feedback successfully (Heyde, 2013; Holland, 2005). The school did devote this time to the implementation of the teacher evaluation system. The school spent “a lot of time at the start of the school year just going over the different components [of the teacher evaluation rubric]” (Kathy) so that teachers were “completely clear, like crystal clear, on how they would be evaluated as professionals” (Andre). Surveys reflect that 91% of teachers who completed Elmwood School’s 2016 Fall culture and climate survey felt that they “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. Though administrative training on the formative system has been extensive, substantive, and very structured, teacher training has not been as comprehensive. For example, teachers reported that they do not spend much time talking about the rubric later in the year. Also, although training in the beginning of the year helps teachers to understand the components of their evaluation, information on how they would be observed and given feedback was not provided. As a result, teachers were unaware of many components of the evaluation process; 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. As identified in the literature review, successful implementation of formative teacher evaluation systems requires that schools prepare teachers to utilize teacher evaluation feedback by providing training on reading and analyzing evaluation data and on how to reflect on instructional practices (citation). In all, though the school provides more hours and higher quality training for administrators than that which is minimally recommended in the literature, teachers training to prepare them to engage in meaningful formative conversations was lacking (Jerald, 2012; Murray, 2014; Sartain et al., 2011; Shaha et al, 2015).

Building a Growth-Mindset

The observation schedule and administrative training set an expectation for frequent feedback sessions, but it is the school’s commitment to helping teachers “grow their practice” (Andre) that opens an opportunity for teachers to use evaluation data to continually improve their practice. As described in the literature review, focusing on formative feedback that can drive growth is a difficult process that requires that teachers and administrators view practice as changeable and that administrators create learning opportunities for teachers that are supported by high quality resources aligned with change initiatives (see: for example: Shah et al., 2015). It is not a process that is well utilized in most schools, but Elmwood administrators have worked hard to fit formative evaluation to the culture of their school. The difference between Elmwood and other schools is apparent from new teachers’ feedback about evaluation. As Andre described, new hires at Elmwood come to the school with previous experiences of observation as an event that induces fear of judgment and reproach. This fear is heightened in new teachers’ initial experiences at Elmwood due to the sheer number of observations that are conducted. As

such, the administrative team spends time “shifting the mindset” of new hires “into the space of learning as opposed to a space of performance.”

Many teachers described getting immediate feedback that helped them to implement changes to their practice, especially during Year 2. This was done both in writing and in person. One teacher reported that she got notes within five minutes of an observation and would meet with her coach later on that day or the following day to discuss everything. Teachers confirmed that in Year 2 frequent formative check-ins occurred multiple times throughout the week. Wendy, an elementary teacher, explained that if she needs feedback from an administrator on a lesson that was not already part of the weekly schedule, she would “just reach out to [the administrators] and they can come in and observe us.”

Feedback also helped teachers work towards the three instructional improvement goals that they chose with their designated administrator at the beginning of the year, which is another process that is described as best practice in the literature (Chappuis & Chappuis, 2007; Cogshall 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). Many interviewed teachers confirmed that they felt the informal evaluation system supported them to work on their goals until they were met, especially during Year 2. As the principal reported, administrators “just work through [the goals] with [the teachers] until we just get [the teachers] where we need them to be.” Having and working toward instructional improvement goals in this incremental fashion helps teachers develop mastery over time (Ericsson, 2009; Shute, 2008; Westerberg, 2013) because “the feedback [each week] is just bite-sized and the

expectation is that [teachers are] going to get to the end point in a gradual way” (Leo). By setting small, immediately attainable goals, administrators could create a constant movement toward improvement. Rather than overwhelming teachers with multiple or multi-layered changes they needed to implement, administrators focused on what would “be the best leverage of that teacher’s time in that moment,” while mentally mapping out how bigger goals, more difficult tasks, and more nuanced learning would “tie into” that initial piece of the goal (Leo). In addition to incremental instructional improvement for teachers, this goal-orientation contributes to a more positive school climate because “it’s a very genuine way to build practice and build better professionals without playing the blame game or coming across as antagonistic” (Andre). Teachers confirm that the goal-oriented culture is positive for them. As Wendy explained, “It’s a different dynamic. I like just the whole goal-setting and everything.”

Administrators also encourage teachers’ growth mindset by regularly asking teachers to reflect on their practice (Anast-May et al., 2011; Chappuis & Chappuis, 2007; Costa & Garmston, 2002; Danielson, 2001; Marzano et al., 2012; Ovando, 2005; Range et al., 2013). For example, Emily explained that after each formal (summative) observation, administrators “send you your rubric in advance and you rate yourself on each category...and then [when you have your post-observation conference]...you compare notes...and that’s [how] the discussion [is structured].” She further described that for informal (formative) conferences, the administrators “typically asks you first for you to be reflective and talk about how you thought the lesson went and rate yourself on the rubric.” Heather similarly described that she is encouraged to reflect on her practice during formal and informal meetings. She describes being first asked “ ‘what do you think worked in the lesson?’ and ‘what could you have changed?’ ” during these sessions. Recently, the school also invested in a video technology that teachers can use on their own to

record classroom videos. The administrators hoped this would encourage teacher use of classroom video to support reflection. For example, Lisa explained how Victoria, the elementary AP, used video to encourage reflection and dialogue. She said, “[the elementary AP] 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 what you had been doing.” Teachers explained that video recording had become such a normal extension of the observation process that they began to view videos of themselves teaching like a resource. For example, Janelle said that she watched video of herself teaching “to see was I using good questions and [if] the class was engaged when I was asking these questions.” As all of these experiences illustrate, teacher self-reflection in relation to the evaluation process was encouraged through many supportive administrative structures.

Linked Professional Learning

In addition to identifying specific areas of improvement, administrators often provide resources that relate to observation feedback and can help them reach their goals, including linked professional learning opportunities. Linking professional learning to teacher evaluation data is another important process that the literature presents as best practice (Archibald et al., 2011; Coggshall et al., 2012; Danielson & McGreal, 2000; Darling-Hammond, 2012& 2013; Darling-Hammond, Amerein-Beardsley, & Haertel, 2012a; Hill & Grossman, 2013; Holland, 2005; Looney, 2013; Marshall, 2013; Smylie, 2016; Zepeda, 2007). These can be informal to highly-structured learning opportunities. Informally, for example, administrators send emails with links to certain articles or resources to support teacher learning and needs. They also gave teachers the opportunity to observe other teachers which allowed them to “see those teachers in

action and... apply [learned strategies] to my classroom” (Lori). Additionally, teachers are sometimes sent to more formal professional development sessions. When relevant, administrators were willing to invest large sums of money for professional trainings and travel expenses. For example, Emily described attending a professional learning conference in California that the school financed.

The literature also recommends that schools link professional learning opportunities to teacher evaluation feedback in order to maximize the educative potential for teachers (Archibald et al., 2011; Cogshall et al., 2012; Danielson & McGreal, 2000; Darling-Hammond, 2012 & 2013; Darling-Hammond, Amerein-Beardsley, & Haertel, 2012a; Hill & Grossman, 2013; Holland, 2005; Looney, 2013; Marshall, 2013; Smylie, 2016; Zepeda, 2007). Elmwood School linked professional learning to the teacher evaluation system at the individual level, through goal setting and individualized feedback, the department level, by using observation data to inform department meeting agendas, and at the school-wide level through its weekly Professional Development Thursday sessions. This process begins in August when administrators work with teachers to collaboratively choose three individual teaching goals for the year. Heather describes this as a very systematic process:

We had the [evaluation] rubric, and we would pick goals [from there], almost as if for semesters. You pick three goals you're going to want to work on for the year. But then, the first nine weeks of school... [you] focus on [one], and [later] we... get to the rest of them... And then we would meet weekly [to discuss]. [The elementary AP would ask] ‘How are you doing on this one? Do you feel like you have it or how can we assist with that?’

Teachers described that in Year 2 they would work towards each goal until the administrator felt they had reached a level of mastery in that particular area of practice before moving on to another appropriate goal. Teachers also have the opportunity to provide direct input on the focus of PD. They respond to needs assessments throughout the year and are able to request learning opportunities. They also meet weekly in their houses, during which time they have an opportunity for extra and specific professional learning. For example, Angela described that as a house leader, she would bring information back to the assistant principal about areas that teachers in her house wanted to work on. Then the assistant principal “would find a PD [from a vendor or expert outside of the school] to support it or she would conduct a PLC [in-house] that would support” it. Being focused on individualized goals based on teacher evaluation data and teachers’ felt needs increases the likelihood that professional development will be relevant, an important prerequisite for positive teacher learning (Almy, 2011; Anast-May, et al., 2011; Contreras, 1999; Curtis & Weiner, 2012).

Finally, at the department and staff level, administrators use multiple data sources to decide upon professional development offerings that are responsive to teacher needs based on their goals and observation data. The administrative team used this data to make decisions about the types of professional development that they would provide on-site and outside of the school. For example, administrators use observation data to decide on school-wide workshops that they provide during Professional Development Thursdays or to contract with a professional development vendor to provide school-wide support in a certain area. They also send select teachers to professional workshops outside of the school site. This was usually reserved for either high-performing teachers who were expected to attend the workshop, then turn-key their knowledge or for new or struggling teachers who presented a significant need that could not be

addressed at the school but that administrators felt could be remedied by attending a workshop. These findings indicate that the administrative team spent time developing and sustaining professional development opportunities linked directly to observation data and teacher input. Unfortunately, though supports for linked professional learning were in place at the beginning of Year 3, shifts in administration led to fewer PD offerings linked to teacher needs, and there was a complete cessation of PD offerings around February when everyone's focus shifted to preparation for standardized testing. Figure 3 displays some of the key professional learning programs that support the teacher evaluation system at Elmwood School.

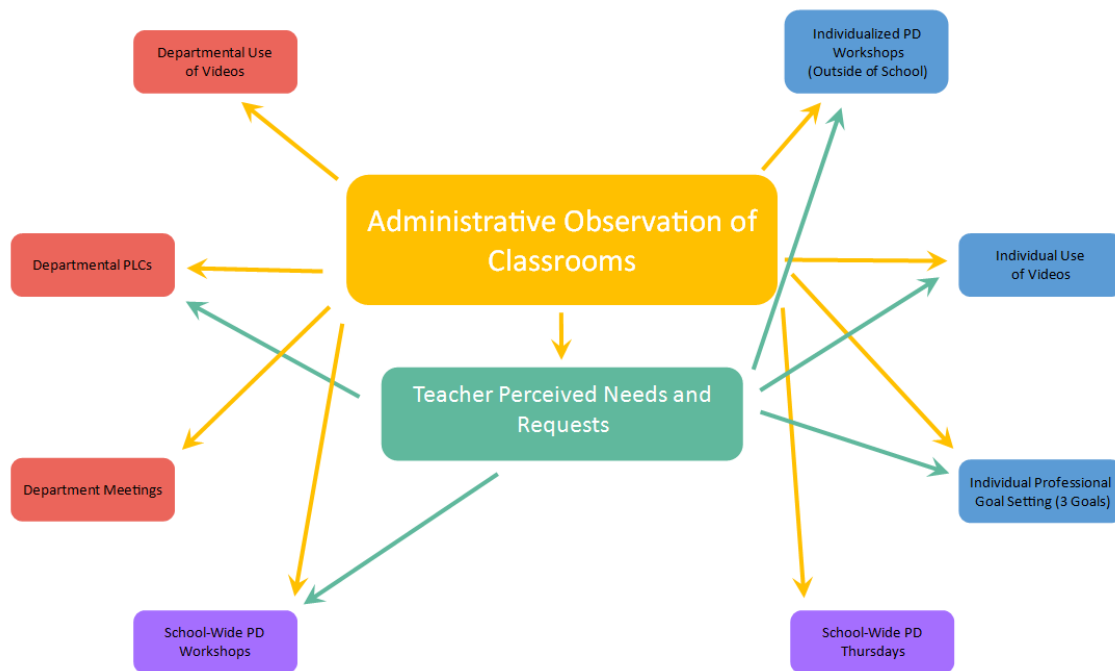


Figure 3: Elmwood School Professional Learning Sources

Inductive Findings: Sustaining the Teacher Evaluation System through Programming

The previous sections describe findings that are aligned with research on the five key structures that support formative teacher evaluation systems. In addition, two related findings emerge from Elmwood's system that may be relevant for effective implementation of formative teacher evaluation at other schools. First, administration is committed to continually improving its practice. As a result, administrators use knowledge gained through this process to create ongoing program improvement by integrating new practices into the framework of the formative evaluation system to increase its potential to support teacher growth.

A key element of the Elmwood School that contributed to what seemed to be effective implementation of the teacher evaluation system is that it is responsive to feedback from stakeholders within the system. This played out in two important ways. First, teachers reported that administrators were open to their ideas and critiques. They overwhelmingly reported that they felt they could voice concerns regarding ineffective evaluation practices and that changes would eventually be made as a result. For example, Emily, a master teacher, reported that one positive aspect of the school's culture is that "[administrators] are pretty open" to recommendations and ideas of new programs to utilize. She discussed how she had been encouraged to record videos of her instruction for the master teacher program and felt that it was a powerful way for her to reflect on her practice, but was underused with other teachers. She hoped that if she pointed this out, the administrators would use this practice in more classrooms. In fact, she gave some specific examples of how administration had already begun to use her suggestion including purchasing a classroom video recording tool and creating a protocol (sign-out sheet) for its use. In relation to the formal teacher evaluation system, other teachers relayed that they viewed their ratings as fair because administrators were willing to reevaluate them

when they felt that ratings were not an accurate reflection of their best efforts. In all, teachers at the school felt comfortable enough, not only to open a dialogue around their own ratings, but to offer suggestions and critiques that would ultimately better the whole system.

Furthermore, administrators sought out opportunities to understand teacher perspectives as a way to improve the process. As outlined previously, the school administers a commercially-created school climate and culture survey to all staff twice a year. The administrative team uses this survey as a means of reflection on their own evaluation practices. As a result of the May administration of the survey in Year 3, the principal learned that the elementary assistant principal, Dana, had not implemented formative teacher evaluation structures as well as she could have. Results of the climate survey showed that teachers were getting less support compared to previous years. Dana was a new administrator; she began her new position late in the summer and could not attend the New York training. It seems that she was not given enough training to be able to understand how to implement evaluation feedback. Andre believed that this problem was the result of unclear communication on his part. As he recalled, “I think that what her perspective of a check-in or feedback conversation with a professional was is not what they expected and not what they got before... Meeting with the new AP was much more of like a conversation as opposed to a meeting with action steps... so they walked out of it with a very loose perception of what they were supposed to do next.” Though he recalls checking weekly with both APs, without other sources of information, the principal incorrectly assumed that teachers would view the AP’s feedback positively and, she was surprised to learn that teachers felt differently in May. He reported that “the information that I was getting from the elementary AP was completely different than the feedback that I got from the teachers at the end of the

year.” Though Andre was dealing with strong feelings of guilt and frustration about the situation, he was determined to use this experience as a growth opportunity. He concluded:

I want to make sure that that doesn't happen at Elmwood again, and also that it doesn't happen under my watch anywhere else. I want to make sure that our systems and procedures and everything is still going like it needs to go even though the people [change]. And so that's definitely something that I'm really thinking about to make sure that we don't do it again.

He named specific action steps to avoid a similar situation in the future, including that he would conduct “more frequent surveys, ... hold administrators more accountable in the moment, and ... [encourage] the teachers [to] hold [him] accountable for holding administrators accountable.”

This quote is a powerful example of Andre’s commitment to both the teacher evaluation system and to ensuring the positive experiences of his teachers. It is also evidence of his scrutiny of his own practice as a leader which undeniably led to a continuous system of improvement.

While openness to feedback from stakeholders is an important prerequisite to ongoing improvement, the change process is more likely to be implemented and sustained with the addition of programs and procedures that formalize it (Scott & Davis, 2007). This is evident at Elmwood School because, in addition to being reflective about their successes and shortcomings, the administrative team was able to drive change by actually incorporating new understandings to improve programs that could better support the teacher evaluation system. As the principal described, these changes often led to a more streamlined process. He stated, “You know we think of different things everyday to make the process even easier than it was before.” Many of the changes were seemingly minor modifications, but substantially impacted the evaluation experience for teachers. For example, administrators aligned Washington D.C.’s required

CLASS evaluation for early learners with the domains of the school's evaluation tool so that teachers could feel confident when representatives from Washington D.C.'s Early Childhood Evaluation Team visited Elmwood School to assess the early childhood program twice a year. Another example was that through the evaluation process, administrators became aware that they needed to provide better learning opportunities for high-performing teachers so they sought out appropriate professional development programs for them, such as the National Academy for Advanced Teaching Program. Administrators also recognized that one teacher preparation program in particular, Urban Teachers, was more aligned with the evaluation practices they were trying to foster at Elmwood School, and so they increased the number of novice teachers that they hired from that program. Many teachers also appreciated that administrators provided coverage so that they could visit other teachers' classrooms to observe effective practice related to their evaluation feedback.

The school also integrated improvements that were more complex and difficult to implement. The data indicate that the school implemented three such programs that were linked directly to the teacher evaluation system. The first is the adoption of a tracking system that all administrators use to input teachers' progress toward reaching their goals over time. The tracking system was created a few years ago when administrators noticed that teachers felt overwhelmed by the sometimes conflicting feedback they were receiving. Andre explained its evolution:

For [the teachers] 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

[teachers] was very fluid across [different observers]. 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 you saw, 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 a spot-check and to know [what] they're focusing on right now ... [It] just gives 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 [members of the leadership team] are 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 they know we're talking ... and so when I'm giving you feedback I'm only giving you feedback around that... We found that was a lot more fruitful ... It's just a lot more planned and structured.

The system consists of excel sheets that are updated in real-time on Google Drive (Figure 4). The format is uncomplicated, and the language plain. Yet the tool has allowed for efficient communication between administrators and has supported Elmwood School's approach to incremental teacher improvement by providing a platform through which teacher learning can be charted over time. The tracking system has become so useful at the school that its implementation was made mandatory for the entire district and now streamlines communication between central office administrators and school-level leaders. Andre described that when central administrators "come in to observe [teachers in our building] they look at the Google Doc to know what we're working on with [each] teacher. So it makes everything more fluid on the

district level as well.” It was also adopted by a national principal preparation program for use by all of its upcoming principals.

Teacher A									
Current PD/Instructional Goals					Previously Met Goals				
Chap.4: Planning for Success (Technique 16) Beginning with the end in mind (planning)									
O=Observation; V= Video; P=Planning mtg or lesson plan review; D=data talk (benchmark, biweekly assessments, exit tickets, etc.)									
Date	Time/Class	ENGAGE 1	ENGAGE 2 (Optional)	Praise	1-3 Key Levers (small, measurable, targeted)	Evidence of Change from Previous Observation	Date of In-Person Coaching Meeting	Type of Coaching Meeting	Observation Notes (Optional)
8/25/2015	10:15 AM	ECL2		Ms. Teacher A has great relationships with his scholars and takes no time to build solid group relationships	Plan critical routines and procedures moment-by-moment		8/31/2015	O	
8/26/2015	1:05 PM	ECL2		Ms. Teacher A has great relationships with his scholars and takes no time to build solid group relationships	Plan critical routines and procedures moment-by-moment	Routines and procedures teacher document	8/31/2015	O	Ms. Teacher A submitted her routines and procedures document and received feedback about the explicitness of the routines and
9/1/2015	8:25 AM			Ms. Teacher A is	Plan with the end		9/1/2015	P	Ms. Teacher A will

Figure 4: Google Tracking Spreadsheet

Though originally constructed as a way to streamline communication and help teachers focus on one area of improvement at a time, the tool also helps the principal to “track and hold every [administrator] accountable for observing three times a week” (Andre). In all, the online system was created to keep all administrators accountable to a shared investment in teacher learning and allow all administrators to communicate consistently with each other. Teachers also benefit because they receive feedback that relates directly to their current area of focus and can help them heighten their practice without being overwhelmed by multiple goals at the same time.

Just as the Google Doc spreadsheet helps increase communication, consistency, and accountability for administrators, teachers are similarly held accountable and supported through

“teacher binders.” The binders were created as a way for teachers to keep track of all the programs that the school adopted to support the teacher evaluation system in one place. These include their professional growth records from both the formal teacher evaluation system and the informal formative feedback system and other related learning systems documents like peer observation and analysis of student achievement data. Each teacher receives a one- inch binder and a pack of dividers. Teachers divide the binder into five sections for weekly meeting notes, the teacher’s instructional action plan, peer observation forms, faculty meeting notes, and student data analysis (see Figure 5).

