

HOW PRINCIPALS' COGNITIVE SCHEMAS IMPACT THEIR IMPLEMENTATION OF
TEACHER EVALUATION POLICIES AND TEACHER EVALUATION SYSTEMS

By

David B. Reid

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Educational Policy – Doctor of Philosophy

2017

ABSTRACT

HOW PRINCIPALS' COGNITIVE SCHEMAS IMPACT THEIR IMPLEMENTATION OF TEACHER EVALUATION POLICIES AND TEACHER EVALUATION SYSTEMS

By

David B. Reid

Since 2009, the United States (US) federal government has spearheaded a nationwide teacher evaluation reform effort, encouraging states to change their process of evaluating teachers with a focus on evaluations that better distinguished teacher performance as well as provide better information on what makes a high-quality teacher (US Department of Education, 2009). The US Department of Education enacted this reform through the Race to the Top (RTTT) initiative and by granting many states Elementary and Secondary Education Act (ESEA) waivers if they changed their teacher evaluation systems to align with ESEA priorities, such as evaluating teachers in part based on student assessment data. These two levers are the primary reason that since 2009 more than two-thirds of states have made significant changes to how teachers are evaluated (The Center for Public Education, 2014).

However, the passage of the 2015 Every Student Succeeds Act (ESSA) allows for an increased role of states and districts in shaping teacher evaluation policies and teacher evaluation systems. All of these changes have the potential to impact how states, districts, schools, and individual principals make sense of and use teacher evaluation policies and systems.

Organizational and individual sensemaking of teacher evaluation policies and systems is of particular importance due to the high-stakes most states attach these policies and systems (e.g. using these policies and systems for the hiring, firing, and tenure decisions of teachers). Because school principals play a pivotal role in how teacher evaluations look in practice, I argue policymakers, researchers, and practitioners must better understand how principals think about and ultimately enact these complex policies and systems. In this dissertation, I answer the

following research questions: 1) How do principals' cognitive schemas influence how principals come to understand and implement teacher evaluation policies and systems?; 2) What role does external pressure play in shaping principal learning and enactment of teacher evaluations policies and systems?; and 3) In what ways, if any, do principals' experience and external pressure interact during the implementation process?

This exploratory multi-case study examines data from school principals (N=12) in Michigan, including interviews (n=36), observations (n=24), and questionnaires (n=12) collected in 2016 and 2017. Additionally, teacher interviews (n=12) and specific teacher evaluation district documents inform this study. Results show: (1) principals in high-pressure environments perceive a pressure to differentiate their teachers' final evaluation ratings, which typically results in these principals rating their teachers more critically than their peers who work in low-pressure environments; (2) principals with high-levels of experience engage in situational leadership, while principals with low-levels of experience engage in relational leadership, which impacts how teacher evaluation policies and systems look in practice; and (3) principals with high-levels of experience are less likely than their more inexperienced peers to use previous teacher data (evaluations/test scores) when evaluating teachers. Implications for policy and practice are discussed.

Copyright by
DAVID B. REID
2017

ACKNOWLEDGMENTS

Country music artist Kenny Chesney wrote a song called, “I Didn’t Get Here Alone” in which he thanks the many people who helped make him successful. I am in no way comparing myself to Mr. Chesney, but many of the lyrics from the song ring true when I think about my career in education and my time at Michigan State University (MSU). I certainly did not get to this point of my life and career alone, so in the paragraphs that follow, I aim to thank the individuals who have supported me along the way. Without the support of these people I likely would not have been able to complete this program and dissertation. I will apologize in advance, as I am sure I will leave out the names of some people who were very supportive and influential during my time working in and studying the field of education.

First, I would like to thank my wife, Molly, who left her job and moved across the country to support me. For that I will be forever grateful. Second, I would like to thank my advisor, Anne-Lise Halvorsen, whose guidance and support is unparalleled. Her constant mentorship during my time at MSU was more than I ever could have asked for or imagined. Additionally, I would like to thank Josh Cowen, who provided great guidance as the co-chair on this dissertation and always helped me consider ideas and concepts in a more critical way. I also want to thank Rebecca Jacobsen, who, although I was not her advisee, treated me as one, and took countless time out of her schedule to support me throughout my entire time at MSU. Additionally, I would like to thank Chris Torres for joining my dissertation committee and always providing thoughtful feedback on my writing. Finally, thank you to Michael Sedlak, who supported me greatly during my time at MSU.