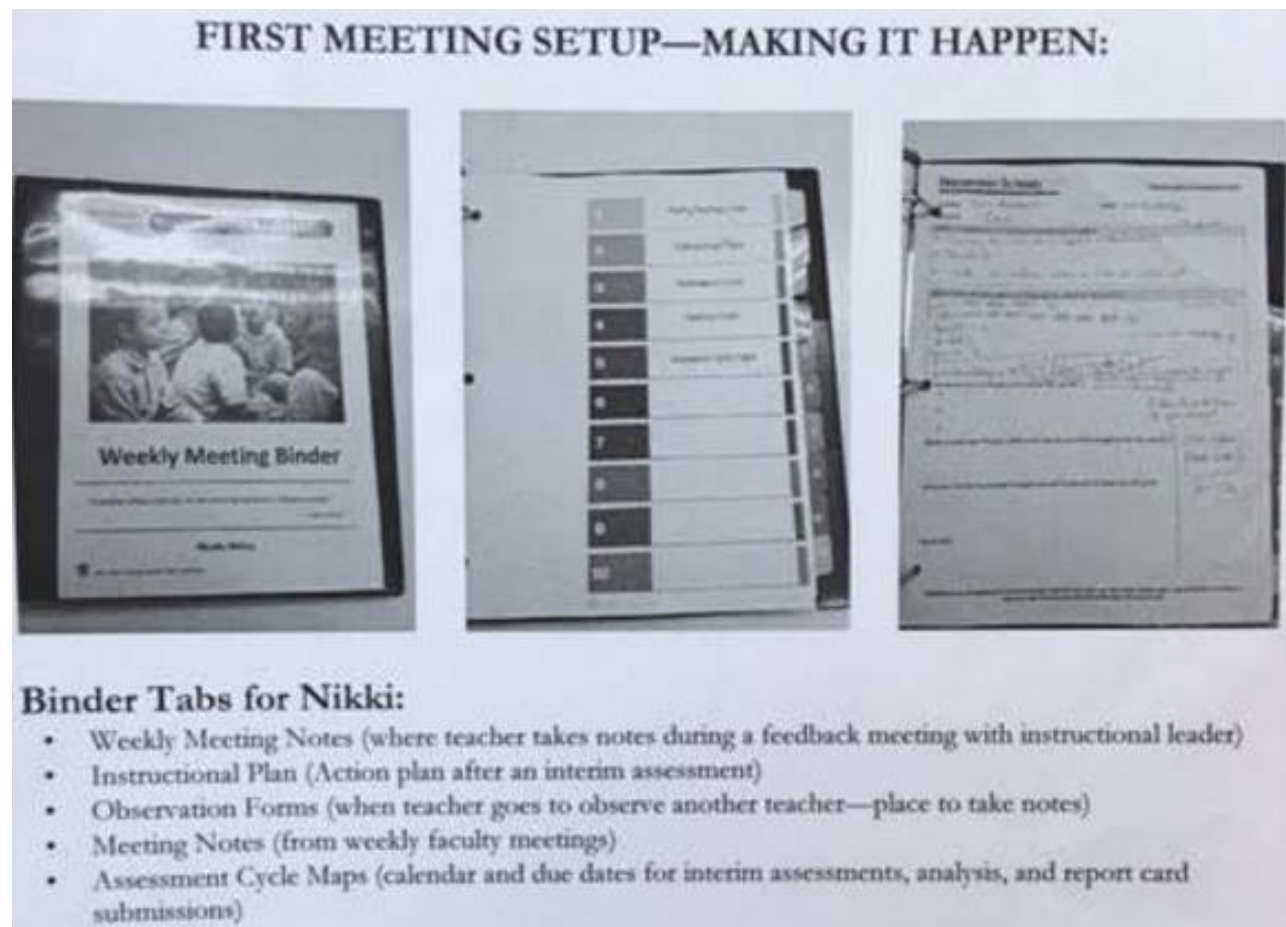


Figure 5: Teacher Binder

Teachers are expected to use the binders to track progress toward meeting their goals, with the support of administration. Heather reported that her administrator distributed the binders in the beginning of the year and that teachers kept copies of the teacher evaluation rubric and their feedback in the binders. She also explained that after each of the three formal observations, the teacher and administrator would look at the teacher's three goals, measure progress towards reaching those goals, and then discuss next steps and how the administrator could support progress. Heather also used the binder after informal observations to track her own feedback. Other teachers explained that they were expected to add classroom artifacts that showed growth towards the goal, like pictures or student work, to the binders. When assistant principals consistently asked teachers to bring them to conferences and used them as a point of reference to talk about teacher growth, the binders encouraged formative, reflective dialogue related to evaluation. Unfortunately, the binders are another initiative that was not sustained in Year 3 as administrators came and went. The two new assistant principals did not require that teachers reference the binders in the same ways that previous administrators had; their relevance was diminished. Andre explained they became less important in Year 3 because "if your AP is not requiring you to bring it, it's not going to be used with fidelity."

Finally, the administrative team used data from the teacher evaluation system to select and implement a guided reading program that was responsive to teacher learning needs. Leo detailed the meeting when the administrative team decided to choose guided reading as an area of focus for PD:

Monday, 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. So that's how we selected guided reading to be an

instructional focus. We saw that was the trend that needed attention and so we decided to make that a priority.

Ultimately, the team decided that this learning need was so wide-spread that they hired a consultant to support teachers. Subsequently, every teacher who worked with the guided reading coach described her support as helpful. For example, Olivia recounted that the coach not only helped her to more effectively understand her practice through guided questions, but gave immediate resources to support her growth. She described her feedback as “great” because it helped her reflect on her practice, was always connected to immediately useable resources, and helped her to map out what needs would be important to attend to next. The teachers indicated that the coach is unquestionably an adept, resourceful mentor. However, it is likely that school leadership enhanced her capacity to be effective by providing group professional development time, scheduling time for her to access individual teacher classrooms and to work with teachers during their individual planning periods, clearly explaining the guided reading coach’s role to teachers, and encouraging teachers to utilize the coach. For example, Wendy describes how PD Thursdays were dedicated to work with the coach:

We used to have half-day Thursdays where we would have different types of professional development. This year, [the administrative team has] been pushing guided reading. We've been doing guided reading the past few years, but they've been really pushing it this year. So they had a consultant come in, and the Thursdays would be her days. She would come and give feedback all week, but then on Thursday, if there was something she wanted to talk about, or if she wanted to do different professional development, she would help us to really get a stronghold on guided reading.

As both quotes reveal, the administrative team clearly linked the coaching program to the teacher evaluation system and appropriately allocated time and resources to the success of the program.

Lessons Learned from Implementation of the Elmwood School Teacher Evaluation System

As previously described, the school established many programs and procedures that support administrators' ability to implement formative evaluation efficiently. The intent of programming was to heighten teacher instructional effectiveness in order to increase student achievement. Although Years 1 and 2 showed a lot of positive motion forward regarding the implementation of a formative evaluation process, when personnel turbulence in Year 3 threatened this, it caused a shift in implementation efficacy which was coupled with a weakened school climate. As described earlier, during Year 1, administrators had focused on practicing how to collect observation data and providing useful feedback, and by Year 2 stakeholders reported positive impacts of the program. By that point, leaders had "bought-in fully" (Andre) and teachers described feeling well-supported to make positive changes in their classrooms. Though, at times, teachers reported feeling "over-burdened" (Pamela) by the plethora of feedback, the general feeling was that they were being supported and their practice was being nurtured. However, during Year 3, the school was unable to fully sustain the evaluation system as new leaders came and left. As will be discussed below, the implementation (or lack thereof in Year 3) of these structures impacted stakeholders at Elmwood School in many ways. This section describes teacher experiences of the teacher evaluation system, how implementation of the teacher evaluation system contributed to shaping the school's climate, and issues related to sustainability of the teacher evaluation system that are of interest to wider formative supervisory practices.

Teacher Experiences of Formative Evaluation

Some teachers felt that receiving feedback allowed them to improve aspects of their instruction, while others felt that the process had no impact on their instruction. This contrast seemed to be due, in part, to uneven provision of support and feedback to teachers. However, teachers who were given similar supports also reported varied impact on their learning. The following short narrative illustrates a very positive example of a teacher who felt she received significant support through the formative teacher evaluation system at Elmwood School and that this helped her improve her practice. This will be followed by a discussion of how this case relates to other teachers' experiences at the school.

Heather is a seasoned teacher. She has worked with kindergarten students in three different Washington D.C. schools over the past ten years. She knows curriculum, she knows 5-year olds, she knows pedagogy, and she loves teaching. However, in her ninth year of teaching, she was a new kindergarten teacher at Elmwood School during Year 2 in its implementation of the formative teacher evaluation system:

[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." 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 [of other classrooms] 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 [the Elementary AP] at the time. And she 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... and I was so impressed that she came in Saturday morning and sat and literally cleared the table. We made anchor 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 Elmwood to me.

As a result of this support, Heather spoke about making many changes to her teaching. Most importantly, she spoke about a shift in her mindset when thinking about students. She gave many examples of how her experiences at Elmwood School helped her to view students as capable of more independence than she had previously given. For example, she talked about how she no longer felt she needed to structure student center choices. She switched grade levels after her first year at Elmwood, but she has tried to integrate what she learned from that experience in her new teaching context.

Clearly, Heather's experience describes a high level of commitment to students from both the administrator and the teacher. It also indicates that administrators had created structures to support the teacher evaluation system, and that there was a positive school climate that encourages teacher growth through the evaluation system. The teacher was able to pinpoint

necessary changes to her instruction and was confident enough to ask for help. The administrator was a responsive leader who was prepared with pedagogical knowledge and resources. She also went beyond her professional duty by providing time outside of school hours to help the teacher. It is evident that Elmwood's climate generally supported this behavior. For example, there is an expectation that students should be provided with a rigorous academic program which led Heather to recognize a need in her practice. It is likely that Heather felt that the school climate was one in which she could safely question her practice and be vulnerable with her administrator. As described earlier, Andre felt that it was important for all teachers to think of their practice as a "place of learning" rather than a "place of performance," so it may be that Heather felt that it was, indeed, okay to admit that she needed learning support. Heather's colleagues were also willing to allow her to observe their classrooms, which is another indicator of the professionally safe climate at Elmwood. In addition, the school's provision of a budget for professional learning opportunities heightened Victoria's (the former elementary AP) ability to provide learning opportunities.

Provisions like these are not common in most schools, and so it is unsurprising that the experience was both affirming and effective for Heather. Yet, this was not everyone's experience of formative evaluation at Elmwood School. Novice teachers reported more positive impacts of the teacher evaluation system than did experienced teachers. For example, Kathy reports that she got the most helpful and substantive feedback during her "first year [of] teaching" which was Year 1 of the formative teacher evaluation system. At that time, she "had lower rankings in certain areas. A lot of it had to do with the rigor or the timing of [my] questioning [students]." Because the administrator was able to give her "specific help and modeling [of] what that looks like," Kathy was able to immediately implement changes to her

instruction. It may be that novices feel that the feedback they receive can easily improve their practice because at least some of their areas in need of improvement are easily addressed through concrete suggestions and modeling. This may contribute to them being more willing to engage in evaluation (Hargreaves, 2005). However, the data also indicate that novices at Elmwood School may have been receiving more substantive and consistent feedback than other teachers which may be why novices at Elmwood found the feedback to be more helpful.

Additionally, as Heather's vignette illustrates, teachers who are self-reflective, able to specify an area of concern, and willing to ask for help are also well-served by the Elmwood teacher evaluation system. For example, Wendy, an elementary teacher, felt that she needed specific guidance on question stems for reading and was immediately able to get help. Kathy, a middle school teacher, reported that she identified a specific area of classroom management that she needed help addressing and got "quick fixes." She described how "[the assistant principal] would give me ... just quick things I could implement right away that would have a quick return on them and it would curb distractions or behaviors I was trying to avoid." Kathy appreciated these quick fixes because they helped her improve her practice immediately. Teachers who were proactive in asking for help, like Heather, also received support more readily. For example, Emily asserted "If you're proactive [in asking for learning opportunities], they'll support [you] as long as it benefits your students." Altogether, formative feedback was most likely to lead to change for teachers who were self-reflective, specific about their needs, and proactive. Nonetheless, not all teachers have these skill-sets or feel able or willing to name their shortcomings to their supervisors. For those teachers, changing practice based on teacher evaluation was mostly possible when administrators provided meaningful feedback with practice opportunities, which did not happen consistently in Year 3.

The teachers least likely to express ways that the teacher evaluation system helped them learn and enact improved practices were high-performing teachers. This finding was especially pronounced during Year 3, but was also apparent in high-performing teachers' descriptions of Years 1 and 2. Guskey (2002) asserts that veteran teachers are motivated to try something new only when they believe it will lead to improved outcomes for their particular student population. Some veteran teachers at Elmwood demonstrated this in their attitudes about formative evaluation and professional learning. For example, Kathy reported that she would usually try each suggested strategy from her observation feedback meetings right away. She would then determine whether she would continue to implement or abandon the strategy after just one attempt. If she found it successful, she would find many other instructional moments to incorporate the strategy, and if not, she would not try it again. Other researchers have discussed how resistance to change can be related to career stages in teaching; they find that as years of experience increase, teacher willingness to adopt new practices decreases (see for example: Hargreaves, 2005). It might be logical, therefore, to conclude that more experienced teachers at Elmwood would be less likely to find formative evaluation feedback actionable and helpful because they see no need to change their instruction. However, interviewed teachers expressed quite the opposite feelings. For example, Emily, a highly proficient and distinguished teacher and one of the district's Master Teachers, explained, "I think regardless of how long you've been teaching or whether you're a master teacher or the lead teacher, I love feedback. I think that's the only way that you can grow... whether you have been teaching for 15 years or 20 years you're still constantly being evaluated and observed because that's how you grow and that's how you become great in your craft." Denise, another highly proficient teacher, also appreciates feedback. She said, "I've got it under control, but that doesn't mean that I don't need support."

In spite of this openness toward and interest in feedback, veteran and proficient teachers were less likely to feel that the evaluation system served their needs well. This confirms findings from the literature review that veterans have different learning needs than do developing teachers (Donaldson, 2012; Halverson et al., 2004). At Elmwood School, experienced teachers were more likely to get generic and rather empty feedback, perhaps because the teacher evaluation rubric and system was not well-suited to engaging teachers with above-basic needs. For example, Wendy felt that the “rubric is too general” to help with her attempts to understand difficult areas of her practice. When she was observed informally, she was simply told that she was doing a “great job”, but had “no idea” what it was she was doing so well. Such generic feedback was unlikely to stimulate teachers’ growth and caused teachers to feel unsupported. Veterans also received less feedback. For example, Denise, who is a new middle school teacher at Elmwood, but had several previous years of experience, described that her administrator:

Came into the classroom probably pretty heavily the first three months, and then that was it, and I didn't see him very much after that... And he said, “Oh, you're doing great. And the reason I have not been in your room is that I have bigger fires to put out.” He said, “I'm not worried about you.” But I was so worried about myself because I felt like I wasn't being the teacher that I had been before at the other school.

Denise described feeling deeply isolated as she was left to figure things out on her own. For example, she discussed how, although the evaluation rubric itself is “very thorough... I feel like I'm not getting enough out of it.” Specifically, she wanted support from administrators and next steps for improving her practice. Even when administrators provided veterans with areas to improve upon, they did not give resources or support tools to help them improve. For example, Kathy described how, in her most recent formal observation, one area of practice was identified

as proficient (3 of 4) rather than distinguished (4 of 4). She agreed that this was an area of teaching that she felt challenged by and asked for help. However, she found “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 talked about feeling that the responsibility for change fell squarely on her even though one of her three yearly goals aligned with a school-wide goal of better stakeholder communication and involvement. She explained that her administrators provided some suggestions, but no “physical resources.” Wendy also felt that when trying to move from an evaluation rating of 3 to 4, she was not given “tangible direction” and was instead simply told she should try to do better.

In all, although identifying and working toward teacher goals was important to administrators, veteran teachers seem to have been expected to move toward meeting them with little direction or support. It may be that, for these reasons, experienced teachers were more likely to describe their teacher evaluation feedback as not contributing to their learning process. For example, when asked about feedback that helped her grow, Wendy said, “Honestly, I can't think [of anything] off the top of my head. I'm sure that there are because I feel like I've grown because I am getting feedback, but I can't think of anything.” Later, she more decidedly concluded “I have [learned how to engage readers on a difficult task], but not from feedback... I kept thinking as a third grader. Like, what I can do to help my students ... So I developed a strategy for them.” Kathy also made clear that evaluation feedback had not helped her improve her instructional practice:

For example, if you're teaching surface area and how to make a net, [it] is a much harder lesson than comparing two ratios. And so [for] some lessons in math that [teacher evaluation] rubric [does not address] the quickness at which kids are going to get it...And the feedback I got was, “Well, how could you have slowed it down? Is there a list of steps

you could have given?” And the answer is no. This is a day one of a five-day intro. No, making an anchor chart won’t help.

Instead, veteran teachers, especially those that were concentrating on learning more complex pedagogical strategies, identified collaboration with peers and self-reflection as much more impactful than the evaluation feedback they had received.

When questioned about providing high-performing teachers opportunities for learning, the principal described other types of structures that he thought would increase high-performing teachers’ buy-in to the school. Not surprisingly then, high-performing teachers felt that the opportunities they were offered were more a means of keeping them employed at Elmwood for longer and of rewarding their commitment to the school than opportunities for learning and growth. For example, veteran teachers are encouraged to attend trainings that will benefit the school in some way. In essence, investments in high-performing teachers are meant to “develop [other] adults in the building” (Kathy), not necessarily to better the teacher’s own craft. Kathy observed that the offerings were not “necessarily a way to give me feedback on the goal that I want [to achieve].” Instead, she was given the opportunity to “be a coach or to affect school culture.” High-performing teachers could also apply to be master teachers. As Kathy described, the training and master teacher opportunities came with two-year contracts, so it was also “a way to keep [veterans] in the organization for longer.” As these examples illustrate, the school is working to create structures that encourage veteran teachers to invest in the organization, but it does not necessarily support veterans’ individual growth as teachers, which suggests that differential support for teachers in this stage of their craft is necessary.

“I want to feel supported, not just evaluated”: School Climate and Teacher Evaluation

This section describes how elements of the school’s climate and the teacher evaluation system were intricately connected so that when one was strengthened, so too was the other, and when one faltered, so too did the other. As a result, stakeholders’ feelings about how the evaluation system was implemented impacted how they felt more generally at the school. When evaluation was well-implemented, teachers felt like they were supported to learn and grow, but when it was not well implemented, teachers felt like they were subject to surveillance. The relationship impacted how teachers felt as professionals in regards to safety, urgency, and buy-in.

Professional Safety. In order for teachers to feel comfortable learning from a teacher evaluation system, they need to feel safe about taking pedagogical risks and deprivatizing their practice by opening their classroom to scrutiny and engaging in conversations about their instruction. In this study, all teacher participants seemed to be comfortable sharing their practice with administrators and other teachers and many voiced that they wanted more opportunities to get feedback from both. For example, Heather’s vignette reveals that she felt it was safe to ask for help and to open up her classroom to immediate scrutiny, even though she was a new teacher in the building who was also concerned about job security. Across job roles, members of Elmwood School communicated their understanding that the informal teacher evaluation system was a truly formative process meant to support teacher growth, not to catch flaws. Andre, the principal, described the system as “a very genuine way to build practice and build better professionals, without playing the blame game or coming across as antagonistic.” Teachers seemed to feel this; they openly embraced the feedback process because they felt that administrators were “not coming to find wrong, [but were] actually coming to support my instruction and to 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” (Pamela). Teacher assertions that a positive climate created an environment more conducive for learning is supported by research that finds that teacher change is an affective, rather than cognitive process, very much linked to the school environment (Kotter & Cohen, 2002). Additionally, the focus on a growth mindset allowed teachers to feel safer being transparent about their challenges and asking questions. The structures and consistency of the evaluation system helped teachers feel they could comfortably talk about growth areas. For example, Emily described how her assistant principal was not just “coming in this one time, and I haven't seen [him] all year, but this is just what we do. It's an everyday thing, so I'm more comfortable discussing with [him] my flaws and my personal growth areas, and I'm also more comfortable receiving [constructive feedback] from [him].” Administrators further supported this sense of safety by maintaining “an open door policy” and by allowing for teacher choice in how they implement suggestions that come from the evaluation data. When the Elmwood School's teacher evaluation system works well, it is coupled with a supportive environment linked to a truly formative approach to teacher learning. Thus, it encourages teachers to both take risks and seek out administrative support for taking those risks and when they experience difficulty.

Urgency. Elmwood School can be characterized as a cognitively busy place where all adults in the building are focused on the academic outcomes of the students and all students are engaged in demanding work. This phenomenon is often referred to as a sense of urgency in school leadership research (Kotter, 2008). In particular, when researchers discuss closing the achievement gap for students living in poverty, they often encourage schools to adopt a sense of urgency as a necessary requisite to providing instruction that changes the trajectory of student

achievement. At Elmwood School, a school that serves a high population of students living in poverty, teachers did not speak explicitly about a sense of urgency to effectively meet the needs of students who, by demographics, would be predicted to be most likely to underachieve.

However, this seemed to be communicated implicitly in many ways. Instead, teachers referred to consistently seeking ways to increase the rigor of their instruction, seemingly with the goal of increasing performance for their student population. A sense of urgency was also communicated throughout the school environment in many ways. For example, data displays of aggregate student achievement data were posted in many hallways and all over the staff rooms. Slogans and captions encouraged viewers to recognize student growth or analyze areas where students could immediately become more proficient. Additionally, teachers spoke with pride about their students' growth in relation to achievement scores. Further, professional dialogue centered on the importance of engaging students in learning through rigorous expectations and activities.