I most certainly need to thank my graduate student colleagues who have been there for

me through the entire program. The number of people to whom I owe a thank you is many and I am sure I will not mention everyone, but I would like to mention a few by name. I am indebted to Jason Burns, Ben Creed, Kim Jansen, Alyssa Morley, John Lane, and Rachel White. Thank you all for your constant engagement and support. To my parents, brother, and sister, who have supported me for as long as I can remember. It is extremely reassuring to know I always have your unwavering support. Thank you. Finally, thank you to my former students from Phoenix, Arizona. You are the reason I decided to pursue this degree. Your perseverance, humor, and brilliance serves as a constant reminder of why I do this work.

TABLE OF CONTENTS

LIST OF TABLES	ix
Chapter 1: Introduction.....	1
Research Questions.....	7
Contribution of the Dissertation.....	8
Outline of the Dissertation.....	9
Chapter 2: Literature Review.....	11
Part I: Education Policy Implementation Overview.....	12
Part II: Teacher Evaluation Policy Implementation.....	16
Part III: Principal Cognition and Policy Implementation.....	19
The Co-Evolution of Teacher Evaluation Policies and the Role of Principals.....	22
The Role of Principal Cognition During Teacher Evaluations.....	26
Experience and Context: Why and How They Matter for Policy Implementation.....	27
Gap in the Literature.....	32
Chapter 3: Framing the Research.....	34
Cognitive Schemas.....	35
Sensemaking Theory.....	40
Individual vs. Collective Sensemaking.....	43
The Usefulness of Sensemaking.....	47
Chapter 4: Research Design and Methodology.....	51
Research Design and Research Questions.....	51
Rationale for a Case Study.....	52
Study Context: Educator Evaluations in Michigan.....	53
Participants and Sampling Strategy.....	58
Data Collection.....	60
Data Analysis.....	64
Establishing Validity.....	67
Limitations.....	69
Chapter 5: How Principals' Cognitive Schemas Impact Their Implementation of Teacher Evaluation Systems.....	70
Overarching Theme: Individual vs. Collective Sensemaking.....	72
Subtheme One: Principal Leadership.....	72
Nuances.....	80
Subtheme Two: Use of Prior Evaluation Data.....	83
Nuances.....	86
Subtheme Three: Accurate Reflection of Teacher Effectiveness.....	89

Nuances.....	93
Subtheme Four: Hiring Decisions.....	94
Nuances.....	96
Chapter Summary.....	97
 Chapter 6: The Role of External Context and Experience in Principal Learning and Implementation of Teacher Evaluation Systems	100
Theme One: Differentiating Teacher Evaluation Ratings.....	101
Nuances.....	106
Theme Two: What do Teacher Evaluations Measure?.....	107
Nuances.....	111
How do Experience and External Pressure Interact during the Implementation Process?.....	112
Chapter Summary.....	123
 Chapter 7: Discussion, Implications, and Conclusions.....	125
The Goals of Teacher Evaluation Policy.....	125
Principals' Role in Teacher Evaluations.....	131
Implications: For Policymakers.....	137
Implications: For Practitioners.....	140
Limitations and Future Research.....	143
Conclusions.....	145
 APPENDICES	150
Appendix A Principal Questionnaire	151
Appendix B Principal Interview Protocol 1.....	156
Appendix C Principal Interview Protocol 2.....	158
Appendix D Principal Interview Protocol 3.....	160
Appendix E Teacher Interview Protocol.....	161
Appendix F Observation Protocol.....	163
 WORKS CITED.....	164

LIST OF TABLES

Table 1.1. Dissertation Participants (N=12).....	6
Table 4.1. Timeline of Educator Evaluation Changes in Michigan Since 2009.....	56
Table 4.2. Principal Participant Sample.....	59
Table 4.3. Principal Background Information (TPS or Charter, Principal Experience).....	60
Table 4.4. Principal Data Collected.....	64

Chapter 1: Introduction

Researchers and program evaluators often find that in practice policies and systems differ significantly from what policy designers had envisioned (Elmore, 1980; Honig, 2006; Honig & Hatch, 2004; Lipsky, 1980; Odden, 1991; Spillane, 2000; Spillane, Reiser, & Reimer, 2002; Weatherly & Lipsky, 1977). Initially researchers concluded inconsistent policy and system implementation was due to a lack of practitioner will and capacity (McLaughlin, 1987; Odden, 1991), but more recently education scholars have concluded the process of how policies and systems enter organizational environments is much more complex (Honig, 2006; Spillane, 2000; Spillane et al., 2002). In an effort to better understand this complex process, scholars who study education policy and system implementation have turned to cognitive and learning science theories to examine what factors shape how policies and systems are interpreted and ultimately implemented by practitioners (Cohen & Hill, 2001; Halverson & Clifford, 2006; Spillane et al., 2002).

Originating in cognitive psychology, theories of cognition examine the interaction between one's psychological processes and the information that comes in contact with and passes through one's psychological network (Grider, 1993). Through the lens of cognitive theory, an individual's ability to learn is dependent upon how he or she receives, organizes, and processes new and existing information (Grider, 1993). Research that uses the cognitive frame finds a complex cognitive process occurs when local actors (e.g. school principals, teachers, or district administrators) attempt to reconcile their previous understandings, habits, and current situational context with new policy demands (Halverson & Clifford, 2006; Honig, 2006; Spillane et al., 2002). Theories of cognition can help scholars of policy and system implementation understand how and why there is often a disconnect between policymakers' desired outcomes

and goals and what happens in practice.

One specific cognitive learning theory that is particularly useful when looking at how individuals interpret evolving policies and systems is sensemaking theory. Sensemaking theory acknowledges that past experiences and prior knowledge shape learning and that learning occurs through our social and situational context (Greeno, 1998; Weick, 1995). Scholars use sensemaking theory to attempt to explain how individuals and organizations interpret the policies, systems, and reforms with which they come in contact (Coburn, 2005; Halverson & Clifford, 2006; Spillane, 2000; Spillane et al., 2002). This research generally finds individuals' prior knowledge, beliefs, and experiences greatly impact how individuals think about and make sense of new and changing information. Central to using a sensemaking theory frame in policy research is the idea that cognition does not simply explain how individual actors interpret information, policies and systems, but also explains how these individuals respond to changes in their environment (Spillane et al., 2006). Spillane et al. (2006) further describe this idea by concluding that an individual's prior knowledge, experience, and beliefs all serve as a lens for what he or she notices in their environment and how information is processed, organized, and interpreted (p. 49). In short, how an individual makes sense of information has a strong relationship to how this information ultimately enters, remains in, or disappears from, this individuals' environment of practice.

Education is a particularly interesting and timely field to study how cognition impacts policy and system interpretation and implementation because it is arguably one of the most active policymaking arenas today with multiple agencies, including federal, state, and local governments, creating new policies at an ever-increasing pace (Honig, 2006; Spillane & Kenney, 2012). These governments ask school leaders and their faculty to implement a dizzying amount

new and reformed policies and systems year after year. For example, the No Child Left Behind (NCLB) act of 2001 required all students in grades K-8 be tested annually in reading in math. This resulted in new testing policies, standards policies, and accountability policies for both teachers and school leaders.