During Years 1 and 2, the teacher evaluation system was a natural complement to the sense of urgency regarding student achievement in the school and positively contributed to the focus on rigorous instruction in the school. When consistently implemented, teachers felt that it was a useful practice that helped them increase their effectiveness and that it held all teachers accountable for meeting high standards for student instruction. This is not always the case in school settings. For instance, in their study of teacher evaluation practices, Weisberg et al. (2009) assert that although teachers and administrators were aware that some teachers were not effective, those teachers never received feedback about their instruction that could help them improve or to reevaluate whether teaching was an appropriate choice as a profession. As a result, many teachers felt burdened to carry the workload of ineffective teachers in order to make sure that student achievement did not decline. In contrast, teachers at Elmwood felt that the

teacher evaluation system encouraged all teachers to support each other in providing high levels of instruction. The rubric clearly communicated what teachers and students would be able to do in classrooms that provide rigorous instruction. Further, when the teacher evaluation system was consistently and effectively implemented at Elmwood School, teachers were able to clearly and immediately understand areas of concern related to their practice and were given resources and supports to improve those areas. This, created a culture where teachers feel a sense of urgency to continually improve their own practice. As Heather's vignette demonstrates, because of the school's focus on rigorous standards, all of the students were much further along than she had come to expect in another setting. Thus, she immediately recognized a need to modify her practice because she was aware that her approach to teaching would not meet the more advanced needs of her students and was not well aligned with the school's expectations for rigor. Through her meetings with administration and her observations of other teachers at the school, Heather began to understand that she could expect a lot more independence and higher-order thinking from her students and that her practice would need to reflect that.

One outcome of this sense of urgency was a widespread understanding of and commitment to using student data to inform instruction. Teachers spoke quite confidently about engaging student achievement and their steps to doing so. They often linked their own progress in meeting professional goals to student achievement data and referenced making this connection as a core value of the school. For example, Wendy described how as part of the formative evaluation system, teachers set student growth goals at the beginning of the year. Teacher binders, described earlier, were used to track teacher goals and student achievement data in relation to those goals. This was such a pervasive part of the school culture that many teachers were often able to quantify their students' progress without referencing any notes, indicating that

scores were analyzed often. Though one might infer that memorization of scores might represent compliance, I did not get the sense that this was the case, but that teachers truly felt vested in understanding, analyzing, and celebrating student growth and achievement tests. Generally, such a school culture urges teachers to continually align their teaching with high standards for their students and often is linked to teachers holding high professional standards for themselves.

Another outcome of a school-wide sense of urgency is that teachers who are not performing well are immediately made aware that they need to improve their practice. As Andre described, the sense of urgency primed administrators to be able to immediately intervene with low-performing teachers. Administrators committed themselves supporting better instructional practices for those teachers. He further explained, “If we're saying that this teacher is struggling, what are we doing to support her? Is our support not enough, or are we not supporting them in the right way? So it's not a blame game.” Andre believed that this provided a more rigorous system in which students could have access to the best teachers and effective teachers could have other effective teachers to support their work. He further reported that, because teachers were aware of the expectation that they meet student and professional standards, many teachers who knew that they would not be able to contribute effectively to Elmwood left the school on their own. Others recognized a need for growth and dedicated themselves to improving their practice by working closely with their administrators. For example, Andre describe how in the past three years, he has only needed to terminate one teacher because that teacher, though passionate, was unable to improve his practice. It was a difficult decision for the administrative team, but in the end, Andre felt encouraged because:

The best part was that he said he understood why I was letting him go, and he understood that he was not the best person to have in front of the kids. And that's all that mattered to

me. For him to take away that it was nothing personal, it's all about making sure that we have what's best for the kids...As long as anybody in my building walks away from any conversation that I have with them with that, I'm okay.

The administrative team kept formative evaluation at the heart of the mission and vision of their schools, such that a sense of urgency to improve was a way for school culture and teacher growth to be simultaneously supported. With vision and mission central, teachers were able to immediately gauge how their practices aligned with the requirements of the teacher evaluation system and personnel decisions like hiring, maintaining, and terminating teachers felt less personal.

Buy-in. When well-implemented, the teacher evaluation system also generated teachers' loyalty because they felt that engaging with it would be good for their practice. When recalling events in Year 2, teachers talked about it as an accurate and helpful process that they felt was worthwhile and would, therefore, happily buy-in to. Heather's vignette demonstrates buy-in from both the teacher and the administrator who devoted a lot of time to using it as a way to create classroom change. Other teachers describe that, because feedback felt like a true reflection of their work, they felt that engaging with it would help them improve their practice. For example, Angela stated that her improvement goal was appropriate, which she felt was a clear indication that Victoria, the assistant principal who had constructed the goal, was actually in her class enough to understand her practice. Pamela, another early childhood teacher, expressed her willingness to participate in the evaluation system because it felt like a formative rather than a punitive process. As she recounted, "When [the assistant principal is] coming in, she's not coming to find wrong. She's actually coming to support my instruction and to 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.” This made Pamela feel that the administrator was qualified to comment on her instruction. While Pamela described her interactions with the administrator as “welcome,” Heather went so far as to describe a familial warmth toward her administrator because she was so acquainted with her and her classroom. She said, “We're more like a family ... I think it helps when they know your students and then you're able to take that feedback differently.” Kathy was also certain that whatever feedback she received was based on things that were “present” in her classroom and that she was given the opportunity to show that she had grown in an area before her final, formal evaluation. As these experiences illustrate, teachers were open to the process and to scrutiny because they felt that the school’s climate related to evaluation was one focused on growth and encouragement, and that administrators seek an accurate interpretation of what is happening in their classrooms as a way to help them improve their practice.

Others discussed how they had bought into the evaluation process because they felt a part of the change process. As such, teachers felt that they could all get behind and support the implementation of the teacher evaluation system. By the end of Year 2, stakeholders seemed to so fully buy-in to the evaluation system that it became a normal extension of teacher practice and an integral part of the school culture. For example, Lisa reported that the evaluation rubric became so familiar and integrated into the teachers’ thinking and practice that after a while they did not need to reference the physical paper on which it was printed. They had many conversations about it and talked so much about how to align their practices with it and they had had many opportunities for feedback and guided practice. In essence, teachers were simply doing “what's found in that rubric” without needing to reference it and had incorporated elements of the rubric so fully into their practice that they just became “mindful of it” (Angela). Angela

described that when her administrator filled out the rubric it was more of a “formality.” She seemed to indicate that she felt it was no longer necessary to talk about the rubric because it had become a lived part of her practice. As she explained, by the end of the year, other documents and practices were created to align with the rubric. For example, the early childhood team created an “environmental checklist” that aligned with the rubric so that all early childhood teachers could create positive classroom spaces based on it. Over the course of Years 1 and 2, teachers moved from actively attempting to align their practice with the rubric’s to seamlessly integrating it fully into their daily instruction.

However, in Year 3, administrators’ overwhelming caseloads left little time for follow-up feedback. Teachers felt a tangible shift in Elmwood School’s climate related to the teacher evaluation system which caused them to invest less. For example, Kathy described feeling rushed by what felt like a bureaucratic process:

Sometimes I would say [the evaluation system] feels very, very rushed. Kind of like administrators have to get through this stuff. A lot of times, it’s not viewed as- at least this past year -my experience was- it’s not viewed as a tool teachers should really be getting a lot [out] of. More as for administrators to communicate what they think or what they saw for a small period in your classroom. It’s not really used as a tool for me to learn about my teaching. It is very much like ‘here it is, check it off, here is your evaluation.’ But there’s not a lot of follow-through on ‘you scored this and next year we want you to be this.’ Those types of conversations don’t happen very often, or that’s just been my experience this year.

Kathy’s comments point out that in addition to classroom observations being too brief in Year 3 to capture her practice, administrators were often unable to keep up with giving feedback and

related evaluation activities. Written feedback from the evaluation rubric, alone, was not enough to help improve Kathy's practice. Her experience is reflected in evaluation research (see for example: Danielson, 2011). Just as in other settings, teachers at Elmwood report that they needed immediate feedback and conversations that extend learning from the tool after every observation session. When the school provided just written feedback or just a rubric without the opportunity for conversation, teachers questioned the purpose of the evaluation process and were less likely to view teacher evaluation feedback as helpful for practice. For example, Kathy wondered about her experiences in Year 3:

Who uses it? Who uses the actual feedback that they get out of these very long intricate rubrics they use to rank teachers? ...I don't know that if nobody pushes you to follow through on it, that anything comes out of it, because maybe you're okay with seeing you scored ones or twos. If nobody pushes you to do anything about it, then its ineffective, but if that's what's going to motivate you to better yourself [maybe it will be beneficial], which I doubt scores on a piece of paper will.

Kathy's reference to the rubric as just a "piece of paper" is a very telling indicator of how distant evaluation and teacher practice became in Year 3. This finding further reinforces that the administrative personnel changes were detrimental to teacher's ability to learn from the teacher evaluation system and, thus, lessened teacher buy-in.

In addition, whereas in Year 2, teachers received written feedback within minutes, and verbal feedback within days after an observation, when teachers were observed in Year 3, they often did not receive feedback at all, or had to follow up several times before receiving feedback. For example, Kathy described how after following up with the principal several times, she finally received written feedback a few minutes before she was set to have her post-observation

meeting. This meant that she did not have time to look at the feedback or reflect upon it in time to use the meeting as an opportunity to deepen her understanding of her practice. Teachers reported that administrators also used the post-observation evaluation meetings to attend to other work. For example, Denise reports that the principal told her that her observation was “fine” and spent the rest of the meeting talking about summer school plans. Teachers felt that this communicated that using evaluation feedback was not worthwhile. For teachers, who had the most to lose from inconsistent support, the lack of observation and feedback led to feeling isolated, which, in turn, led to anger and resentment. The surveyed teacher’s comment about feeling evaluated and not supported is indicative of the shift in climate. At times, teachers felt that they needed to be responsible for following up with leaders, which led to even more resentment. For example, Denise described her anger towards the math coach, Joan, because she “made a schedule [of] when she was supposed to come to my classroom, and she didn’t [come].” Denise gave up on trying to work with Joan “Because it’s not my job to hold her accountable.” Overall, due to changes in practices related to the teacher evaluation system in Year 3, teachers experienced a climate shift around accountability practices in the school. They described feeling like they were no longer supported. Instead they felt evaluated, and that they had to take on responsibilities that were not theirs. Teachers were no longer being held accountable for working toward meeting their goals or encouraged and helped to learn and grow, leaving them alone to be “baptized by the fire” (Denise, Middle School Teacher). As the climate shifted away from true formative feedback and the opportunity for professional growth, teacher buy-in suffered.

Leadership Lessons Learned

The study site offers many lessons about sustaining a meaningful and constructive teacher evaluation system including the importance of training new leaders and sustaining innovation during periods of turbulence. Although many opportunities for teacher learning were lost in Year 3, the principal spoke with a lot of depth and resolve about how this year had helped him to understand the process of sustaining teacher evaluation more deeply. The principal has come to understand the importance of upholding “systems and procedures... and equipping all leaders with the ability to do the same level of fidelity of work, so that if we lose somebody” teachers do not lose learning opportunities.

The data reveal that many actions are necessary to support the work of teacher evaluation, starting at the district level. It seems intuitive that adequately staffed schools would be a prerequisite to full implementation. It was no different at Elmwood where “the data [from the teacher surveys] shows... considerable growth” in teachers’ perceptions of the school’s climate when two APs stayed in their roles over a few years (Andre). Yet, Elmwood School was only sufficiently staffed for two of eleven months in Year 3. The district level inhibited implementation of the teacher evaluation system by failing to hire, train, and retain effective leaders and to plan for staffing changes. Another issue was that the district assumed that administrators could wear many hats. For example, the math coach was also the testing coordinator. Additionally, both interviewed administrators reported having to restructure their evaluation and feedback schedules to attend last-minute meetings at the district level. Though this cannot always be avoided, if the district values an 80% commitment to observation and feedback, district administration needs to be just as dedicated to this time as school level administration is (Kraft & Gilmour, 2016a) by keeping last minute meetings to a minimum.

Changes in the administrative team's workloads had a significant impact on the implementation of the evaluation process. However, there were other school-level factors that also contributed to a less robust evaluation program during Year 3 including effective leadership induction, better ways to monitor implementation, and ways to safeguard systems in the event that administrative personnel changes. It was apparent that the principal needed to more thoroughly initiate new leaders into the expectations, values and procedures of Elmwood's teacher evaluation system. New administrators were not trained with the same level of intensity as the first cohort of leaders had been. For example, they did not attend the Leverage Leadership Institute. As a result, the elementary AP, Dana, and math coach, Joan, who were both new to the district, were not providing the same kind of feedback and support that teachers had come to expect in previous years.

It seems that in addition to a lack of training, the principal did not fully communicate his expectations for formative evaluation to new administrators. This was apparent from the ways that teachers described the activities in which new administrators engaged. When teachers described the school's International Baccalaureate (IB) and English, Language Arts (ELA) coaches, who had been part of the initial implementation of the teacher evaluation system in Year 1, they talked about receiving constant feedback and having support to plan effective lessons and think through student achievement in relation to observation data. In contrast, the math coach, Joan, who technically had the same job requirements, but was new to the district, was described as pulling small groups of students from the large group setting for remedial instruction and creating classroom resources. While the IB and ELA coaches were providing instructional leadership, the math coach was providing supplemental instruction for students and assistant teacher-like supports. Maybe as a result, whereas the IB and ELA coaches were always

described as positive influencers of practice, almost all participants had very strong, yet contrasting, feelings about Joan. Wendy, an elementary teacher, described her as “great” three times and a surveyed middle school teacher commented that Joan’s support has been “highly beneficial.” Yet, Denise, a middle school teacher, referred to her as “worthless” and felt that she “wasn’t doing anything.” Lisa, an early childhood teacher, indicated that she had not interacted with Joan at all and believed that this was because her focus for the year was to support middle school teachers. Together, these comments indicate that Joan (whom I did not interview), may have been attempting to give support that was outside of her purview while failing to follow through with requirements of the formative teacher evaluation system or the expectations of teachers based on prior experiences with formative feedback.

Descriptions of Dana, the new elementary AP, indicate that she too was not fulfilling the formative evaluator role as it had been previously defined. When looking at each administrator’s workload on the online Google tracking system, I saw that she had input many more entries than other administrators in the building and was logging meetings about twice weekly with each teacher. I felt that she was clearly keeping up with Andre’s 80% requirement for observation and feedback sessions. Andre also felt that she had met the requirements of his mandates based on frequent spot-checks of the Google tracking system and check-ins with her during weekly meetings. However, I observed that Dana’s ability to construct continuous plans for improvement seemed lacking. While Michael focused on the same goal from week to week until it had been reached or introduced new goals that were related to previous goals, Dana was not as systematic. Andre did not seem to notice this during his spot-checks. Moreover, without asking teachers about their interactions with her, he did not know that teachers were not getting the cohesive, structured feedback over time to which they had become accustomed in previous years.

This only became apparent to Andre later as a result of the teacher survey that was administered in January. However, by the time results were analyzed, there was little time for him to intervene because Dana had already accepted another position that would start in early March.

During my interviews in late March, teachers also relayed that they missed some aspects of the feedback process that they had experienced in the previous years. Whereas the previous elementary AP, Victoria, gave resources, demonstrations, and opportunities for practice after each feedback session, as described in Heather's vignette, Dana only gave feedback in the moment. Although she may have been following the guidelines dictated by the principal and the Six Steps of Feedback process, she was not integrating feedback with other learning processes. Also she did not give feedback with continuity over time, as Victoria had done. For example, Angela commented about a lack of follow through from one meeting to the next:

Last year when we chose our goals on [the rubric], we actually revisited it mid-year and at the end of the year. So we knew our goal because the AP made sure that we knew our goal and we were hitting the mark... And we also met a lot as a group with the AP and talked about the goals and what we could do to improve as a Pre-K Department. So it was a little more organized and we knew the expectation.

Angela's description of a previously more organized and systematic formative evaluation approach was also clear when elementary teachers described their access to resources and collaborative learning practices. Though teachers described the use of structures that supported the formative teacher evaluation system such as teacher binders, opportunities to observe each other's classrooms and use the evaluation rubric to talk about those observations, and videotaping in detail during conversations about Year 2, they never mentioned any of these when talking about Year 3.

It is evident that both of the new administrators were very unclear about their roles and responsibilities within Elmwood's teacher evaluation system, and so they relied on their own understandings of what they thought their role should entail, often unsuccessfully. In Year 3, new leaders had been introduced to the Six Steps of Feedback and the teacher evaluation rubric in August, and had taken part in the weekly norming meetings through which they understood the practices of observing classrooms and giving feedback. However, they did not seem to understand the deeper elements of creating a cohesive and supportive evaluation system. As the literature review describes, creating systems that provide high-quality feedback is "more challenging than simply increasing the frequency of observations" (Aldeman & Chuong, 2014, p. 7) or increasing the amount of training observers receive on increasing reliability (McKay & Silva, 2015). Though the principal had trained new administrators on how to enact a set of evaluative behaviors, he had not understood the importance of clearly communicating and supervising their role so that they understood that evaluation at the school meant more than simply providing feedback. Since the new leaders were unaware of how to support instruction through practice opportunities and through linking teacher evaluation feedback to other systems (like the binders, the collaborative house meetings, and videos), there was uneven implementation in Year 3. Andre realized that he could have created a more standardized protocol for principal induction to alleviate these gaps. As he admitted, "had I been more accountable of holding up the expectation of what the Six Steps of Feedback meeting looks like from the beginning, I don't think it would have ever gotten that loose."

This finding may also indicate that Andre was not yet able to transform his vision for the teacher evaluation system into a fully-integrated system at the school. For example, Andre spoke frequently about his requirement that leaders spend 80% of their time in classrooms doing

observations or giving feedback. However, Leo described his observation and feedback loop as taking place on Tuesdays. Further, even that time was often infringed upon because, as Leo explained, “things happen.” Clearly, one day of the week does not equate to 80% of time dedicated to formative feedback. Though Andre describes meeting with administrators to set up yearly calendars and block out time for formative evaluation, Leo’s description is a clear indication that Andre’s vision was not translating into practice. Even Andre seemed unsure of what had happened when he described the changes in Year 3. He said, “The information that I was getting from [Dana on a weekly basis] was completely different than the feedback that I got from the teachers at the end of the year.” This illustrates that when implementing formative evaluation systems, administrators must go beyond compliance with basic expectations by shifting their perspectives about the ways in which to give formative feedback. This disconnect had not been bridged at Elmwood school where most administrators were able to keep up with behavioral expectations (such as implementing weekly meetings, reviewing the online Google tool, following the principal-created calendar), but did not, or could not, follow through with the more difficult process of engendering teacher learning through the process. In all, because there was not a system in place to adequately train new leaders and administrators were not yet fully implementing all aspects of the teacher evaluation system as it was conceived, the teacher evaluation system was not well-sustained when the school experienced changes.

In addition to training, clearly communicating and, supervising administrative work, there needed to be strategies in place to more adequately understand how teachers were experiencing the evaluation system. Many stakeholders mentioned that school leaders should be held accountable through administrative evaluations (sometimes called “reverse evaluations”). Andre also talked extensively about utilizing the Insight Surveys more frequently to understand

teacher perceptions of the administration's work and of the school's climate. More frequently surveying teachers will be an important way for the principal to gauge teachers' perceptions of the school's climate and the evaluation system. However, it seems that there was also an issue with ways in which Elmwood's administration used the survey data. Although Andre held a meeting to discuss the school's goals based on the teachers' feedback, one teacher reported that she felt that teacher voices were only taken seriously on "that day" because there has been no follow up about "what [administrators] are going to do with the [goals from the survey feedback] next year" (Denise). It seems that other measures of teacher perceptions, like informal check-ins from each department might help administrators better understand teachers' reactions to the evaluation system. These could also be a more consistent way of communicating to teachers that their voices are heard and they have an "opportunity to tell admin what [they] need ...and [are] not getting" (Denise).

Additionally, the Google tool could be used more efficaciously. While Andre recounted using the tool frequently to track that administrators were getting their observations and teacher meetings done, it seems that he should also have tracked the ways in which administrators were using observation data to construct feedback. As an outside observer, it was clear to me that the feedback that Dana, the new elementary AP, was inputting was qualitatively different from that of other administrators. Quantitatively, she was following through with her observation and meeting schedules, and so this may have been interpreted as successful administrative practice. However, it may be that by focusing on quantity of observations more than the ways in which feedback was constructed, Andre was inadvertently communicating that simply being in the classroom for 80% of the day was the most important administrative action and that the quality of subsequent interactions that resulted was less important. This may have been the reason that

teachers were left feeling that the actual feedback “wasn't the highest caliber as they [had] received in the year prior” (Andre).

Finally, the principal signified a need for distributed leadership, so that if personnel change, systems are “safeguarded.” For example, Andre talked about how, in response to the changes in personnel, he initially responded by taking on the full responsibility on his own to “make [evaluation] happen across the entire campus.” He recognized quickly that this was unsustainable because “there's only one of me.” Instead he began to take a more realistic look at what is possible when systems experience turbulence and concluded that distributed systems of leadership must be in place when attempting to safeguard such weighty and detailed processes as the formative evaluation program. Because “every school loses people,” Andre began to think of ways to create “a plan that supports consistent observational feedback no matter how many people you have” in order to keep the “momentum of learning” through “continuity... and fidelity.”