Of the many policies and systems schools must implement one that is a prime candidate for study is teacher evaluation. Currently within the field of education there is perhaps no more polarizing issue than teacher evaluation policies and teacher evaluation systems. The polarizing nature of these policies and systems is due in part to the increasing acknowledgement that teacher quality can positively impact student outcomes, such as achievement and attendance (Aaronson, Barrow, & Sander, 2007; Chetty, Friedman, & Rockoff, 2014; Rockoff, 2004) and in part to research suggesting teacher evaluation systems have historically done a poor job distinguishing teacher performance (Donaldson, 2009; US Department of Education, 2009; Weisberg, Sexton, Mulhern, & Keeling, 2009). In an effort to address deficient teacher evaluation systems, the 2009 Race to the Top (RTTT) initiative encouraged states to change the process of evaluating teachers, with focus on evaluations that better distinguished teacher performance as well as provided better information on what makes a high-quality teacher (US Department of Education, 2009). Additionally, the US Department of Education granted many states Elementary and Secondary Education Act (ESEA) waivers if they changed their teacher evaluation policies to align with ESEA priorities, such as evaluating teachers in part based on student assessment data. These two levers are the main reasons since 2009 more than two-thirds of states have made significant changes to their teacher evaluation policies and systems (The Center for Public Education, 2014). This nationwide reform effort has resulted in school leaders having to make sense of new teacher evaluation policies and systems at a rapid pace.

Additionally, for many school leaders, teacher evaluation policies and systems are continuing to evolve, tasking school leaders with making sense of multiple editions of these policies and systems.

To date, much of the research on teacher evaluation policy reform has focused teachers (Aaronson et al., 2007; Donaldson, 2013; Taylor & Tyler, 2012). This makes sense given the widespread evidence supporting teachers are the most important school-based factor that can positively impact student achievement (Aaronson et al., 2007; Chetty et al., 2014; Rockoff, 2004). However, more recent research has begun to acknowledge the important role principals play in student outcomes, such as attendance, achievement, and graduation rates (Beteille, Kalogrides, & Loeb, 2009; Branch, Hanushek, & Rivkin, 2009; Clark, Martorell, & Rockoff, 2009; Grissom & Loeb, 2009). Additionally, and of particular importance to this work, scholars and policymakers have begun to acknowledge that school principals play a crucial role in how policies and systems, including teacher evaluation systems, play out in practice (Donaldson & Papay, 2014; Halverson, Kelley, & Kimball, 2004; Koyama, 2014; Rigby, 2015). How principals interpret, communicate, and ultimately carry out new teacher evaluation policies and systems has great implications for how these policies and systems look in practice. Although principals have always assumed the responsibility of evaluating their staff, the process is now higher-stakes because in many cases new teacher evaluation policies and systems have tied a teacher's evaluation score to career defining decisions, such as hiring, firing, and tenure decisions.

How principals make sense of teacher evaluation policies and systems matters because: (1) this sensemaking process demonstrates to policymakers and researchers the ways in which policies and systems may be interacting and working in schools; (2) principals' sensemaking has the potential to highlight the unintended consequences of these policies and systems; and (3)

perhaps most importantly, studying how principals make sense of teacher evaluation policies and systems offers feedback to policymakers regarding the ultimate impact the policy is having in practice. Examining principal policy and system interpretation and implementation through the lens of cognition is a useful approach to better understand how different people charged with implementing the same policy make sense of this process. Although school leaders undoubtedly think about and ultimately implement policies and systems differently (Lipsky, 1980; Weatherly & Lipsky, 1977) focusing on the cognition of these leaders has the potential to shed light on how they make sense of and use the policies that are entering their systems of practice.

Specifically, in this dissertation I study principals' cognitive schemas, which scholars define as a cognitive framework that helps individuals interpret and organize information (Weick, 1995). Schemas are useful when studying how individuals interpret and implement policies and systems because schemas focus on how characteristics of individuals impact how they process, interpret, and make sense of new and evolving information (Spillane, 2000; Spillane & Lee, 2014; Spillane et al., 2002). Given the amount of information school leaders are asked to understand and make sense of in their environments, using the cognitive frame to examine this process has the potential to yield difficult to capture findings. Additionally, in educational research the cognitive frame is being used with increasing frequency to better understand how individuals, including principals, shape, mold, influence, ignore, prioritize, and interpret information, policies and systems.

In this study I focus on principals who have high-levels and low-levels of experience and who are facing high-levels and low-levels of external pressure. For the purposes of this dissertation, I define low-levels of experience as an individual who has held the position of school principal for four years or fewer. I define a principal with high-levels of experience as an

individual who has been a principals for at least nine years. Like teachers, research shows principals improve their practice significantly in the first five years of their practice (Leithwood, Seashore-Louis, Anderson, & Wahlstrom, 2004; Seashore-Louis, Wahlstrom, Leithwood, & Anderson, 2010). Given this research I chose to study principals with fewer than five years of experience and more than five years of experience. Additionally, I wanted to leave a gap in years of experience because I felt this would strengthen the results of this work (as opposed to studying principals with four versus five years of experience).

I define principals who face high external pressure as principals whose schools are labeled either red or orange on Michigan’s 2014 Accountability Report Card. Principals who work in schools with green, lime, or yellow ratings are considered to have low external pressure (See Table 1.1). A more detailed rationale for using this scorecard is provided later in this work. The goal of using these attributes of school leaders is to make the process of sensemaking more predictive. For example, do principals with low-levels of experience in high-pressure-environments make sense of policies and systems differently than principals with low-levels of experience in low-pressure environments? We know that sensemaking happens and impacts policy and system implementation (Coburn, 2005; Halverson et al., 2004; Rigby 2015; Spillane et al., 2002), but we do not have a predictive theory beyond saying “sensemaking happens.”

Table 1.1.

Dissertation Participants (N= 12)

Principal	High-Experience	Low-Experience
High-Pressure	3	3
Low-Pressure	3	3

Research Questions

The goal of this dissertation is to better understand how principal experience and external pressure impact how principals implement evolving teacher evaluation policies and systems. To assist in answering these questions I reviewed education policy and system implementation research, which helped me construct an analytic framework from which I developed my research questions. Specifically, this study asks the following:

- (1) How do principals' cognitive schemas (i.e., highly developed background knowledge due to experience) influence how they come to understand and implement teacher evaluation policies and systems?
- (2) What role does external context (i.e. high-pressure vs. low-pressure environments) play in shaping principal learning and enactment of teacher evaluations policies and systems?; and
- (3) In what ways, if any, do principals' experience and external pressure interact during the implementation process?

Throughout this dissertation I refer to both teacher evaluation policies and teacher evaluation systems. When referring to teacher evaluation policy I mean the specific legislation of the state of Michigan. For example, at the time of data collection, Michigan required all teachers be evaluated annually (in most circumstances), and that 25 percent of a teacher's final evaluation score be based on student growth and assessment data. Subsequent references to teacher evaluation policy (and principals' sensemaking and cognition of this policy) refers to the actual state legislated policy in place at the time of data collection. A requirement of Michigan's teacher evaluation policy was that schools must use a teacher evaluation system. At the time of data collection the four state approved systems were: Charlotte Danielson's Framework for

Teaching, the Marzano Teacher Evaluation Model, The Thoughtful Classroom, and 5 Dimensions of Teaching and Learning. Although a majority of district's opted to use one of these system, districts were allowed to choose another teacher evaluation system if this system met certain requirements laid out by Michigan's teacher evaluation policy. Any subsequent reference to a principal's thinking, use, or enactment of their teacher evaluation system refers to their specific system, and not necessarily the policy. At times I use both "policy" and "system" when discussing principals' cognition and sensemaking, but other times I intentionally use either "policy" or "system" when describing the results of this work. This is a small, but important difference to keep in mind as you continue to read this dissertation.

Contribution of the Dissertation

Although research on how cognition impacts policy and system implementation and is increasing, additional research should seek to better understand how and why specific characteristics impact how these policies and systems play out in practice. Few studies look at the impact of specific implementer characteristics and how the cognitive schemas of individuals with these characteristics impact how these individuals interpret policies and systems. By exploring these nuanced and complex ideas, this dissertation will make two primary contributions.

First, this dissertation contributes to the work of sensemaking theory by providing insights and hypotheses of how to two specific implementer characteristics—school leader cognitive schemas (particularly focusing on background knowledge developed due to experience) and the amount of external pressure facing a school impact policy and systems interpretation and implementation. Research shows cognitive schemas impact policy implementation (Halverson et al., 2002; Rigby, 2015) and external pressure impacts policy implementation (Koyama, 2014; Spillane, 2000). However, while we know these characteristics

matter for policy interpretation and implementation we lack an understanding of how and why they matter. With the goal of moving beyond the notion that “sensemaking happens,” examining these specific implementer characteristics separately constitutes key contributions to this study, but looking at how these two important characteristics interact during the implementation process is the main theoretical contribution of this work.