He saw two processes as being necessary for continuity. First, he recognized that he could not have possibly increased his presence in teacher classrooms without redistributing other, less critical, responsibilities. Andre felt that there were many areas of school leadership that could have been handled by other members of the staff so that he could focus on the more important work of teacher learning. Some examples include “literally sitting someone in my place” for non-essential meetings and empowering others to handle student discipline concerns and parent meetings. The second was to train more people to give feedback to teachers. For example, Andre put plans in place to have Master Teachers observe and support advanced proficient teachers in 2016-17, so that school administration could spend more time with new and struggling teachers. He was also able to encourage the district to take on a similar initiative

and to provide an extensive observer training program to Master Teachers in the summer of 2016 (pre-Year 4). In all, though Year 3 brought many losses and changes, the principal used the opportunity for reflection that helped him regroup and begin to create structures that would lead to more effective and more sustainable implementation of the evaluation system in the future.

Chapter Review and Conclusion

In this chapter, I have used findings from interview data, focus groups, and school artifacts to describe the Elmwood School teacher evaluation system and to argue that shifts in the implementation of teacher evaluation greatly impacted potential for teacher learning and the school's climate. To make this argument, I began by outlining some of the school-level practices that Elmwood administrators employed to create an effective teacher evaluation system that was viewed positively by many of the teachers. These included some that matched the recommendations for best practice from the literature including distributed evaluator caseloads, evaluator training, a focus on formative feedback, and linked professional learning. I also identified two additional processes that are supported by the data, namely, the importance of the administrators' commitment to growth and ability to link other professional learning opportunities to teacher evaluation. Next, I gave examples of how these supports not only led to increased teacher learning through the evaluation system, but also contributed to a positive school climate, including professional safety, professional urgency, and buy-in during Years 1 and 2. I also demonstrated how difficulty sustaining the teacher evaluation system and the resulting breaks in teacher evaluation programming at Elmwood were linked with changes in teachers' perceptions of the evaluation system and of the school's climate. Major changes in the third year of implementation contributed to teacher frustration with the system.

Overall, teacher perceptions of feedback and its potential for impacting learning are mixed at Elmwood School. This is perhaps due to changes in implementation from year to year, a reflection of the difficulty of implementing and sustaining formative teacher evaluation systems, and an indication that formative evaluation systems should not be the only structure for supporting teacher learning. For some teachers feeling supported to take professional risks and be guided to better practices led to huge shifts in their practice and strengthened their commitment to the school. For others, especially veteran teachers, the feedback felt misaligned with their needs and practices. This finding will be expanded further in Chapter Five when I present a final area of concern that emerged from the data regarding teachers' feelings that collaboration with other teachers could have been more beneficial than feedback from teacher evaluation alone.

Finally, I described some of the lessons that the principal learned over the three years of implementation and the ways in which he plans to get the school back on course with a sustained and embedded teacher evaluation system. These findings will also be discussed further in Chapter Five where I will present implications for practice and policy. I will also propose implications for research in several areas of teacher evaluation.

CHAPTER FIVE: DISCUSSION AND IMPLICATIONS

In this final chapter, I discuss the findings from my case study in consideration of my research questions and literature review. Several theoretical and practical implications arise from these findings. Importantly, I modify my theoretical framework by presenting a new theme and related literature about situating teacher evaluation systems within shared teacher learning structures. I conclude by identifying missed opportunities and describing implications for practice, policy, and research, including changes that have already begun to take place at Elmwood School as a result of this research.

Discussion of Findings

Through the research questions in this study, I sought to learn more about school-level structures that support the use of teacher evaluation data and the resulting outcomes for teacher learning and improved practices in their classrooms. Because I hoped to make recommendations about positive implementation of formative systems, I spent many months searching for a school site that implemented best practices in formative supervision, as outlined in the literature review (see Chapter Two). Upon visiting the research site, the data collection process led me in different directions than those originally identified in my research questions. The school had experienced a turbulent year due to the mid-year loss of both the elementary and middle school assistant principals. In addition, the principal seemed to have underestimated, from the beginning of the year, the impact of new staff on program implementation. He incorrectly assumed that new school administrators would easily adopt the norms of the teacher evaluation system through practice, and chose not to provide detailed initial training to new leaders. As he

described, he thought “getting better at [observing and providing] comes with the practice [along with] coming back to the table as a team and norming.” However, without clearly communicating and supervising the work of these leaders, evaluation procedures and outcomes deteriorated. As a result, I will discuss findings in relation to my first research question on school structures by expanding the question to include ways that the leadership team could have been more likely to sustain change over multiple years and with new personnel. In regards to the second research question, I found that teachers who did experience positive learning experiences and improved their practice were more likely to speak about this occurring during Year 2. Therefore, this study provided limited guidance on teachers’ actual process of using teacher evaluation data, but at the same time suggested important areas for improved implementation. Thus, I will discuss the limitations of this study in relation to the second research question.

Research Question 1: By studying a school with a strong commitment to formative evaluation, what can we learn about school-level factors that support teacher use of evaluation feedback?

The first research question focuses on school-level structures that support evaluation: This section provides two frames by which to understand research question 1, utilization-focused evaluation (Patton, 2012) and sustaining practitioner change (Fullan, 2005; Kotter, 2012).

Utilization-focused evaluation. Patton (2012) proposes that in order for evaluation to be purposeful, organizations should judge the merit of evaluation programs not only on their ability to collect, analyze, and present data, but on the degree to which they are actually used. He refers to this as “utilization-focused evaluation.” In order to implement utilization-focused evaluations, he directs organization to “think evaluatively,” by instilling evaluation into organizational culture and everyday activities (p. 144). In so doing, evaluation becomes a central learning function

within schools. Patton's framework is helpful for schools that seek to incorporate formative evaluation because it points out that if evaluation is to actually impact teacher learning, it must be regularly utilized and embedded with the act of teaching (Chappuis & Chappuis, 2007; Scriven, 1967).

Elmwood School provides some important examples of this concept. Findings from this study show some successes in utilizing teacher evaluation that begins with the selection and integration of a well-developed rubric based on best teaching practices to promote shared language and values about instruction. In Year 2, the school had seen so much success with the evaluation system that teachers indicated that the rubric was embedded and automatically present in their thinking about teaching and learning; teachers felt it pervaded their daily understanding of how to enact instruction. At this point, teachers described the rubric itself as simply a "formality," a "reference," and a "fallback" (Angela; Lisa; Pamela) that they believed had become firmly aligned with their understanding of good instruction such that they did not need to talk about the rubric anymore. Instead, they felt that they "were already doing" instruction as described by the rubric and they could easily be "mindful of" incorporating practices from the rubric into their daily instruction without really referencing it (Angela).

In Years 1 and 2, the teacher evaluation system was situated within a fully-conceived and carefully articulated system. Administrators were well aware of their required role in implementing the program and using the structures designed to support it, such as how often they were supposed to evaluate and meet with teachers. They were also gaining confidence in creating cohesive, coherent, and consistent plans for teacher improvement and helping teachers develop their self-reflection skills. To further support this, the school implemented its own leadership training program to support the teacher evaluation system that included weekly

calibration sessions. As a result, administrators felt they could effectively lead instructional change. Leo reported that the system structures helped him “think about what is going to be the best leverage of that teacher’s time in that moment and how other things will tie into that” (Administrator, Interview). Leaders also linked teacher professional development opportunities to the evaluation process which also contributed to teachers’ capacity to improve their instruction. Observation data and professional development planning were explicitly linked; administrators met weekly to discuss what learning needs they were observing across departments and planned professional learning opportunities around those needs within the week. The most powerful example of teacher professional learning through evaluation that was shared during interviews was Heather’s experience during Year 2 when she described how incredibly supported she felt by her administrator who provided time and resources for her to observe other classrooms and talk about best practices, attend professional development, plan more responsive lessons, create materials, and get immediate feedback on her efforts to improve her instruction.

According to Patton (2012), the most important structure a school can provide to support evaluation is the use of “metaevaluation.” In metaevaluation, administrators constantly think through, assess, and improve evaluation practices. Indeed, I observed metaevaluation at Elmwood; I refer to this as administrators’ openness to change in Chapter Four. Elmwood administrators constantly used multiple sources of data to critically analyze outcomes of the teacher evaluation system. For example, they administered school climate survey twice a year to all staff and used it to support analysis and reflection regarding the evaluation process. Further, when Year 3 results revealed that teachers did not receive the quantity or quality of feedback that they had received previously, the principal immediately identified actions he could take to increase administrator understanding of and accountability to the teacher evaluation system.

More generally with regard to metaevaluation, the school administrative team held weekly meetings that provided a platform for them to create and integrate new practices that would support effective evaluation processes and could lead to increased teacher learning. The principal made the value of these sessions clear when he explained that “We think of different things everyday to make the process even easier than it was before.” Three highlights of the metaevaluative process were the creation of the online Google spreadsheet, the use of teacher binders, and the utilization of a guided reading coach to improve in a recognized area of challenge that was revealed through the teacher evaluation system.

Sustained leadership. Kotter’s (2012) conception of sustained leadership is a second lens through which to understand the teacher evaluation system at Elmwood School. Many researchers have proposed that initiating new programs requires a multi-year, intentional commitment to the change process (Abdal-Haqq, 1995; Feiman-Nemser, 2001; Fogarty & Pete, 2007; Little, 1988; McCann, et al., 2012; Reeves, 2009). Kotter (2012) outlines eight linear elements that support school change which are helpful in understanding the successes and limits of implementing the teacher evaluation system at Elmwood School. This analysis will describe elements 1, creating a sense of urgency, 2, forming a central coalition, 4, communicating the vision, and 8, anchoring changes in school culture. Overall, leadership thought through and initiated appropriate activities in each area, but were unable to sustain momentum. Throughout, I will also use Fullan’s (2005) conceptions of sustaining change to support the discussion of Kotter’s (2012) leadership model.

During the first step, Kotter calls on leaders to create a sense of urgency through which all stakeholders see a need for change. Fullan (2005) also believes that this should be a first step to change and describes that urgency should relate to “public service with moral purpose,”

through which educators serve in purposeful pursuits that truly enrich student lives. As described in Chapter Four, it was clear that the teachers at Elmwood School felt a strong sense of urgency towards improving student achievement. This sense of urgency arose from a desire to close the achievement gap for student in their high-poverty setting. One outcome of this urgency was that teachers consistently analyzed student data in order to drive rigorous instruction. They often communicated this through a unique strength of voice when they spoke about improvements in student data and a sense of pride related to student achievement on standardized tests.

In terms of teacher evaluation, the teacher evaluation system strongly aligned with the sense of urgency in Years 1 and 2 when it was well implemented. At that time, teachers felt that the teacher evaluation system helped them meet their student learning goals. However, in Year 3, the principal's continued passion about the importance of teacher evaluation did not translate into a driving mission for administrators or teachers. As a result, teachers seemed to believe that teacher evaluation was not an effective means to supporting high quality instruction for their students, although they continued to believe that high quality instruction was of utmost importance. For example, Kathy admitted "I really didn't look at [evaluation feedback] that much [in Year 3] ... because it was not really referenced that much at all in my final end of the year evaluative meeting. I spent a lot of time looking at [student test] score results, but I don't think that classifies as feedback." As this quote reiterates, this teacher understood that reviewing score results would allow her to serve the needs of her students, but felt that the teacher evaluation feedback was not important for her practice. She expected that her evaluator would help her interpret it in meaningful ways, and disregarded the written feedback when it was not supported by a verbal conversation. Although Andre had rallied teachers and leaders around his

vision for evaluation in previous years, stakeholders could not or did not want to partake in evaluation practices when they were no longer well-supported.

Kotter (2012) outlines formation of a central coalition as the second step to sustained school change. This element is also one that reveals strengths and limitations of the teacher evaluation system at Elmwood School. In Years 1 and 2, the administrative team built a home-grown evaluation system together. They attended training together, then worked together to apply their learning in a way that would work well for them. They continued to refine this process over two years, using a “metaevaluative” approach (Patton, 2012), as described above. In Year 3, the school had to hire new staff in three core positions, the elementary and middle school assistant principals, and the math coach. The principal assumed that accountability structures that were in place, like the online Google system, and practice opportunities like the weekly administrative meetings, would be enough to integrate new leaders into the evaluation system. Unfortunately, without attending initial (or participating in some replacement for it) and understanding the process with the same depth as the original core did, new leaders did not continue to develop teacher learning in the same ways as previous administrators had. Replacing members of the central coalition without similar training to what original members had experienced was extremely detrimental to the effective implementation of the teacher evaluation system because new leaders did not understand Elmwood School’s version of the initiative. Teachers did not fully understand the process either. For example, no teachers could describe The Six Steps of Feedback. Teacher evaluation seemed to be best understood by the principal. It seems that in addition to Kotter’s model of sustaining learning for a central coalition, the school needed to distribute understanding and leadership of the teacher evaluation system to many stakeholders (Fullan, 2005; Spillane, 2012). It may have been that with a more

standardized training protocol for new leaders and more distributed involvement in the teacher evaluation system, the teacher evaluation system at Elmwood School would have been less vulnerable to staffing changes. These are two areas of change that the principal has already planned for Year 4.

A central step in the change process is communication. Communication ensures that all stakeholders are aware of the process and potential products of change, which can increase stakeholder buy-in (Fullan, 2005; Kotter, 2012; Reeves, 2005). It also allows for the change vision to be fully understood and keeps momentum going for change. It seems that this was an area of need for the Elmwood School teacher evaluation system as well. As outlined above, Andre failed to successfully communicate the vision for teacher evaluation to new members of the administrative team and did not communicate key aspects of the system to all staff. In addition, administrators had not fully communicated to teachers the outcomes of how they were using the teacher evaluation system to drive professional learning for teachers. An important example was the use of the guided reading coach. Though both interviewed administrators spoke about choosing guided reading as an area of focus during their weekly meetings in order to meet a perceived need identified through classroom evaluations, teachers did not link the learning from the guided reading coach to larger evaluation structures at the school. Every teacher who taught guided reading referenced the guided reading coaching as instrumental to their learning. For example, Olivia described the guided reading coach as giving the best feedback, but in reference to feedback that she gets from the teacher evaluation system, she said “I guess it was decent. I mean it was okay.” In essence, she saw the guided reading coach as separate from her experience with the teacher evaluation system. Had the administrators been more communicative about how guided reading aligned with the vision of supporting instruction

through the formative teacher evaluation system, teachers might have recognized how teacher evaluation was impacting their daily instruction. Instead, teachers felt that nothing came out of “the actual feedback that they get out of these very long intricate rubrics they use to rank teachers.” The principal also recognized that he needed to do a better job of branding the teacher evaluation system by “reshaping the narrative of what observational feedback is” because teachers did not recognize “observational feedback could come from a coach, could come from an administrator, [or] it could come from a reading specialist.” Altogether, it seems that administration needs to spend a lot more time at all levels communicating about the process of teacher evaluation, its impacts on school-wide and individual outcomes, and how it connects to other school programs. More generally, the need for continued communication around evaluation was a significant indicator that the teacher evaluation change initiative had not yet become a sustained process, and administrators needed to continue building understanding and commitment.

Kotter (2012) recommends anchoring changes in organizational culture as a final step in creating sustained change initiatives. Fullan (2005) also maintains that addressing the environment within which the change occurs is necessary before organizations can adapt to change. As previously described, the school administration began this work by incorporating utilization-focused evaluation structures into the work of school leaders and classroom teachers. This was initially successful, as evidenced by teachers referring to the evaluation rubric as a lived document that was completely integrated into the learning culture of the school in Year 2. Individual uses of the formative evaluation system had become engrained in the school culture. However, the structures behind the formative evaluation system were not widely-understood. Primarily, the lack of administrative training in Year 3 led to uneven implementation, which

eroded the culture around formative teacher evaluation that had been built in previous years. In addition, teachers had only learned how to use their own feedback data with support, not independently. And, though teachers had been encouraged to observe each other's practice, they did not know about the administrative structures that were put in place to support the evaluation system such as the Six Steps of Feedback. In a more ideal scenario, teachers may have been able to support each other's use of feedback or supported each other's learning process through a peer-feedback process, had those elements of the program been more strongly developed in Years 1 and 2. The evaluation system was further compromised due to administrative actions at the district level. Central office contributed to overwhelming workloads for Elmwood's administrative staff by not hiring replacements for open positions, giving administrators "several hats" to wear, and calling last-minute meetings that often conflicted with observation and feedback time. Overall, this study confirms previous research about sustaining change and helps us to understand that institutionalization requires deep and long-term commitment to the change effort from local and district leadership and that leaders must continually train and communicate until change initiatives become anchored in school culture and knowledge and skill is well distributed throughout the building.

Research Question 2: In what ways do teachers use feedback from a formative teacher evaluation system implemented in their school to inform their practice?

The study's second research question seeks to understand how teachers use evaluation feedback to improve their instructional practices: Originally, I intended to analyze how feedback impacted the ways that teachers thought about their practice, planned for practice, and enacted practice. However, this question was derailed as teachers overwhelmingly reported that teacher

evaluation feedback had become extraneous to their thinking about teaching during Year 3. Though some teachers had been able to link the teacher evaluation system to their professional growth as described in Chapter Four, in all, there seemed to be little use of teacher evaluation feedback data during the study year. It may be that this was due to the mid-year personnel changes and limited administrative training during Year 3 or to a lack of clear communication about how administrators were using teacher evaluation data to increase teacher learning at the school. However, this may simply be because, as Nordin (2014) describes, there is an unclear connection between how feedback translates to practice, which makes describing this connection difficult for teachers and other school practitioners. Or, less teacher use of feedback might be attributable to a combination of all three reasons. Regardless, findings seem to indicate that, with the exception of teachers in need of extensive support, few teachers were given sustained opportunities for individualized professional growth through the teacher evaluation system in Year 3. Though many stakeholders used language to indicate a culture of learning for all (for example, Andre often referred to the teaching staff as “professional scholars”), the data indicate that as they struggled to keep up with their caseloads, leaders focused on only those teachers who were most in need of improvement. For example, Denise was told that her administrator “had bigger fires to put out” when she asked for more learning support. In general, teachers felt that other opportunities for learning would be more helpful in developing and improving their practice. The next section will describe how teachers reported needing more opportunities for collaborative conversations about practice with their peers.

Revision to the theoretical framework: Reevaluating the nature of teacher evaluation feedback. The theoretical framework of this study describes how most researchers agree on the need to create valid teacher evaluation tools and reliable procedures to sustain them,

but are divided about the best uses for the data that these tools generate. Some favor summative approaches through which administrators can sort teachers in terms of excellence in order to reward and punish them for the quality of their instructional practices. Others believe that teacher evaluation data is best used to help teachers learn about and improve their practice. Few studies argue that teacher evaluation should serve both functions, though many policy makers and school leaders favor both uses of data. This study adds to the available research by proposing that there are other possible uses of teacher evaluation data that may be more suited to the work of teachers. Specifically, I present what I term “collaborative uses” as a new approach to teacher evaluation that has implications for researchers and practitioners.

At the start of this study, my views aligned strongly with those that support formative approaches to teacher evaluation. My experiences and reading had led me to believe that formative teacher evaluation systems are the most important commitment that school leaders can make to support individual teacher growth because they provide embedded, continuous, and relevant feedback for practice. Yet, the summative-formative dichotomy primes schools to utilize teacher evaluation systems as a top-down data collection tool through which the administrator is central to the process and the teacher is the receiver of administrative knowledge. The summative-formative dichotomy also overlooks a larger construct in the literature on teacher learning that emphasizes the importance of creating opportunities for teachers to collaborate around and generate new professional knowledge. For example, Darling-Hammond (2013) argues “Although individual teacher evaluation can be a part of an educational improvement strategy, it cannot substitute for ongoing investments in the development and dissemination of profession-wide knowledge through ... work in professional learning communities” (p. 150). Although I had read this and other works urging for the inclusion of

teacher evaluation as one piece of a more collaborative approach to teacher learning several times before conducting this study, the importance of this assertion did not register until I saw it in my own data.

What seems to emerge from this study is that focusing only on one-on-one, top-down feedback to teachers can actually inhibit teacher growth. First, individualized sessions often privatize teacher challenges such that teachers are expected to rely only upon the administrator as a source of knowledge. This can be very limiting because it depends on the knowledge, experiences, and perspectives of that individual. For example, Denise, a veteran middle school teacher who had taught for ten years before moving to Elmwood, described how she felt comfortable with content and classroom management, but needed help understanding the new school culture. Yet, the administrator only gave her suggestions about classroom management and organization. In this case, her needs were not met by the administrator-to-teacher feedback flow. Although it is likely that other teachers would have been helpful to Denise's understanding of her own teaching practice, she was unable to access that knowledge due to a strong focus on individual feedback over collaborative approaches. Additionally, the administrator-to-teacher model seems to favor a singular approach for all teachers that was successful for some, but not others. In order for the administrator-to-teacher feedback flow to be effective, administrators need to be skilled in engaging and supporting all teachers in all aspects of practice.