Second, the overarching goal of this work is to be able to describe and explain individual principal sensemaking by better understanding how individual principals with certain attributes interpret and implement teacher evaluation policies and systems. The results of this work have the potential to inform future practice as schools and districts may be able to better anticipate how certain individuals will interpret teacher evaluation policies and systems. Practitioners may be able to use the results of this work to better design teacher evaluation professional development and training for principals within their school district. Put differently, the results of this work can serve as a template for districts who can make strategic decisions and individualize training and development to best address the needs of all principals in their district.

Outline of the Dissertation

The chapters that follow describe the relevant literature, frame the research, discuss the design and methods of the study, show the findings of the work, and finally examine the implications and conclusions of this work. Chapter two presents a review of the literature, specifically focusing on the different waves of educational policy implementation in general and then moving in to a more nuanced review of policy and system implementation and cognition. This chapter then reviews literature on cognition and teacher evaluation policy and system implementation, focusing on the importance of teacher evaluations and school principals in the implementation process. Chapter three frames the research, including providing a conceptual framework of policy implementation. The theoretical framework is then discussed, as is the

importance of cognitive schemas and sensemaking theory in this work. The chapter concludes with the rationale of why experience and external pressure are important variables to look at when studying policy and system implementation in education. Chapter four discusses the design and methodology of the study, including the research questions, research design, the rationale for using qualitative methods, and for using a case study approach for this work. This chapter then discusses the site and participant selection, data collection methods, data analysis, and the validity of this data collection. Chapters five and six examine the findings of this work. Finally, chapter seven discusses the findings, focusing on the implications of and the conclusions drawn from this work.

Chapter 2: Literature Review

The purpose of chapter two of this dissertation is to describe and synthesize research that has examined education policy implementation generally as well as review research which examines how cognition and sensemaking impacts policy and system implementation. Additionally, this chapter reviews the literature on principal sensemaking and principal teacher evaluation policy and system implementation. This literature review provides context for my study by providing a historical overview of the different waves of policy implementation research, including how this research has evolved over the past several decades to sharpen its focus on how individual sensemaking impacts policy implementation. This chapter also sets up chapter three of this dissertation, which establishes the theoretical framework that guided my data collection and analysis. Finally, this chapter grounds this dissertation in a stream of research and scholarship that extends over many decades that will help situate the results of this work in the current literature on how sensemaking impacts policy and system implementation.

Part one of this literature review examines this history of education policy and system implementation, including how scholars have increasingly focused on individual cognition when examining what factors impact how policies and systems play out in practice. Part two of this review focuses specifically on research that examines how cognition and sensemaking impact teacher evaluation policy and system implementation. Finally, part three of this chapter reviews literature that examines how principals make sense of and implement policies and systems generally and how principals make sense of and implement teacher evaluation policies and systems specifically. Additionally, part three of this review examines research on how the two primary variables of this study, individual experience and external pressure, impact individual interpretation and implementation of policies and systems.

Part I: Education Policy Implementation Overview

There is a broad literature on how individuals and organizations make sense of and implement policies and reforms. Over the past fifty years scholars have attempted to better understand how and why policy implementation varies in certain contexts and with certain individuals. This has become an increasingly complicated endeavor as Honig (2006) writes, “the policies under investigation on the whole are significantly more comprehensive and varied than in previous decades” (p. 4). However, despite the complexity of studying how individuals and organizations interpret and make sense of policies and reforms, researchers generally agree on three waves of education policy implementation research (Honig, 2006; Odden, 1991).

Researchers from one wave (approximately 1960-1969 (Odden, 1991)) of educational policy implementation research generally found that the policies and reforms handed to schools and school leaders often lacked clear expectations and directions for these local implementers, resulting in policy ambiguity (Lipsky, 1980; Weatherly & Lipsky, 1977). This lack of clarity often resulted in unsuccessful attempts by individuals and organizations to implement these policies in ways envisioned by policy designers (Honig, 2006; Lipsky, 1980; Weatherly & Lipsky, 1977). Furthermore, research from wave one concluded even if local implementers understood how a policy or reform was expected to be implemented, these individuals often lacked the will and capacity to implement these policies and reforms as intended by the designers of the policy (Honig, 2006; McLaughlin, 1987; Weatherly & Lipsky, 1977). The most widely studied policy from wave one was the Elementary and Secondary Education Act (ESEA) passed in 1965. Most of the research examining the implementation of ESEA found local implementers did not faithfully attempt to implement the aspects of ESEA as envisioned by the designers (Murphy, 1971).