When the administrator is positioned as the sole source of new professional knowledge, the administrator becomes the source of professional learning. This played out in two ways at Elmwood School. First, the administrator's perceived priorities more than those of the teachers were what drove her day-to-day decisions about how to allocate her time. For example, the administrator would focus her attention on the teachers which she believed to be most in need of

her support. This left many teachers, especially veterans, feeling as though the teacher evaluation system was not fulfilling their needs for constructive support. Denise illustrated how this left her out. She reported that the middle school AP, Michael, said, “I’m not worried about you.” Denise continued “but I was so worried about myself.” The system prioritized teachers that the administrator viewed as in more need of support, so Denise’s perceived needs for feedback could not be addressed. Emily, another veteran teacher, also shared a similar experience of her need for feedback not being met. She said, “I love feedback. I think that’s the only way that you can grow.... but, of course, administrators give their time and resources and professional development to those who are most in need.” This quote shows that, in such a system, the administrator is the one who chooses how to delegate time and resources and professional development opportunities, again centralizing the administrator in the process. Emily seems to describe resources and learning opportunities as belonging to administrators, an idea that will be explored in more detail below.

Second, administrator knowledge was given precedence in the administrator-to-teacher flow. One basic way this played out was that administrators were viewed as the givers of knowledge. Denise described how she wished that her administrator could just “Be listening to you, so that way you have someone who you can rely on who’s listening to what you’re saying... But I feel like I didn’t get that and I really needed it last year” (during Year 2, which was her first year at the school). More generally, teachers reported that feedback matched the perspective of the administrator. For example, Kathy explained that her administrator could not give her any feedback “content-wise” because her teaching area was outside of his area of expertise. As such, the administrator’s comfort level with various potential areas for feedback dictated the focus of feedback rather than the individual teacher’s learning needs. As discussed in Chapter Four,

Heather and Denise's experiences of the teacher evaluation system were very different although both were experienced teachers that started in Elmwood School during the same year (Year 2). Whereas Heather's needs were easily addressed by her administrator because they aligned with her administrator's background in early childhood reading instruction, Denise's needs were outside of her administrator's area of comfort, and, therefore, were not addressed.

Relatedly, when teacher learning was dictated by data available through the teacher evaluation tool, only the practices that were measured were viewed as important for improvement. In essence, accepted dialogue about teaching and teacher learning became limited to the categories that were the focus the teacher evaluation tool. The fact that the observation tool drove all feedback discussions was especially impactful for the learning of veteran teachers who often did not need support in the basic teaching practices that are highlighted there, including classroom management and general pedagogy. Instead they were in need of support with pedagogical content-knowledge. For example, Kathy described that when teaching difficult content, basic suggestions like "Anchor charts... won't help." Wendy similarly asserted "The rubric is too general to help me understand [the difficult pedagogical skill of helping students to find the main idea]."

When learning about teaching was dictated by the perspectives of the administrator and the confines of the evaluation rubric, teachers often viewed it as outside of their experiences, needs, and questions about teaching. They described feedback as if it belonged to the administrator and was at the administrator's discretion to create and disseminate. This practice may have contributed to teachers taking less ownership of the teacher evaluation system than might have been necessary for a more sustained process. For example, Kathy reported that she did not spend time looking at her feedback because it was "not really referenced that much in the

final end of the year evaluative meeting,” although she felt it was important to analyze data from her students’ standardized tests on her own. Her description indicates that she felt the feedback was only relevant when discussed by the administrator and that otherwise she should not be responsible for implementing it. The practice also contributed to teachers feeling like their own professional knowledge was not valued in the same ways as knowledge sanctioned by the evaluation tools or the administrators’ point of view. For example, Heather relayed how her administrator’s view of how to engage students was different from her own because his experience was in upper elementary school. She felt that with “the younger students, it’s very different.” Though she had more expertise with the particular student population than the administrator who was giving her feedback, was not asked to contribute to the conversation based on her own understandings of the student population. This example illustrates how teachers were rarely invited to participate in problem-solving, knowledge-sharing, and knowledge generation through the teacher evaluation process.

Altogether, feedback was an experience that was often viewed as flowing in just one direction, from the administrator, and often misaligned with the teachers’ immediate needs. Therefore, it is perhaps unsurprising that when asked about helpful feedback, many teachers discussed experiences with shared teacher learning rather than administrator-directed feedback, indicating that opportunities for collaborative learning were more supportive and impactful. For example, Olivia described:

The most effective one [feedback] that I can recall, wasn't [from an administrator] actually coming into the classroom and observing me...[Our department] used to... bring in work from that week ... and I would bring in my exit tickets and we would go over that data from those exit tickets. It was data day protocols or something like that... I

would give it to them and we would just go through different phases of it. So, first any clarifying questions. Then, I would be silent while they gave me “warm” feedback and “hard” feedback. ... So I really like that type of protocol, that type of feedback, that kind of system.

Here, Olivia describes the activities of a professional learning community (PLC) as providing the most salient feedback that she received at the school. Wendy and members of the early childhood focus group also talked about PLCs as most relevant to their learning. Teachers also spoke about other forms of collaboration as important to improving their practice, such as weekly planning meetings, observing other teachers, or simply having a colleague to “bounce ideas off.” Teachers report that these practices were good resources that allowed teachers to “utilize each other in a safe space” (Emily), reflect on and apply learning, and feel more confident about their own instruction. Unfortunately, teachers spoke about how adequate time had not been allocated to any of these practices and that many of the practices had ended as administrators focused on preparing for state tests. Though teachers recognized that collaboration was a promising learning experience, they felt unable to do this systematically on their own without dedicated time to do so. For example, Kathy explained that they had not engaged enough times in peer-to-peer learning activities “to develop a relationship with somebody that you could trust and open up to” and so doing this kind of teacher learning without formal structures and support would be difficult and unsustainable.

As described earlier, viewing teacher evaluation systems as successful when they are accompanied by collaborative practices is not a new finding in the research on teacher evaluation. However, most studies only focus on exploring the dichotomy of summative and

formative uses of data. Instead, this study confirms that “Evaluations should trigger continuous...opportunities to share expertise, as part of recognizing teachers' strengths and needs” (Darling-Hammond, 2013, p. 99). Many authors have proposed that teaching improves more through collegial than individualized efforts and that individualizing teacher learning leads to isolation and competitiveness which does not contribute to a culture of school as a learning organization (Darling-Hammond, 2013; Fullan, 2007; Moore-Johnson, 2015). Others find that important teacher qualities, like self-efficacy (Siciliano, 2016; Wilhelm et al., 2016) and self-reflection (Kremer-Hayon, 1993) increase with the help of colleagues and that school change is not possible when teachers work in isolation (Fullan, 2005).

This study adds to the available research. It reveals that by conceptualizing teaching practice as individual, rather than collective, administrators at Elmwood School were in control of teacher evaluation feedback and evaluation was a process that was driven solely by administrators, rather than one in which teachers wanted to engage (Patton, 2012). Though it was helpful and even transformational for individual teachers, at times, the evaluation system was not always viewed as central to teachers' learning. Yet all teachers believed that the practices of observing each other and providing feedback to each other were instrumental to their learning.

Thus, I propose that formative evaluation should include “collaborative uses” of teacher evaluation data. Though many have proposed that teacher learning should be situated in collaborative approaches to understanding teacher-determined areas of practice (Cochran-Smith & Lytle, 1999; Darling-Hammond & McLaughlin, 1995; Franke, et al., 2005; Little, 1988; McCann, Jones, & Aronoff, 2012; Wood, 2007), using teacher evaluation data in a collaborative way is a new approach that can enhance the value, power, and relevance of formative evaluation.

Collaborative uses of data could include peer observation and feedback, using data generated from administrative observations in teacher-led explorations of department- or school- level practices, or having teachers work together to generate collaborative improvement plans based on analysis of aggregated observation data. Through collaborative uses of teacher evaluation data, the focus would shift from an administrator-directed, one-way dialogue to include more collaborative and teacher-centered approaches to using teacher evaluation data to improve instructional practices. Focusing on individualized uses of data has led to teacher isolation, whereas collaborative approaches may help teachers feel more confident about their practice and more able to understand and respond to problems that arise in their practice. In addition, collaborative approaches have much greater potential to be sustained because they are not contingent upon the leadership of a few highly-trained staff members. Instead, knowledge is more wide-spread and so changes in personnel are less likely to cause major shifts in implementation. Most importantly, collaborative approaches to data use allow teachers to be central in the use of teacher evaluation data.

Conclusions

As described in Chapter Four and the discussion in this chapter, the data from this study reveal several important findings regarding initiating, sustaining, and maximally-utilizing formative teacher evaluation systems. I conclude by reviewing some missed opportunities that limited my study and providing some important implications for practice, policy, and research, including practice implications that the school has already begun to implement.

Missed Opportunities

The data provided an abundance of rich findings leading to many lessons. However, there were some missed opportunities to deepen my insight of Elmwood School's teacher

evaluation system. First, I was unable to collect enough data regarding my second research question about teacher learning, especially during Year 3. I originally hoped to understand how teacher evaluation feedback contributed to teachers' thinking about, planning for, and enacting practice. However, the data revealed that the uptake of feedback during Year 3 was not as robust as I had hoped, perhaps because the school had not effectively sustained practices related to the teacher evaluation system after Year 2. It seems that my data would have told a different story had I had the opportunity to collect it during Year 2. I also recognize that I lost sight of this original question when important descriptions regarding sustainability began to emerge. During the transcription process, I realized that during my first round of interviews, I kept trying to push teachers to describe their experiences with professional learning related to the evaluation system, even as teachers steered the conversation towards other issues like sustainability and the need for collaborative learning. Over time, I found myself becoming more interested in both concepts as well, and utilized much of my second round of interviews to try to understand differences between years of implementation and among teacher responses in relation to their years of experience as teachers and in the school. In all, I could have pushed the learning dialogue more or in better ways in order to use the case to facilitate a deeper understanding of teacher learning, however it seems that maybe there was not much potential for learning during Year 3. In either case, more research would be necessary to begin to describe teacher learning in relation to teacher evaluation feedback.

A second missed opportunity in my study was that I only interviewed school administrators; I should have included teacher leaders and coaches in my administrative sample to make my data richer. In addition, as I began data collection after both assistant principals had left the school, I did not have any data from them about their experiences with the training

process as new administrators in an existing system. The lack of these key staff perspectives left many gaps in my understanding of the process of learning how to give and support feedback and the experience of being a stakeholder in the formative teacher evaluation system. Instead, I had to piece together what I think are sound assessments of the process based on indirect information.

Outcomes of Research at Elmwood School

As a result of sharing findings from this research, the principal has committed to incorporating new structures to support the teacher evaluation system including training, differentiated teacher supports, and accountability processes. In terms of training, the principal recognized that there needs to be more consistency in the administrative training process for new administrators. Therefore, the school has dedicated a budget for initial administrative training. Teachers also lacked clear understandings of the formative teacher evaluation system, even though they were very aware of the evaluation criteria for job performance. So, in the summer before Year 4, the whole staff was introduced to the feedback process, including the Six Steps of Feedback. Department heads and Master Teachers were further trained on coaching techniques using this framework. The assumption is that when all staff members fully understand the process and possible outcomes of formative evaluation, teachers can utilize feedback more fully. Also, given feedback from this study, the school has incorporated more differentiated supports for teachers. First, the district dedicated time in the professional learning calendar to give teachers networking opportunities based on feedback about teachers' desire for more opportunities to collaborate. Last year, teachers were only afforded about four opportunities to meet with teachers from other schools, while this year fourteen collaborative meetings are scheduled. The school will also set up a more tiered approach to observation that is more

supportive of high-performing teachers, including staff who will be “responsible for pushing the thinking and learning and the practice of those mid-tier teachers” (Andre). The school had already planned to better develop the video catalogue to display teaching benchmarks. The school can also use its Google tracking system more effectively by tracking not only frequency of observations, but using it to follow progressions of teachers’ learning over time. By doing so, the principal can provide better feedback to leaders about their implementation. Finally, in terms of accountability, the principal has also committed to more frequent surveys of staff in the hopes that any system-level issues will come to his attention earlier.

Implications for Practice

Overall, this study affirms previous studies’ findings that institutionalization of teacher evaluation systems requires deep commitment and sustained practice at all organizational levels. To stimulate teacher learning through the teacher evaluation system, leaders are tasked with creating a vision for how they will purposefully link evaluation feedback to teacher learning. Then they must continually evaluate their progress in light of that vision, clearly communicating progress towards that vision, and initiating and aligning school programs and policies with that vision. Elmwood School provides many positive examples in these areas. For example, the school’s program illustrates the importance and potential for success through “metaevaluation” (Patton, 2012). By using home-grown policies and procedures to support the teacher evaluation system, Elmwood School was able to tailor teacher evaluation practices to their unique needs and community. Elmwood School’s use of video to support the teacher evaluation, though it is as yet a work in progress, also showed a lot of promise for supporting teacher learning. Teachers who had used videos of their own or others’ classrooms reported an increase in self-reflection and confidence. Findings also imply that comprehensive teacher evaluation systems necessarily

impact the school's climate, especially the elements of professional urgency, safety, efficacy, and buy-in and visa versa.

There were also some findings that provide recommendations for schools. First, there are many easily-implemented changes that can improve formative systems, such as developing a professional resource library that can be used to provide teachers with immediate examples of practice that can help them address challenges to reaching their goals identified during the evaluation process. In addition, to sustain change, multi-faceted programs will be necessary. For teachers, schools need to provide training on how to use feedback, continually communicate the outcomes of the evaluation program, provide differentiated supports for teachers across years of experience, increase teacher collaboration, and invest in mentoring programs. Leadership tasks need to be strategically distributed to make room for the difficult and time-consuming work of formative evaluation. Perhaps most importantly, principal induction programs are necessary when implementing complex formative systems. Principals need training on how to carry out program procedures, but they also need to understand the more nuanced and complex work of supporting teacher's growth over time. School leaders also structure the induction of new administrators by clearly communicating necessary roles and responsibilities. This work is critical because formative teacher evaluation systems are not traditionally implemented in most schools, and therefore will be a new approach for many newly hired administrators. At the district level, the case highlights the importance of central-leadership's commitment to the teacher evaluation system, without which the difficult and time-consuming work is unsustainable. By providing adequate administrative staffing, relieving administrators of extra-duties or restructuring job roles, and respecting the time commitments that administrators should devote to evaluation in order to be effective, central leadership can contribute to the success of

school-level leaders. Generally, schools that are moving towards more formative systems must recognize that such systems require many moving pieces. It is not just a matter of training, or just providing resources, or just hiring good leaders; instead it is critically important to create a synergistic system that is consistent and constantly reevaluating progress and improving procedures. For “turnaround” schools like Elmwood, this case study demonstrates that this work is even more important because changes in personnel can be a continuous challenge (Stuit & Smith, 2012).

Implications for Policy

As this study and others have argued, policy success is largely dependent on school implementation (Honig, 2006; Lipsky, 2010; Spillane & Kenney, 2012). As Race to the Top took hold at the state level, districts and schools had to find ways to incorporate, and sometimes circumnavigate, its many requirements. To make teacher evaluation policies more effective, policy makers should dedicate and sustain funding for strategies that may increase high quality implementation. First, this study indicates that initially learning how to observe and provide feedback on instruction is a difficult task that requires ongoing training and practice opportunities. States should consider adding coursework about this process to requirements for principal licensure. States can also enhance this learning by mandating extended study opportunities for school leaders to support each other in this key practice, whether through multi-year principal induction or principal professional learning networks. Second, states must address the demands on administrators’ time if leaders are to do this work thoroughly. This may include restructuring the role of principals, reevaluating the necessity of other time-consuming state requirements, or establishing another way to support the time principals need to implement formative systems.

Implications for Future Research

The study also raises many questions that can be addressed through future research. Its original intent was to use a case example of a school where formative teacher evaluation is being used effectively to uncover and describe the process by which teachers learn from feedback based on observations of their classrooms and implement those learnings in practice. The school where this study was conducted did not, in the end, provide a strong example of this process. Therefore, more research is warranted. Instead, the study uncovered the importance of a sustainability plan that has already impacted implementation at the school site. A follow up study or studies at Elmwood School that takes place in subsequent years when lessons learned are implemented could lead to other practical learnings and implications for policy and practice.

More generally, studies of positive case examples, especially longitudinal studies, can provide more guidance to schools and researchers seeking to improve teacher evaluation implementation. However, as noted in Chapter Three, finding a research site may be a very difficult process. For this study, I was unable to find a site that was implementing positive formative evaluation practices in NJ, and only found a school in Washington D.C. by coincidence. As this study further shows, even schools that are purposefully implementing good practices are often unable to sustain them. Given the difficulty of finding research sites, studying them is all the more important. This may lead to better understandings about why such structures are so rare in schools and how they can become more commonplace. The goals of this research might also be to develop and test the impact of evaluation tools and practices that support teacher learning. For example, at Elmwood School, professional tracking systems like binders and the online Google tool supported the implementation process. More research on these and other support tools can be of use to practitioners.

This study also uncovers the importance of collaborative uses of teacher evaluation data. Though some researchers (see for example, Darling-Hammond, 2013) make a case for the importance of collaborative approaches to teacher evaluation, few describe actual systems that employ them effectively. Additionally, more research is needed to understand how principals hone the craft of giving feedback and supporting teachers, and the different models for doing so, such as top-down and collaborative approaches. With better understandings, these qualities can be replicated in principal preparation programs. One important practice at Elmwood School was weekly administrative calibration sessions. Although there is now a considerable research base on teacher professional learning communities (see for example: DuFour & Eaker, 2005) and video study groups (see for example: Sherin & van Es, 2005), the process of collaborative learning for leadership is not well documented. Finally, though some researchers have described how the way in which the coach's role is framed can help or inhibit the effectiveness of instructional coaching (see for example: Mangin & Dunsmore, 2015), this study reveals that such framing is also necessary for school leaders' work, especially when trying to establish complex and comprehensive systems such as formative teacher evaluation. More research about the ways in which this is successfully done can enhance policy and practice.

Personal Reflection

This research effort began with a strong belief that teacher evaluation is the most important way that leaders can initiate individual teacher learning at the school level. Because feedback from observations of teacher practice has potential for individualized and relevant learning for teachers, I believed that a focus on teacher evaluation would strongly impact professional practice. That belief led me to a wonderful school, staffed by passionate, hard-working teachers and administrators, where teacher evaluation was used to serve teacher learning

needs. However, through my study of the school, I learned that teacher evaluation systems are very difficult to sustain. I also learned that, even when sustained, teacher evaluation systems are not enough to activate and continually support every teacher's learning needs. When leaders make a commitment to providing formative teacher evaluation, they must also commit to embedding the evaluation system into school culture by providing ongoing training to teachers and leaders, dedicating time and resources to the evaluation process, continually calibrating their own practice, and frequently revising programs, policies, and procedures that support teacher evaluation practices. The work is arduous and interminable, yet, as positive examples from the school showed, worthwhile for impacting teacher, and thus, student growth.

This study also revealed a finding that was transformative for my own belief system about evaluation. Although I once believed that personalized feedback could increase teachers' self-reflection and self-efficacy, I was surprised to learn that a singular focus on conversations about individual classrooms may actually contribute to teacher isolation and may limit what teachers can learn from teacher evaluation feedback. Data revealed that Elmwood School's implementation led to heightened focus on individualized feedback which further privatized teaching and privileged administrator perspectives and the instructional values encoded in the feedback tool. Some unanticipated outcomes were that teachers, especially those who were proficient or highly-proficient, did not feel that they were encouraged to grow and develop in meaningful ways. Relatedly, the limits of the teacher evaluation tool and of administrator time and knowledge contributed to a triage model of teacher professional development, whereby only teachers in need of extensive or obvious support were well-served. As a result, this study fundamentally contributes to the available literature on teacher evaluation because it challenges the idea that teacher evaluation can only be used by administrators to summatively assess

practice or to provide formative feedback. Available research creates a dichotomy that seems to force school leaders to adopt a fixed stance towards individual uses of data, however I propose that researchers should also include another use of teacher evaluation data that is not focused on the individual teacher, but is used to drive the work of collaborative teacher learning. These types of systems may also be more sustainable and cost-effective because they allow all teachers to contribute to and gain from conversations about practice and are not contingent on centralized staff.

References

- Abdal-Haqq, I. (1995). Making time for teacher professional development. Washington, D.C.: *ERIC Clearinghouse on Teaching and Teacher Education (Digest 95-4)*.
- Aldeman, C., & Chuong, C. (2014). *Teacher evaluations in an era of rapid change: From “unsatisfactory” to “needs improvement.”* Washington, D.C.: Bellwether Education Partners.
- Almy, S. (2011). *Fair to everyone: Building the balanced teacher evaluations that educators and students deserve. Teacher quality.* Washington, D.C.: 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, D.C.: National Comprehensive Center for Teacher Quality.
- Baker, B. D., Oluwole, J., & Green, P. C. (2013). The legal consequences of mandating high stakes decisions based on low quality information: Teacher evaluation in the Race-to-the-Top era. *Education Evaluation and Policy Analysis Archives*, 21(5), 1-71.
- Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., ... & Shepard, L. A. (2010). *Problems with the use of student test scores to evaluate teachers. EPI Briefing Paper# 278.* Washington, D.C.: Economic Policy Institute.
- 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. (2012). Beyond the scoreboard. *Educational Leadership*, 70(3), 26-30.

- Barr, A. S., & Burton, W. H. (1926). *The supervision of instruction: A general volume*. New York, NY: D. Appleton.
- Baxter, P., & Jack, S. (2008). Qualitative case study methodology: Study design and implementation for novice researchers. *The Qualitative Report*, 13(4), 544-559.
- Beerens, D. R. (2000). *Evaluating teachers for professional growth: Creating a culture of motivation and learning*. Thousand Oaks, CA: Corwin Press, Inc.
- 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.
- Bell, C. A., Qi, Y., Croft, A. J., Leusner, D., Mccaffrey, D. F., Gitomer, D. H., & Pianta, R. C. (2014). In Kane, T., Kerr, K., & Pianta, R. (Eds.), *Designing teacher evaluation systems: New guidance from the measures of effective teaching project* (pp. 50-97). Somerset, NJ: John Wiley & Sons.
- 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.
- Bill & Melinda Gates Foundation. (2013). *Feedback for better teaching. Nine principles for using measures of effective teaching*. Seattle, WA: Bill & Melinda Gates Foundation.
- Bogdan, R. C., & Biklen, S. K. (1998). *Qualitative research in education. An introduction to theory and methods*. Needham Heights, MA: Allyn & Bacon, A Viacom Company.
- Bolton, D. L. (1973). *Selection and evaluation of teachers*. Berkeley, CA: McCutchan Publishing Corporation.

- Bonavitacola, A. C. (2014). *Teachers' perceptions of the impact of the McREL Teacher Evaluation System on professional growth*. (Doctoral dissertation). Capella University, Minneapolis, MN.
- Boser, U. (2012). *Race to the Top: What have we learned from the states so far?* Washington, D.C.: Center for American Progress.
- Brophy, J. (1986). Teacher influences on student achievement. *American Psychologist*, 41(10), 1069-1071.
- Bryk, A. S. (2010). Organizing schools for improvement. *Phi Delta Kappan*, 91(7), 23.
- Bryk, A. S., Sebring, P. B., Allensworth, E., Easton, J. Q., & Luppescu, S. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago, IL: University of Chicago Press.
- Callahan, K., & Sadeghi, L. (2014). *TEACHNJ: An evaluation of two years implementation*. Miami, FL: Florida International University.
- Camburn, E. M., & Han, S. W. (2015). Infrastructure for teacher reflection and instructional change: An exploratory study. *Journal of Educational Change*, 16(4), 511-533.
- Casabianca, J. M., Lockwood, J. R., & McCaffrey, D. F. (2015). Trends in classroom observation scores. *Educational and Psychological Measurement*, 75(2), 311-337.
- Casabianca, J. M., McCaffrey, D. F., Gitomer, D. H., Bell, C. A., Hamre, B. K., & Pianta, R. C. (2013). Effect of observation mode on measures of secondary mathematics teaching. *Educational and Psychological Measurement*, 73(5), 757-783.
- 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.

- Chaplin, D., Gill, B., Thompkins, A., & Miller, H. (2014). *Professional practice, student surveys, and value-added: Multiple measures of teacher effectiveness in the Pittsburgh public schools*. Washington, D.C.: Mathematica Policy Research.
- Chappuis, S., & Chappuis, J. (2007). The best value in formative assessment. *Educational Leadership*, 65(4), 14-19
- Charalambous, C. Y., Komitis, A., Papacharalambous, M., & Stefanou, A. (2014). Using generic and content-specific teaching practices in teacher evaluation: An exploratory study of teachers' perceptions. *Teaching and Teacher Education*, 41, 22-33.
- Charters, W. W., & Waples, D. (1929). *The commonwealth teacher-training study, Volume 9*. Chicago, IL: University of Chicago Press.
- Clayton, C. (2013). *Understanding current reforms to evaluate teachers: A literature review on teacher evaluation across the career span*. New York, NY: Pace University.
- Clipa, O. (2015). Roles and strategies of teacher evaluation: Teachers' perceptions. *Procedia-Social and Behavioral Sciences*, 180, 916-923.
- Cochran-Smith, M. (2010). Foreword. In M. Kennedy (Ed.), *Teacher assessment and the quest for teacher quality: A handbook*. (xiii-xix). San Francisco, CA: Jossey-Bass.
- Cochran-Smith, M., & Lytle, S. L. (1999). Relationships of knowledge and practice: Teacher learning in communities. *Review of Research in Education*, 249-305.
- Coggshall, 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, D.C.: National Comprehensive Center for Teacher Quality.

- Cohen, D. K. (2010). Teaching quality: An American educational dilemma. In M. Kennedy (Ed.), *Teacher assessment and the quest for teacher quality: A handbook*. (pp. 375-401). San Francisco, CA: Jossey-Bass.
- Cohen, D. K., & Moffitt, S. L. (2010). *The ordeal of equality: Did federal regulation fix the schools?* Cambridge, MA: Harvard University Press.
- Cohen, D.K., Raudenbush, S.W., & Ball, D.L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis*, 25(2), 119–142.
- Cohen, J., & Goldhaber, D. (2016). Building a more complete understanding of teacher evaluation using classroom observations. *Educational Researcher*, 45(6), 378-387.
- Coimbra, M. (2013) Supervision and evaluation: Teachers' perspectives. *International Journal of Humanities and Social Science*, 3(5), 65-71.
- Collins, A. B. (2004). Teacher performance evaluation: a stressful experience from a private secondary school. *Educational Research*, 46(1), 43-54.
- Conley, S., Smith, J. L., Collinson, V., & Palazuelos, A. (2014). A small step into the complexity of teacher evaluation as professional development. *Professional Development in Education*, 42(1), 168-170.
- Contreras, G. (1999). Teachers' perceptions of active participation, evaluation effectiveness, and training in evaluation systems. *Journal of Research and Development in Education*, 33(1), 47-59.
- Cook, M. A., & Richards, H. C. (1972). Dimensions of principal and supervisor ratings of teacher behavior. *The Journal of Experimental Educational*, 41(2), 11-14.
- Cor, K. (2011). *The measurement properties of the PLATO rubric*. Paper presented at the American Educational Research Association annual meeting, New Orleans, LA.

- Costa, A. L., & Garmston, R. J. (1994). *Cognitive coaching: A foundation for renaissance schools*. Norwood, MA: Christopher-Gordon Publishers
- Curby, T. W., Stuhlman, M., Grimm, K., Mashburn, A., Chomat-Mooney, L., Downer, J., ... & Pianta, R. C. (2011). Within-day variability in the quality of classroom interactions during third and fifth grade: Implications for children's experiences and conducting classroom observations. *The Elementary School Journal*, 112(1), 16-37.
- Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five designs*. Los Angeles, CA: Sage Publications.
- Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. Los Angeles, CA: Sage Publications.
- Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into Practice*, 39(3), 124-130.
- 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. (2001). New trends in teacher evaluation. *Educational Leadership*, 58(5), 12-15.
- 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. (2016). It's time to rethink teacher evaluation. *Education Week*, 35(28), 20-24.
- Danielson, C., & McGreal, T. L. (2000). *Teacher evaluation to enhance professional practice*. Alexandria, VA: Association of Supervision and Curriculum Development.

- Danielson Group. (2013). *The Danielson Group: Promoting effective teaching and professional learning*. Retrieved from <https://www.danielsongroup.org/>.
- Darling-Hammond, L. (2000). Teacher quality and student achievement. *Education Policy Analysis Archives*, 8(1), 1-44.
- 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. (2013). *Getting teacher evaluation right: What really matters for effectiveness and improvement*. New York, NY: Teachers College Press.
- Darling-Hammond, L. (2014). One piece of the whole: Teacher evaluation as part of a comprehensive system for teaching and learning. *American Educator*, 38(1), 4-13.
- Darling-Hammond, L. (2015a). Can value added add value to teacher evaluation? *Educational Researcher*, 44(2), 132-137.
- Darling-Hammond, L. (2015b). *The flat world and education: How America's commitment to equity will determine our future*. New York, NY: Teachers College Press.
- 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).
- Darling-Hammond, L., & McLaughlin, M. W. (1995). Policies that support professional development in an era of reform. *Phi Delta Kappan*, 76(8), 597-604.

- Darling-Hammond, L., Wise, A. E., & Pease, S. R. (1983). Teacher evaluation in the organizational context: A review of the literature. *Review of Educational Research*, 53(3), 285-328.
- 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.
- Dee, T., & Wyckoff, J. (2013). Incentives, selection, and teacher performance: Evidence from IMPACT. *Journal of Policy Analysis and Management*, 34, 267-297.
- Delvaux, E., Vanhoof, J., Tuytens, M., Vekeman, E., Devos, G., & Van Petegem, P. (2013). How may teacher evaluation have an impact on professional development? A multilevel analysis. *Teaching and Teacher Education*, 36, 1-11.
- DeMatthews, D. (2015). Getting teacher evaluation right: What principals need to know. *The Educational Forum*, 79(1), 81-89.
- 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.
- Derrington, M. L., Campbell, J. W., Schools, A. C., & Alcoa, T. N. (2015). Principal concerns and superintendent support during teacher evaluation changes. *Editorial Review Board*, 12(3), 11-22.
- Dewey, J. (1933). *How we think: A restatement of the reflective thinking to the educative process*. Lexington, MA: Heath.

- Doherty, K. M., & Jacobs, S. (2015). *State of the states 2015: Evaluating teaching, leading and learning*. Washington, D.C.: National Center on Teacher Quality. Retrieved from: <http://www.Nctq.org/dmsView/StateofStates2015>
- Donaldson, M. L. (2009). *So long, Lake Wobegon. Using teacher evaluation to raise teacher quality*. Washington, D.C.: Center for American Progress.
- Donaldson, M. L. (2012). *Teachers' perspectives on evaluation reform*. Washington, D.C.: Center for American Progress.
- Donaldson, M. L. (2013). Principals' approaches to cultivating teacher effectiveness constraints and opportunities in hiring, assigning, evaluating, and developing teachers. *Educational Administration Quarterly*, 49(5), 838-882.
- Donaldson, M. L., Cobb, C., LeChasseur, K., Gabriel, R., Gonzales, R., Woulfin, S., & Makuch, A. (2014). *An evaluation of the pilot implementation of Connecticut's system for educator evaluation and development*. Hartford, CT: Connecticut Department of Education.
- 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.
- DuFour, R., & Eaker, R. (2005). *Professional learning communities at work: Best practices for enhancing students' achievement*. Bloomington, IN: Solution Tree Press.
- DuFour, R., & Marzano, R. J. (2009). High-leverage strategies for principal leadership. *Educational Leadership*, 66(5), 62-68.

- Dynarski, M. (2016). *Teacher observations have been a waste of time and money*. Washington, D.C.: Brookings Institute.
- 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* (405-431). New York, NY: Cambridge University Press.
- Evans, B. R., Wills, F., & Moretti, M. (2015). Editor and section editor's perspective article: A look at the Danielson Framework for teacher evaluation. *Journal of the National Association for Alternative Certification*, 10(1), 21-26.
- Feiman-Nemser, S. (2001). From preparation to practice: Designing a continuum to strengthen and sustain teaching. *The Teachers College Record*, 103(6), 1013-1055.
- Firestone, W. A., Blitz, C. L., Gitomer, D. H., Kirova, D., & Shcherbakov, A. (2013). *New Jersey teacher evaluation, RU-GSE external assessment, Year 1 Report*. New Brunswick, NJ: Rutgers University.
- Fisher, S. L., & Ford, J. K. (1998). Differential effects of learner effort and goal orientation on two learning outcomes. *Personnel Psychology*, 51(2), 397.
- Fogarty, R. J., & Pete, B. M. (2007). *The adult learner: Some things we know*. Thousand Oaks, CA: Corwin Press.
- 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.

- Franke, M., Kazemi, E., Shih, J., Biagetti, S., & Battey, D. (2005). Changing teachers' professional work in mathematics: One school's journey. *Understanding Mathematics and Science Matters*, 209-229.
- 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. (2005). *Leadership & sustainability: System thinkers in action*. Thousand Oaks, CA: Corwin Press.
- 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.
- Gabriel, R. & Allington, R. (2012). The MET Project: The wrong 45 million dollar question. *Educational Leadership*, 70(3), 44-49.
- Gage, N. L. & Needels, M. C. (1989). Process-product research on teaching: A review of criticisms. *The Elementary School Journal*, 89(3), 253-300.
- Gall, M. D., Gall, J., & Borg, W. (2013). *Applying educational research: How to read, do, and use research to solve problems of practice*. Upper Saddle River, NJ: Pearson Higher Ed.
- Gallagher, H. A. (2004). Vaughn Elementary's innovative teacher evaluation system: Are teacher evaluation scores related to growth in student achievement? *Peabody Journal of Education*, 79(4), 79-107.
- Gardner, D. P. (1983). *A nation at risk*. Washington, D.C.: National Commission on Excellence in Education.

- Garrett, R., & Steinberg, M. (2015). Examining teacher effectiveness using classroom observationscores: Evidence from the randomization of teachers to students. *Educational Evaluation and Policy Analysis*, 37, 224–242.
- 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, D.C.: 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: Chicago, IL.
- Gitomer, D. H. (2008). Crisp measurement and messy context. In D. Gitomer (Ed.), *Measurement issues and assessment for teaching quality* (223-233). Thousand Oaks, CA: Sage Publications.
- Gitomer, D. H., Bell, C. A., Qi, Y., McCaffrey, D. F., Hamre, B. K., & Pianta, R. C. (2014). The instructional challenge in improving teaching quality: Lessons from a classroom observation protocol. *Teachers College Record*, 116, 1-32.
- Glasman, N. S., & Paulin, P. J. (1982). Possible determinants of teacher receptivity to evaluation. *Journal of Educational Administration*, 20(2), 148-171.
- Glazerman, S., Goldhaber, D., Loeb, S., Raudenbush, S., Staiger, D. O., Whitehurst, G. J., & Croft, M. (2011). *Passing muster: Evaluating teacher evaluation systems*. Washington, D.C.: Brookings Institute.
- Goe, L. (2013). Can teacher evaluation improve teaching? *Principal Leadership*, 13(7), 24-29.

- Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis*. Washington, D.C.: National Comprehensive Center for Teacher Quality.
- Goe, L., Biggers, K., & Croft, A. (2012). *Linking teacher evaluation to professional development: Focusing on improving teaching and learning. Research & policy brief*. Washington, D.C.: National Comprehensive Center for Teacher Quality.
- Goe, L., Holdheide, L., & Miller, T. (2011). *A practical guide to designing comprehensive teacher evaluation systems: A tool to assist in the development of teacher evaluation systems*. Washington, D.C.: National Comprehensive Center for Teacher Quality.
- Gold, Y. (1996). Beginning teacher support: Attrition, mentoring, and induction. In J. Sikula (Ed.), *Handbook of research on teacher education* (pp. 548-594). New York, NY: Simon & Schuster Macmillan.
- Goldhaber, D. 2002. The mystery of good teaching. *Education Next* 2(1):50–55
- Goldhaber, D. & Brown, N. (in press). Teacher policy under ESEA and HEAL A convergent trajectory with an unclear future. In C. P. Loss & P. J. McGuinn (Eds.), *The convergence of K-12 and higher education: Policies and programs in a changing era*. Cambridge, MA: Harvard Education Press
- Goldhammer, R. (1969). *Clinical supervision*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Gordon, R. J., Kane, T. J., & Staiger, D. (2006). *Identifying effective teachers using performance on the job*. Washington, D.C.: Brookings Institute.
- 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.
- Grossman, P., Cohen, J., & Brown, L. (2014). Understanding instructional quality in English Language Arts: Variations in PLATO scores by content and context. In T. Kane, K. Kerr, & R. Pianta (Eds.) *Designing teacher evaluation systems: New guidance from the measures of effective teaching project*. (pp. 127-164). Somerset, NJ: John Wiley & Sons, pp. 303-331.
- Grossman, P., Loeb, S., Cohen, J., Hammerness, K., Wyckoff, J., Boyd, D., & Lankford, H. (2010). *Measure for measure: The relationship between measures of instructional practice in middle school English language arts and teachers' value-added scores*. New York, NY: National Bureau of Economic Research.
- Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching*, 8(3), 381-391.
- Haefele, D. L. (1993). Evaluating teachers: A call for change. *Journal of Personnel Evaluation in Education*, 7(1), 21-31.
- Hall, G. E., & Hord, S. M. (2001). *Implementing change. Patterns, Principles, and Potholes*. Upper Saddle River, NJ: Pearson.
- Hallinger, P., Heck, R. H., & Murphy, J. (2014). Teacher evaluation and school improvement: An analysis of the evidence. *Educational Assessment, Evaluation and Accountability*, 26(1), 5-28.
- 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.
- Hanna, R. (2013). *Common strategies for uncommon achievement: How districts enable and support high-performing schools*. Center for American Progress: Washington, D.C..
- Hanushek, E. A. (2010). *The economic value of higher teacher quality* (NBER Working Paper No. 16606). Cambridge, MA: National Bureau of Economic Research.
- 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
- Hargreaves, A. (2005). Educational change takes ages: Life, career and generational factors in teachers' emotional responses to educational change. *Teaching and Teacher Education*, 21(8), 967-983.
- Harris, D. N., Ingle, W. K., & Rutledge, S. A. (2014). How teacher evaluation methods matter for accountability a comparative analysis of teacher effectiveness ratings by principals and teacher value-added measures. *American Educational Research Journal*, 51(1), 73-112.
- Harvey, J., & Holland, H. (2013). *The school principal as leader: Guiding schools to better teaching and learning*. New York: Wallace Foundation.

- Hatch, T., Hill, K., & Roegman, R. (2016). Investigating the Role of Instructional Rounds in the Development of Social Networks and District-Wide Improvement. *American Educational Research Journal*, 53(4), 1022-1053.
- 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.
- Hazi, H. M., & Rucinski, D. A. (2009). Teacher evaluation as a policy target for improved student learning: A fifty-state review of statute and regulatory action since NCLB. *Education Policy Analysis Archives*, 17(5), 1-22.
- 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, Chicago, IL.
- Hill, H. C., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough: teacher observation systems and a case for the generalizability study. *Educational Researcher*, 41(2), 56-64.
- Hill, H. C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, 48(3), 794-831.
- 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.

- Ho, A. D., & Kane, T. J. (2013). *The reliability of classroom observations by school personnel. Research Paper*. Seattle, WA: Bill & Melinda Gates Foundation.
- 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.
- Honig, G. (2006). *New directions in education policy implementation: Confronting complexity*. Albany, NY: State University of New York Press.
- Horng, E. L., Klasik, D., & Loeb, S. (2010). Principal's time use and school effectiveness. *American Journal of Education*, 116(4), 491-523.
- 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* (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.
- Hunter, M. C. (1982). *Mastery teaching*. El Segundo, CA: Tip Publications.
- Hunter, M. C., & Russell, D. (1995). *Mastering coaching and supervision*. Thousand Oaks, CA: Corwin Press.
- Ilggen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology*, 64(4), 349.
- Ing, M. (2010). Using informal classroom observations to improve instruction. *Journal of Educational Administration*, 48(3), 337-358.

- 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.
- Jiang, J. Y., Spote, S. E., & Luppescu, S. (2015). Teacher perspectives on evaluation reform Chicago's REACH students. *Educational Researcher*, 44(2), 105-116.
- Joe, J.N., McClellan, C.A., & Holtzman, S.L. (2014). Reliability and the length and focus of classroom observations. In Kane, T., Kerr, K., & Pianta, R. (Eds.) *Designing teacher evaluation systems: New guidance from the measures of effective teaching project* (pp. 415-443). Somerset, NJ: John Wiley & Sons.
- Joe, J. N., Tocci, C. M., Holtzman, S. L., & Williams, J. C. (2013). *Foundations of observation: Considerations for developing a classroom observation system that helps districts achieve consistent and accurate scores. MET Project Policy and Practice Brief*. Seattle, WA: Bill and Melinda Gates Foundation.
- Johnson, S. (1984). Merit pay for teachers: A poor prescription for reform. *Harvard Educational Review*, 54(2), 175-186.
- Johnson, S. M. (1990). *Teachers at work: Achieving success in our schools*. New York, NY: Basic Books.
- K12 Insight (2016). *K12 Insight: About Us*. Retrieved from <http://www.k12insight.com/>.

- Kane, T., & Cantrell, S. (2011). *Learning about teaching: Initial findings from the Measures of Effective Teaching project*. Seattle, WA: Bill & Melinda Gates Foundation.
- Kane, T., Kerr, K., & Pianta, R. (2014a). *Designing teacher evaluation systems: New guidance from the Measures of Effective Teaching Study*. San Francisco, CA: Jossey-Bass.
- Kane, T., Kerr, K., & Pianta, R. (2014). Why measure effective teaching? In Kane, T., Kerr, K., & Pianta, R. (Eds.) *Designing teacher evaluation systems: New guidance from the measures of effective teaching project* (pp. 1-6). Somerset, NJ: John Wiley & Sons.
- Kane, T., & 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.
- Kennedy, M. (2010a) Approaches to annual performance assessment. In M. Kennedy (Ed.), *Teacher assessment and the quest for teacher quality: A handbook* (pp. 225-249). San Francisco, CA: Josey-Bass.
- Kennedy, M. (2010b). Attribution error and the quest for teacher quality. *Educational Researcher*, 39(8), 591-598.

- Kennedy, M. (2010c). The uncertain relationship between teacher assessment and teacher quality. In M. Kennedy (Ed.), *Teacher assessment and the quest for teacher quality: A handbook* (pp. 1-6). San Francisco, CA: Josey-Bass.
- Kennedy, M. M. (2008). Sorting out teacher quality. *Phi Delta Kappan*, 90(1), 59-63.
- 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.
- Kimball, S. M., White, B., Milanowski, A. T., & Borman, G. (2004). Examining the relationship between teacher evaluation and student assessment results in Washoe County. *Peabody Journal of Education*, 79(4), 54-78.
- Koedel, C., Li, J., Springer, M. G., & Tan, L. (2014). *The impact of performance ratings on job satisfaction for public school teachers*. Working Paper: Tennessee Consortium on Research, Evaluation & Development.
- Kotter, J. P. (2012). *Leading change* (New ed.). Boston, MA: Harvard Business Review.
- Kotter, J. P. (2012). *A sense of urgency*. Boston, MA: Harvard Business Press.
- Kotter, J. P., & Cohen, D. S. (2002). *The heart of change: Real-life stories of how people change their organizations*. Boston, MA: Harvard Business Press.
- Kowalski, T. J., & Dolph, D. A. (2015). Principal dispositions regarding the Ohio teacher evaluation system. *AASA Journal of Scholarship and Practice*, 11(4), 4-20.
- Kraft, M. A., & Gilmour, A. (2015). *Can evaluation promote teacher development? Principals' views and experiences implementing observation and feedback cycles*. Working Paper Cambridge, MA: Harvard University.

- Kraft, M. A., & Gilmour, A. F. (2016a). Can principals promote teacher development as evaluators? A case study of principals' views and experiences. *Educational Administration Quarterly*, 52(5), 711-753.
- Kraft, M.A. & Gilmour, A.F. (2016b). *Revisiting the Widget Effect: Teacher evaluation reforms and the distribution of teacher effectiveness. Working Paper*. Providence, RI: Brown University.
- Kremer-Hayon, L. (1993). *Teacher self-evaluation: teachers in their own mirror*. New York, NY: Springer.
- Kuzmic, J. J. (2002). Research as a way of knowing and seeing: Advocacy for the other. In J. Loughran & T. Russell (Eds.), *Improving teacher education practices through self-study*, (pp. 222-235). New York, NY: RoutledgeFalmer.
- Ladson-Billings, G. (2009). Opportunity to teach. In D. Gitomer (Ed.), *Measurement issues and assessment for teaching quality* (pp. 206-213). Thousand Oaks, CA: Sage Publications.
- LaRocco, D. J., Wheeler, D. S., & Sennett, B. A. L. (2015). Connecticut's new teacher evaluation process: The concerns and experiences of educators in two pilot districts. *Current Issues in Education*, 18(1), 1-19.
- LeBoeuf, M. (1998). *The greatest management principle in the world*. Mumbai, India: Jaico Publishing House.
- Legg, K. C. (2015). *The experiences of teachers and administrators with a new teacher evaluation system in a local school division*. (Doctoral dissertation). VirginiaTech University: Blacksburg, VA.

- 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
- Lipsky, M. (2010). *Street-level bureaucracy, 30th Ann. Ed.: Dilemmas of the individual in public service*. New York, NY: Russell Sage Foundation.
- Little, J. W. (1988). Seductive images and organizational realities in professional development. In A. Lieberman, *Rethinking school improvement* (27-53). New York, NY: Teachers College Press.
- Looney, J. (2011). Developing high-quality teachers: Teacher evaluation for improvement. *European Journal of Education*, 46(4), 440-455.
- Louis, K. S., Leithwood, K., Wahlstrom, K., & Anderson, S. (2010). *Investigating the links to improved student learning: Final report of research findings*. New York, NY: The Wallace Foundation:
- Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science*, 5(4), 333-369.
- Mangin, M. M., & Dunsmore, K. (2015). How the framing of instructional coaching as a lever for systemic or individual reform influences the enactment of coaching. *Educational Administration Quarterly*, 51(2), 179-213.
- Marcos, A., Machado, E., & Abelha, M. (2015). Effect (s) of teacher evaluation on collaborative practices: induction or inhibition? *Procedia-Social and Behavioral Sciences*, 174, 3674-3680.
- Marshall, K. (2005). It's time to rethink teacher supervision and evaluation. *Phi Delta Kappan*, 86(10), 727-735.

- 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.
- Martinez, F., Taut, S., & Schaaf, K. (2016). Classroom observation for evaluating and improving teaching: An international perspective. *Studies in Educational Evaluation*, 49, 15-29.
- Marzano, R. J. (2003). *What works in schools: Translating research into action*. Alexandria, VA: Association of Supervision and Curriculum Development.
- Marzano, R. J. (2012). The two purposes of teacher evaluation. *Educational Leadership*, 70(3), 14-19.
- Marzano, R. J., Boogren, T., Heflebower, T., Kanold-McIntyre, J., & Pickering, D. (2012). *Becoming a reflective teacher*. Bloomington, IN: Marzano Research Laboratory.
- Marzano, R. J., Frontier, T., & Livingston, D. (2011). *Effective supervision: Supporting the art and science of teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Maslow, V. & Kelley, C. (2005, April). *Effects of teacher evaluation on equity and system changes in large diverse high schools*. Paper presented at the annual meeting of the America Education Research Association, Montreal, Canada.
- Maslow, V. J., & Kelley, C. J. (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(3), 600-632.
- McCann, T. M., Jones, A. C., & Aronoff, G. A. (2012). *Teaching matters most: A school leader's guide to improving classroom instruction*. Thousand Oaks, CA: SAGE Publications.

- McCreight, C. (2000). *Teacher attrition, shortage, and strategies for teacher retention*. College Station, TX: Texas A & M University
- McGreal, T. L. (1983). *Successful teacher evaluation*. Alexandria, VA: Association of Supervision and Curriculum Development.
- McGuinn, P. (2012). *The state of teacher evaluation reform: State education agency capacity and the implementation of new teacher-evaluation systems*. Washington, D.C.: Center for American Progress.
- McKay, S. & Silva, E. (2015). *Improving observer training: the trends and challenges. Issue Brief*. Stanford, CA: Carnegie Foundation for the Advancement of Teaching.
- McMillan, J. H., & Schumacher, S. (1997). *Research in education: A conceptual approach*. New York: Longman.
- Medley, D. M., & Coker, H. (1987). The accuracy of principals' judgments of teacher performance. *The Journal of Educational Research*, 80(4), 242-247.
- Merriam, S. (2009). *Qualitative research: A guide to design & implementation*. San Francisco, CA: Jossey-Bass.
- Merriam, S., & Bierema, L. (2014). *Adult learning: Linking theory and practice*. 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.
- Mielke, P., & Frontier, T. (2012). Keeping improvement in mind. *Educational Leadership*, 70(3), 10-13.

- Mihaly, K. & McCaffrey, D. (2014). Why measure effective teaching? In Kane, T., Kerr, K., & Pianta, R. (Eds.) *Designing teacher evaluation systems: New guidance from the measures of effective teaching project* (pp. 123-158). Somerset, NJ: John Wiley & Sons, pp. 9-49.
- Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. *Peabody Journal of Education*, 79(4), 33-53.
- 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, February 22). Majority of NJ schools opt for widely used teacher-evaluation method. *NJ Spotlight*, pp. B7.
- Moore-Johnson, S. (2015). Will VAMs reinforce the walls of the egg-crate school? *Educational Researcher*, 44(2), 117-126.
- Morgan, G. B., Hodge, K. J., Trepinksi, T. M., & Anderson, L. W. (2014). The stability of teacher performance and effectiveness: Implications for policies concerning teacher evaluation. *Education Policy Analysis Archives*, 22(95), 1-18.

- Morganstein, D. & Wasserstein, R. (2014). *ASA statement on using value-added models for educational assessment*. Alexandria, VA.: American Statistical Association.
- 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: Pittsburg, PA.
- National Council on Teacher Quality. (2012). *State of the states 2012: Teacher effectiveness policies*. Washington, D.C.: National Council on Teacher Quality.
- Natriello, G., & Dornbusch, S. M. (1980). Pitfalls in the evaluation of teachers by principals. *Administrator's Notebook*, 29(6), 1-4.
- Neumerski, C., Grissom, J. A., Goldring, E., Cannata, M., Drake, T., Rubin, M., & Schuermann, P. (2014, March). *Inside teacher evaluation systems: Shifting the role of the principal as instructional leader*. Paper presented at the annual meeting of the Association for Education Finance and Policy, San Antonio, TX.
- Nordin, T. L. (2014). *Feedback filter: Exploring factors affecting teachers' use of observational data in teacher evaluation* (Doctoral dissertation). Rutgers University: New Brunswick, NJ.
- Organization for Economic Cooperation and Development (OECD). (2013). *Creating effective teaching and learning environments: Results from TALIS*. Paris: Organization for Economic Cooperation and Development (OECD).

- 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. (2005). Building instructional leaders' capacity to deliver constructive feedback to teachers. *Journal of Personnel Evaluation in Education*, 18(3), 171-183.
- 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.
- Papay, J. P., & Kraft, M. A. (2015). Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics*, 130, 105-119.
- Park, Y.S., Chen, J., & Holtzman, S.L. (2014). Evaluating efforts to minimize rater bias in scoring classroom observations. In Kane, T., Kerr, K., & Pianta, R. (Eds.) *Designing teacher evaluation systems: New guidance from the measures of effective teaching project* (pp. 383-414). Somerset, NJ: John Wiley & Sons.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Thousand Oaks, CA: Sage Publications.
- Patton, M. Q. (2012). *Utilization-focused evaluation*. Thousand Oaks, CA: Sage Publications.
- 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.

- Peterson, K. D. (2000). *Teacher evaluation: A comprehensive guide to new directions and practices*. Thousand Oaks, CA: Corwin Press.
- Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38(2), 109-119.
- Polikoff, M. S. (2015). The stability of observational and student survey measures of teaching effectiveness. *American Journal of Education*, 121(2), 183-212.
- Popham, W. J. (1974). *Evaluation in education: Current applications*. Richmond, CA: McCutchan Publishing.
- Popham, W. J. (1988). The dysfunctional marriage of formative and summative teacher evaluation. *Journal of Personnel Evaluation in Education*, 1(3), 269-273.
- Popham, W. J. (2013). On serving two masters: Formative and summative teacher evaluation. *Principal Leadership*, 13(7), 18-22.
- Praetorius, A. K., Pauli, C., Reusser, K., Rakoczy, K., & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction*, 31, 2-12.
- 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. (Doctoral dissertation). Saint Louis University: St. Louis, MO.
- 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.

- Reddy, L. A., & Dudek, C. M. (2014). Teacher progress monitoring of instructional and behavioral management practices: An evidence-based approach to improving classroom practices. *International Journal of School & Educational Psychology*, 2(2), 71-84.
- 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, D.C.: Regional Educational Laboratory.
- Rivkin, S. G., Hanushek, E.A., & Kain, J.F. (2005). Teachers, schools and academic achievement. *Econometrica* 73(2):417–458.
- Robinson, M. (2015). The inchworm and the nightingale: On the (mis) use of data in music teacher evaluation. *Arts Education Policy Review*, 116(1), 9-21.
- Rockoff, J. E. 2004. The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review* 94(2):247–252.
- Rosenshine, B. (1983). Teaching functions in instructional programs. *The Elementary School Journal*, 83(4), 335-351.

- Rosenshine, B., & Furst, N. (1973). The use of direct observation to study teaching. In N. Gage (Ed.), *Second handbook of research on teaching* (122-183). Chicago, IL: Rand McNally.
- Rotherham, A. J., & Mitchel, A. L. (2014). *Genuine progress, greater challenges: A decade of teacher effectiveness reforms*. Washington, D.C.: Bellwether Education Partners.
- Rothman, R., & Toch, T. (2008). *Rush to judgment: Teacher evaluation in public education*. Washington, D.C.: Education Sector.
- Rothstein, J. (2009). Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Education Finance and Policy*, 4, 537–571.
- Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *The Quarterly Journal of Economics*, 125, 175–214.
- 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.
- Sabers, D. S., Cushing, K. S., & Berliner, D.C. (1991). Differences among teachers in a task characterized by simultaneity, multidimensional, and immediacy. *American Educational Research Journal*, 28(1), 63-88.
- 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). ESEA loosens reins of teacher evaluation, qualifications. *Education Week*, 35(15), 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.
- Schmoker, M. (2012). First things first: The madness of teacher evaluation frameworks. *Phi Delta Kappan*, 93(8), 70-71.
- Scott, W. R., & Davis, G. F. (2007). *Organizations and organizing. Rational, natural and open perspectives*. New York, NY: Pearson Education International.
- 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.
- Seidel, S. (1998). Learning from looking. In N. Lyons (Ed.), *With portfolio in hand: Validating the new teacher professionalism* (pp. 69-89). New York, NY: Teachers College Press.
- 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.
- Sherin, M. G., & van Es, E. A. (2005). Using video to support teachers' ability to notice classroom interactions. *Journal of Technology and Teacher Education*, 13(3), 475.
- Shough, L. L. (2010). *Standards-based teacher evaluation: A professional development tool for improving instruction*. (Doctoral dissertation). Arizona State University: Scottsdale: AZ.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153-189.
- Siciliano, M. D. (2016). It's the Quality Not the Quantity of Ties That Matters: Social Networks and Self-Efficacy Beliefs. *American Educational Research Journal*, 53(2), 227-262.

- 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.
- Smith, O. (2015). *Central Florida educational leaders' professional perceptions of Race to the Top components concerning teacher evaluation and compensation*. (Doctoral dissertation). University of Central Florida: Orlando, FL.
- Smylie, M. (2014). Teacher evaluation and the problem of professional development. *Mid-Western Educational Researcher*, 26(2), 97-111.
- Spillane, J. P. (2012). *Distributed leadership (Vol. 4)*. Somerset, NJ: John Wiley & Sons.
- 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.
- Spillane, J. P., & Kenney, A. W. (2012). School administration in a changing education sector: The US experience. *Journal of Educational Administration*, 50(5), 541-561.
- Spina, N., Buckley, P., & Puchner, L. (2014). Shifting practices in teacher performance evaluation: A qualitative examination of administrator change readiness. *NCPEA Education Leadership Review of Doctoral Research*, 1(2), 113-128.
- Stake, R. E. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.
- State Collaborative on Reforming Education. (2012). *Supporting effective teaching in Tennessee: Listening and gathering feedback on Tennessee's teacher evaluation*. Nashville, TN: State Collaborative on Reforming Education.
- Steinberg, M. P., & Donaldson, M. L. (2016). The new educational accountability: Understanding the landscape of teacher evaluation in the post-NCLB era. *Education Finance and Policy*, 11(3), 340-359.

- Steinberg, M. P., & Garrett, R. (2016). Classroom composition and measured teacher performance: What do teacher observation scores really measure? *Educational Evaluation and Policy Analysis*, 38(2), 293-317.
- Steinberg, M. P., & Sartain, L. (2015). Does teacher evaluation improve school performance? Experimental evidence from Chicago's Excellence in Teaching project. *Education Finance and Policy*, 13(9), 535-572.
- Stodolsky, S. S. (1984). Teacher evaluation: The limits of looking. *Educational Researcher*, 13(9), 11-18.
- Stronge, J. (2013). *Effective teachers= student achievement: What the research says*. New York, NY: Routledge.
- “Student Success Schools” (pseudonym). (2015, October). “*Student Success*” public charter school. Retrieved from [<undisclosed website>](#)
- Stuit, D. A., & Smith, T. M. (2012). Explaining the gap in charter and traditional public school teacher turnover rates. *Economics of Education Review*, 31(2), 268-279.
- Sykes, G., & Winchell, S. (2010). Assessing teacher tenure. In M. Kennedy (Ed.), *Teacher assessment and the quest for teacher quality: A handbook*. (pp. 201-224). San Francisco, CA: Jossey-Bass.
- Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *The American Economic Review*, 102(7), 3628-3651.
- 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: The New Teacher Project.

- Thomson, D. L. (2014). *Teachers' perceptions of the new Massachusetts teacher evaluation instrument and process on instructional practice*. (Doctoral dissertation) Northeastern University: Chicago, IL.
- 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. (1997). Lake Wobegon: Where all teachers are competent (or, have we come to terms with the problem of incompetent teachers?). *Journal of Personnel Evaluation in Education*, 11(2), 103–126.
- Tucker, P. D., & Stronge, J. H. (2005). *Linking Teacher Evaluation and Student Learning*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Turnbull, B. J., Haslam, M. B., Arcaira, E. R., Riley, D. L., Sinclair, B., & Coleman, S. (2009). *Evaluation of the school administration manager project*. Washington, D.C.: Policy Studies Associates.
- Tuytens, M., & Devos, G. (2010). The influence of school leadership on teachers' perception of teacher evaluation policy. *Educational Studies*, 36(5), 521-536.
- Tuytens, M., & Devos, G. (2014). How to activate teachers through teacher evaluation. *School Effectiveness and School Improvement*, 25(4), 509-530.
- U.S. Department of Education. (2009). *Race to the Top program: Executive summary*. Washington, D.C.: U.S. Department of Education.

- Vekeman, E., Devos, G., & Tuytens, M. (2014). The influence of teachers' expectations on principals' implementation of a new teacher evaluation policy in Flemish secondary education. *Educational Assessment, Evaluation and Accountability*, 27(2), 129-151.
- Weisberg, D., Sexton, S., Mulhern, J., Keeling, D., Schunck, J., Palcisco, A., & Morgan, K. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. New York, NY: The New Teacher Project.
- Westerberg, T. R. (2013). Feedback for Teachers. *Principal Leadership*, 13(7), 30-33.
- Wexler, A. (2014). Reaching higher? The impact of the Common Core State Standards on the visual arts, poverty, and disabilities. *Arts Education Policy Review*, 115(2), 52-61.
- Whitehurst, G., Chingos, M., & Lindquist, K. (2014). *Evaluating teachers with classroom observations*. Washington, D.C.: Brookings Institute.
- Wiggins, G. (2012). Seven keys to effective feedback. *Educational Leadership*, 70(1), 10-16.
- Wilhelm, A. G., Chen, I. C., Smith, T. M., & Frank, K. A. (2016). Selecting expertise in context: Middle school mathematics teachers' selection of new sources of instructional advice. *American Educational Research Journal*, 53(3), 456-491.
- 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.
- Windish, D. H. (2012). *Central Florida educational leaders' professional opinions of the Race to the Top grant components concerning teacher evaluation and compensation prior to implementation*. (Doctoral dissertation) University of Central Florida: Orlando, FL.
- Wise, A. E., Darling-Hammond, L., McLaughlin, M. W., & Bernstein, H. T. (1985). Teacher evaluation: A study of effective practices. *The Elementary School Journal* 86(1), 61-121.

Wolf, R. L. (1972). *The role of the teacher in classroom evaluation*. (Doctoral dissertation).

University of Illinois at Urbana-Champaign: Urbana-Champaign, IL.

Wood, D. R. (2007). Professional learning communities: Teachers, knowledge, and knowing.

Theory into Practice, 46(4), 281-290.

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). *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.

APPENDICES

Appendix A: Consent Forms

Principal Investigator: Heba Abdo

Project Title: Teacher Use of Evaluation Feedback

INFORMED CONSENT FORM

My name Heba Abdo, and I am graduate student at Graduate School of Education at Rutgers University. You are invited to participate in a research study. The purpose of this research is to understand how your school is using the teacher evaluation process to support teacher learning. All three school administrators and several teachers from your school will participate in this study.

You will be part of one of three groups:

Administrators: You will participate in one interview at the beginning of the study.

Focus Group Teacher Participants: You will participate in one group interview at the beginning of the study.

Other Teacher Participants: Throughout the course of the study, I will be interviewing you on two occasions, after you receive feedback from your evaluator. In addition, I may look at documents such as your lesson plans, feedback documents, or student work.

Participation in all aspects of the study is voluntary, and you may choose to participate in one part, without volunteering in others.

The data collected during this research will be kept completely confidential. Confidential means that the research records may include some de-identified data about you and this information will be stored in such a manner that some linkage between your identity and the response in the research exists. Please note that I will keep this information confidential by limiting individual's access to the research data, keeping it in a secure location, and changing all names in the study.

The research team and the Institutional Review Board at Rutgers University are the only parties that will be allowed to see the data, except as may be required by law. If a report of this study is published, or the results are presented at professional conference, only group results will be stated. All study data will be kept for 3 years.

There are no foreseeable risks to participation in this study, however possible benefits of participation in this study helping your school improve its teacher evaluation practice, increasing your reflection about school and classroom practice, and being empowered to share your opinions.

Participation in this study is voluntary. You may choose not to participate, and you may withdraw at any time during the study procedures without any penalty to you. In addition, you may choose not to answer any questions with which you are uncomfortable.

If you have any questions about the study or study procedures, you may contact myself at heba.abdo@gse.rutgers.edu.

If you have any questions about your rights as a research subject, you may contact the IRB Administrator at Rutgers University at:

Rutgers University, the State University of New Jersey

Institutional Review Board for the Protection of Human Subjects

Office of Research and Sponsored Programs

3 Rutgers Plaza

New Brunswick, NJ 08901-8559

Tel: 848-932-0150

Email: humansubjects@orsp.rutgers.edu

If requested, I will provide a copy of this consent form for your records.

Sign below if you agree to participate in this research study:

Subject (Print) _____

Subject Signature _____ Date _____

Principal Investigator Signature H. C. Allen Date November 10, 2015

Appendix B- Data Collection Protocols

Recruitment Email for Teachers

Hello.

My name is Heba Abdo and I am a graduate student at Rutgers University in NJ. I am conducting my dissertation research on teacher evaluation and your school is the focus of my study. I am particularly interested in how you, as teachers, teacher evaluation in your school and how you use it (or don't) to make decisions about your classroom instruction.

I chose your school site because I heard that administrators are doing some innovative things with the teacher evaluation, including giving frequent and detailed feedback, and using classroom video to support all teachers. By participating in this study, you can share your experiences with it. Your perspectives might support other schools as they try to implement systems that help teachers improve their practice.

Your participation in the study will be completely confidential. Your name will never be used in association with anything you say and I will take every precaution to make your identity.

Neither your administrator, your co-workers, nor the readers of my dissertation will know who said what. Participation should take 40-60 minutes, and will be scheduled in advance. Additionally, your participation is voluntary and you can withdraw at any time.

If you are interested, please click on this [link](#) which will allow you to enter some demographic and contact information. This link should take less than 1 minute to complete and can be completed on a computer, smartphone, or tablet.

If you have any questions, please do not hesitate to contact me (heba.abdo@gse.rutgers.edu).

Thank you so much for your time.

Heba Abdo

Rutgers University

Survey for Teacher Recruitment

How can I contact you to set up an interview appointment (include best email or phone number)?

Years of Teaching Experience:

First Year

2-5

6-10

11+

Years at Student Success District:

First Year

2-5

6-10

11+

Current Grade Level:

K-2

3-5

6-8

Subject Areas taught this Year (Choose all that apply):

All Academic (ELA, Math, SS, Science)

ELA

Math

SS

Science

Art, Music, PE

Support Staff

Other

Teacher Preparation Program:

Undergraduate Program

Graduate Program

Alternate Route

Teach for America

Urban Teachers

Other

Anything else you would like to add (optional)

Initial Administrator Interview

Hello, my name is Heba and I am a graduate student at Rutgers University in NJ. I am interested in understanding your experiences with your teacher evaluation system at Student Success Schools. I would like to begin today's session by giving you time to read through the consent form for this project. It will answer some basic questions you might have. You can also ask me any questions you have after you read it.

- give time to read through consent form, ask questions, and sign.

Let's start with some general questions:

Describe how you believe that teachers learn best?

Describe the Student Success School TE system. **(Principal Only)**

Remember:

how far apart

% unannounced

% informal

how long is each session

what do teachers receive

how do walk-throughs fit in

how is it the same/different for teachers with different levels of experience/ tenure

how do teacher goals fit in

time spent on TE.

Follow up

General procedure

Please describe the Google Drive videos and feedback system (Principal Only)

Probe: how do components support teachers

What training did you receive to conduct TE?

Probes:

-for rating

-for feedback

-what was it like, how often, what did you learn

What is your case-load?

Probe:

-How many teachers?

-How often do you observe, give feedback, other

Next I would like to know some information about your feelings and comfort in regards to the tool.

How do you feel about evaluating teachers with the Student Success Schools' Teacher Evaluation Tool?

Probes:

- classroom management
- classroom instruction
- giving feedback, positive and negative

What kinds of support do you receive to manage your evaluation case-loads and conduct evaluations?

What are your goals, typically, during the pre-conference?

How do you typically prepare yourself to confer with teachers after a classroom observation?

Probe:

- Describe one positive example of feedback

What resources do you provide to help support teacher improvement after TE?

Probes:

- time and personnel and PD?
- embedded, one time?

How does the administrative team use teacher evaluation data?

Finally, I would like to know how you feel that the teacher evaluation system is perceived across the school or your department.

How would you describe your department's view of the teacher evaluation process at Student Success school?

In what ways do you think that teachers are using the TE system to improve their practice?

Are there any ways you feel the teacher evaluation tool could better serve your department.

Is there anything else that you would like to say about the Student Success Teacher Evaluation system or your work with teachers?

Finally, I would like to ask some basic demographic information.

Years of teaching experience:

Teaching content area:

Years of admin experience:

Years since you were a classroom teacher:

Years at Student Success School:

Gender:

Race/Ethnicity:

Thank you so much for your time today. Please feel to reach out to me at any time if you would like further information about the study, or to share new information as the year continues.

Teacher Focus Group with 5-8 Teachers

Demographics Pre-Survey

Your answers from this session will be kept confidential. That means that your name will be removed from all records and that if I later report about something you say, it will be reported with no associated identifying information. Before I begin, I want to get some information about your previous experience.

Please take a moment to fill the following:

Name (this is only for my use when taking notes.):

Years of teaching experience:

Teaching content area:

Years at Student Success School:

Gender:

Race/Ethnicity:

Teacher Education Program (Please circle one of the following)

Alternate Route Certification

Urban Teachers

Teacher for America

Other

University Preparation Program

Focus Group

Hello, my name is Heba and I am a graduate student at Rutgers University in NJ. I am interested in understanding your experiences with your teacher evaluation system at Student Success School. I would like to begin today's session by going through the consent form for this project. It will answer some basic questions you might have. After that we will have a short question/answer period.

-read through consent form

-allow for a maximum of 5 questions

-give teachers a moment to sign

Before we begin, please take a moment to fill in some demographic information.

-allow a few minutes for this

Now that that is all out of the way, I would like to explain how today session will work. I have purposely chosen a group with varied backgrounds because I value diversity in points of view. I want you all to feel comfortable expressing your opinions, even if you feel like they might not be

popular. I also want to stress that since I am a total outsider at your school, anything you say in this room will be kept confidential. You should also know that participating in this group and sharing your opinions, both positive and negative, may lead to better practices in your school when your administrators read the de-identified report. It may also lead to better understanding of the teacher evaluation system in your whole school, which could mean that it becomes a more teacher-friendly experience.

Finally, I would like to ask that we each take turns speaking one at a time so that everyone can be heard. Because I will be transcribing this later, it will be particularly important to be able to distinguish every word. When people talk over each other, this is difficult. In order to facilitate this, we will be using my phone as a “talking stick.” Only the person holding the phone can speak. Please speak clearly into the phone as I will be recording this session.

I would like you to start by just describing some of the aspects of the teacher evaluation system at Elmwood.

I just want to start by getting a general understanding of your teacher evaluation system here at Student Success School. What do you think is a major strength of it?

Describe it. What generally happens?

What type of feedback do you get?

How were goals set? And who decided or how did they decide if your goals were met

Taking a look at the Student Success Teacher Evaluation rubric or based on your own experiences, what do you think it means to be a good teacher at this school?

Next I would like to know more about how the teachers interact with the teacher evaluation system.

What type of support do teachers have to meet the standards outlined in the Student Success School TE system?

Describe how you use feedback from the TE system at Student Success Schools.

If you have experienced professional learning that is linked to your TE feedback, can you describe it?

Can you describe any ways that teacher evaluation system could support you better?

Thank you so much for your time today. Please feel to reach out to me at any time if you would like further information about the study, or to share new information as the year continues.

Initial Teacher Interviews with 6 Teachers

Hello, my name is Heba and I am a graduate student at Rutgers University in NJ. I am interested in understanding your experiences with your teacher evaluation system at Student Success Schools. I would like to begin today's session by giving you time to read through the consent form for this project. It will answer some basic questions you might have. You can also ask me any questions you have after you read it.

- give time to read through consent form, ask questions, and sign.

It is really important that you know that everything that you say during this session will be confidential. I deeply value your opinions and am very interested to hear what you have to say about these questions. I also want you to know that anything you say about your school's teacher evaluation system, both positive and negative, will be a great asset to the whole school community. Your administrators will likely take your recommendations under consideration when they read the de-identified report.

First I would like to hear a little more about you.

Please tell me about your teaching experience.

Remember: Ask about preparation and how it impacts her teaching.

What has been your greatest source of joy in your work with students?

If you could change one thing about your work with students, what would it be? Why?

Let's talk a little bit about the teacher evaluation process at Elmwood:

Please tell me about your most positive experiences with the teacher evaluation system here at Elmwood. Take me from start to finish. What was it like, how did it feel, how long did it take?

If necessary: What is the biggest strength of the teacher evaluation process?

If necessary: What might you change about the teacher evaluation process?

Can you describe a time in which you used feedback from the teacher evaluation process in your classroom?

Do you want to tell me anything else about the observation here or the feedback you get?

Thank you so much for your time today. I look forward to speaking with you again on my next school visit.

Review of Videos and Teacher Evaluation Results for 6 Select Teachers

Feedback as Result of:

- ☐ Informal Observation (such as walk-through)
- ☐ Formal Observation
- ☐ Pre-Conference
- ☐ Post-Conference
- ☐ Announced Observation
- ☐ Unannounced Observation
- ☐ Other _____

Feedback Included:

- ☐ Checklist of Teacher Behaviors
- ☐ General Feedback (such as “effective,” “good”)
- ☐ Lengthy Descriptions of Observed Practice
- ☐ Specific Suggestions for Improvement of Practice
- ☐ Referral to Specific PD or Resources
- ☐ Reference to Other Data (such as student achievement)
- ☐ Opportunity for Teacher to Interpret Classroom Events
- ☐ Collaborative Goal Setting
- ☐ Evidence of Student Learning Included
- ☐ Evidence of Teacher Practice Included
- ☐ Percent of Administrator Talk Time:
- ☐ Percent of Teacher Talk Time:

Other Notes: _____

Review of PD Documents

- ☐ Allow for Teacher Collaboration
- ☐ Allow for Principal Collaboration
- ☐ Allow for School-Wide Collaboration
- ☐ Linked to Teacher Evaluation Feedback
- ☐ Allows Teachers to Set Professional Goals
- ☐ Helps Teachers Engage in Self-Reflection

Other Notes: _____

Final Teacher Interviews Teachers

Teacher 1

Hello, again! As you may remember, my name is Heba and I am a graduate student at Rutgers University in NJ and I am interested in understanding your experiences with your teacher evaluation system at Woodbridge.

I want to remind you that everything that you say during this session will be confidential. I deeply value your opinions and am very interested to hear what you have to say about these questions. I also want you to know that anything you say about your school's teacher evaluation system, both positive and negative, will be a great asset to the whole school community. Your administrators will likely take your recommendations under consideration when they read the de-identified report.

Let's start with talking the feedback you receive during evaluation:

What is your process for understanding the feedback you receive?

How much time do you spend looking at your feedback on your own?

How much time do you spend discussing your feedback with your administrator?

What do you believe is your evaluator's goal when providing you with feedback?

I see that you and other teachers took a lot of time filling out a self-assessment in the beginning of the year. How did you go through that process?

How do you feel your supervisor supported you with your areas of need from that process?

There were areas for which you and your evaluator disagreed. Can you talk to me about that conversation and how you all came to consensus?

Now I would like to ask about how, if at all, you have used the feedback you received. I will be asking about three key parts of teacher work: thinking about your classroom, planning for you classroom, and acting in your classroom.

In what ways have you found that feedback from TE helps you to think about your classroom?

In what ways have you found that feedback from TE helps you to plan your lessons?

Please describe a time that you used feedback from TE to change the way you related to students, enacted instruction, or otherwise acted while you were actually teaching.

In the case that I have a follow up question, do you mind if I reach out by email?

Thank you so much for your time today. Please feel to reach out to me at any time if you would like further information about the study, or to share new information.

Teacher 2

Hello, my name is Heba and I am a graduate student at Rutgers University in NJ. I am interested in understanding your experiences with your teacher evaluation system at Friendship Schools. I would like to begin today's session by giving you time to read through the consent form for this project. It will answer some basic questions you might have. You can also ask me any questions you have after you read it.

- give time to read through consent form, ask questions, and sign.

It is really important that you know that everything that you say during this session will be confidential. I deeply value your opinions and am very interested to hear what you have to say about these questions. I also want you to know that anything you say about your school's teacher evaluation system, both positive and negative, will be a great asset to the whole school community. Your administrators will likely take your recommendations under consideration when they read the de-identified report.

First I would like to hear a little more about you.

Please tell me about your teaching experience.

Remember: Ask about preparation and how it impacts her teaching.

What has been your greatest source of joy in your work with students?

If you could change one thing about your work with students, what would it be? Why?

Let's talk a little bit about the teacher evaluation process at Woodbridge:

Can you please take me through, from start to finish, your most recent observation experience.

Please include your feelings and role throughout the process.

If necessary: What is the biggest strength of the teacher evaluation process?

If necessary: What might you change about the teacher evaluation process?

According to my notes, it seems you had some pretty varied feedback throughout the year. I am interested in how you took that feedback and translated it into practice.

First, how did you go about understanding that feedback?

How much time do you spend looking at your feedback on your own?

How much time do you spend discussing your feedback with your administrator?

Second I want to know how you used feedback to help you to think about your classroom?

In what ways have you found that feedback from TE helps you to plan your lessons?

Please describe a time that you used feedback from TE to change the way you related to students, enacted instruction, or otherwise acted while you were actually teaching.

Finally, how did your supervisors support you in meeting your areas of improvement? Describe some of these examples.

Probes:

- resources you receive

- other supports

In the case that I have a follow up question, do you mind if I reach out by email?

Thank you so much for your time today. Please feel to reach out to me at any time if you would like further information about the study, or to share new information.

Teacher 3

Hello, again! As you may remember, my name is Heba and I am a graduate student at Rutgers University in NJ and I am interested in understanding your experiences with your teacher evaluation system at Woodbridge.

I want to remind you that everything that you say during this session will be confidential. I deeply value your opinions and am very interested to hear what you have to say about these questions. I also want you to know that anything you say about your school's teacher evaluation system, both positive and negative, will be a great asset to the whole school community. Your administrators will likely take your recommendations under consideration when they read the de-identified report.

Let's start with talking the feedback you receive during evaluation:

Last time I was here, we talked about the feedback you received and how you felt it was a bit superficial, however when I looked through school documents, I saw that teachers were getting pretty targeted feedback. Do you think the evaluation process works well for any teachers? How do you think it might work better for you?

I also saw that teachers took a lot of time to think through their personal goals in the beginning of the year. How do you think the school can incorporate teacher goals better into the evaluation process.

Now I would like to ask about how, if at all, you have used the feedback you received. I will be asking about three key parts of teacher work: thinking about your classroom, planning for you classroom, and acting in your classroom.

In what ways have you found that feedback from TE helps you to think about your classroom?

In what ways have you found that feedback from TE helps you to plan your lessons?

Please describe a time that you used feedback from TE to change the way you related to students, enacted instruction, or otherwise acted while you were actually teaching.

In the case that I have a follow up question, do you mind if I reach out by email?

Thank you so much for your time today. Please feel to reach out to me at any time if you would like further information about the study, or to share new information.

Teacher 4

Hello, again! As you may remember, my name is Heba and I am a graduate student at Rutgers University in NJ and I am interested in understanding your experiences with your teacher evaluation system at Woodbridge.

I want to remind you that everything that you say during this session will be confidential. I deeply value your opinions and am very interested to hear what you have to say about these questions. I also want you to know that anything you say about your school's teacher evaluation system, both positive and negative, will be a great asset to the whole school community. Your administrators will likely take your recommendations under consideration when they read the de-identified report.

Let's start with talking the feedback you receive during evaluation:

What is your process for understanding the feedback you receive?

How much time do you spend looking at your feedback on your own?

How much time do you spend discussing your feedback with your administrator?

What are feedback sessions like?

How do your personal goals fit in to your feedback sessions?

Based on our last meeting, I felt that you were incredibly aware of your own accomplishments and limitations as a teacher. I wonder how does the evaluation process help you, in particular, to think about your classroom more deeply?

Now I would like to ask about how, if at all, you have used the feedback you received. I will be asking about three key parts of teacher work: thinking about your classroom, planning for you classroom, and acting in your classroom.

In what ways have you found that feedback from TE helps you to think about your classroom?

In what ways have you found that feedback from TE helps you to plan your lessons?

Please describe a time that you used feedback from TE to change the way you related to students, enacted instruction, or otherwise acted while you were actually teaching.

In the case that I have a follow up question, do you mind if I reach out by email?

Thank you so much for your time today. Please feel to reach out to me at any time if you would like further information about the study, or to share new information.

Teacher 5

Hello, again! As you may remember, my name is Heba and I am a graduate student at Rutgers University in NJ and I am interested in understanding your experiences with your teacher evaluation system at Woodbridge.

I want to remind you that everything that you say during this session will be confidential. I deeply value your opinions and am very interested to hear what you have to say about these questions. I also want you to know that anything you say about your school's teacher evaluation system, both positive and negative, will be a great asset to the whole school community. Your administrators will likely take your recommendations under consideration when they read the de-identified report.

Let's start with talking the feedback you receive during evaluation:

What is your process for understanding the feedback you receive?

How much time do you spend looking at your feedback on your own?

How much time do you spend discussing your feedback with your administrator?

What do you believe is your evaluator's goal when providing you with feedback?

Last time I met with you, you spoke about getting support from your supervisor with quick fixes like incorporating timers for center work, however I see that you set some personal goals for yourself in much more involved areas of instruction like differentiating for learning styles. Were

you getting feedback towards your personal goals, and if so how supported do you feel in meeting your areas of improvement? Describe some of these examples.

Probes:

- resources you receive

- other supports

Now I would like to ask about how, if at all, you have used the feedback you received. I will be asking about three key parts of teacher work: thinking about your classroom, planning for you classroom, and acting in your classroom.

In what ways have you found that feedback from TE helps you to think about your classroom?

In what ways have you found that feedback from TE helps you to plan your lessons?

Please describe a time that you used feedback from TE to change the way you related to students, enacted instruction, or otherwise acted while you were actually teaching.

In the case that I have a follow up question, do you mind if I reach out by email?

Thank you so much for your time today. Please feel to reach out to me at any time if you would like further information about the study, or to share new information.

Principal Final Interview

Hi x.

It is so nice to speak with you again and to have shared in this experience with your staff. It's a great school and I heard so many positive things from so many people in the building about your leadership, about the students and community, and about the support they were receiving as teachers. I also noticed that many teachers spoke with such confidence about your core values that you have communicated to me so that to me was an indicator that school wide professional learning and change efforts have been really strong.

However, I didn't always hear such strength of voice when teachers talked about evaluation and feedback they received. It seemed a bit like a mixed bag. So I think it is kind of perfect that now you have had some distance and time to reflect, and could think about some lessons learned that we can talk about to really substantiate our submission to ascd.

There are a few themes I saw that I would like to get your feedback on:

Principal Work Load

Original set up

Vast Differences between Departments and years- sustainability

Supporting the formative process

Targeting support across the learning continuum

Principal Work Load

I notice that one AP had many more teachers. Why was that?

I also notice that by Feb. Most have dropped use of FIOT. What was happening there?

Types of Comments- SPED talked a lot more about instruction- why? What do you think the result was?

Original set up

No teacher had really heard of the 6 Steps of Feedback, even master teachers and department heads- how is this communicated to teachers?

Even Admin forgot the 6 steps- like was lost for words for a moment

One administrator said: "The system is just a tool to put in the feedback that you are giving, but I didn't get enough training on giving feedback."

Would you say that that is accurate and in what ways have you worked this out or would you work this out in the future?

Vast Differences between Departments and Years/Sustainability

Overall huge difference between last year and this year.

What would you say attributed to that? How were ap's differently supported or differently tasked between the two years?

For aps that were just better, how did they become better?

Somehow results for formative evaluation on the survey were mixed, though everyone felt comfortable with summative- why do you think that is

What about keeping this work sustained?

How about specifically supporting principals- many teachers and admins said that admins were always pulled away from evaluation. Do you think that is accurate and do you see any way to alleviate that?

Supporting the Formative Process

One teacher talked about a goal notebook she kept with MsX- did you have any other way of visually tracking growth for teachers. What could that look like?

Many talked about support coming from other people in the building, and not from admin- UTC, Guided Reading Coach which I know were purposeful choices for admin- do you feel that that is accurate? Why do you think that is?

Targeting Support Across the Learning Continuum

Higher end teachers seemed to get a lot of PD support in order to benefit the school, but not so much for their learning needs. What are your thoughts about that?

What are your plans for targeting support in coming years for all levels?

I also notice that there was a drop, more generally in the followup teachers received. Would you say that is accurate? Can you expand on this idea?

A lot of teachers talked about increasing teacher networking experiences and using each other as a resource being something they would like more of. What are your thoughts?

Follow up questions

It would be nice to talk to other APs- do you think Ms. X would be willing to talk to me now?

There is no fall 14 survey for middle school. Why so? Anything else you have for comparison?

Are there spring findings for this year?

Please send binder guidelines.