Wave two of policy implementation research generally found that given time, local actors often tried to implement policies with the best of intentions and if given enough time and support, in practice policies and reforms often resembled, at least in part, the original vision of policymakers (Honig, 2006). This wave of research moved beyond the idea that local policy implementers lacked the skill, will, or capacity to implement policies and reforms, acknowledging policy implementation was a more complex process. Although researchers from wave two began to acknowledge that people and places mattered greatly when examining policy implementation, this work did not explore how people and places mattered (Honig, 2006).

Eventually, scholars from wave three of policy implementation research began to examine how specific individuals and their characteristics impacted policy and reform interpretation and implementation. Additionally, these researchers began to focus on the impact of place on policy implementation (Honig, 2006). This focus on person and place began to expand throughout wave three of policy implementation research and near the end of the third wave, scholars began exploring specifically how and why interactions among people and places impacted how policies were interpreted and ultimately how these policies played out in practice (Honig, 2006).

Most recently, over the past twenty or so years, education policy implementation research has continued to sharpen its focus on how people matter, specifically how individual cognition and sensemaking impact how policy interpretation and implementation occurs (Coburn, 2001; Coburn, 2005; Halverson et al., 2004; Honig, 2006; Rigby, 2015; Spillane et al., 2002). As Honig (2006) writes, “Whereas past implementation research generally revealed *that* policy, people, and places affected implementation, contemporary implementation research specifically aims to uncover their various dimensions and *how and why interactions among these dimensions*

shape implementation in particular ways” (p. 14). Specifically this line of research has focused on how people draw on their various identities, social situations, and prior knowledge and experiences to shape how they make sense of and implement policies and reforms (Honig, 2006; Spillane, Reiser, & Gomez, 2006).

The evolution of policy implementation research sets the stage for this dissertation. Although earlier policy implementation research suggested factors such as policy design and local resistance may be the primary causes of a disconnect between policymakers and practitioners, as Spillane et al. (2006) write, “This work suggests that viewing implementation failure exclusively as a result of poor clarity or deliberate attempts to ignore or sabotage policy neglects the complexity of the human sensemaking processes consequential to implementation” (p. 47). Today, research that focuses on how cognition and sensemaking impacts policy interpretation and implementation has examined how many different groups of people make sense of policies and reforms including: (1) how teacher cognition affects policy interpretation and implementation (Booher-Jennings, 2005; Firestone, Monfils, Schorr, Hicks, & Martinez, 2004; Kennedy, 2010); (2) how principal cognition affects policy interpretation and implementation (Coburn, 2005; Halverson et al., 2004; Halverson & Clifford, 2006; Rigby, 2015); and how a host of other individuals’ cognition from central office personnel (Honig, 2006) to mayors (Hess, 2008), impacts how policies play out in practice. This line of research generally finds that how individuals make sense of policies and reforms has a strong relationship to how these individuals think about and ultimately implement policies (Spillane et al., 2002).

One prominent study by Spillane (2000) that examined school leader’s implementation of mathematics reform notes:

Cognitive science offers a number of plausible explanations for the dominant patterns in district leaders' understanding of the mathematics reforms, explanations that are not mutually exclusive. Whereas more conventional implementation accounts might focus on district leaders' attempts to sabotage the mathematics reforms or their limited capacity to carry out reformers' proposals, a cognitive frame suggests that implementation failure was due in important measure to what district leaders understood from the reforms (p. 169).

This work suggests individuals make sense of new information through their existing knowledge and beliefs and that individual sensemaking is constantly filtered and updated as new information is learned and processed (Spillane, 2000; Weick, 1995). Subsequent work by Spillane et al. (2002) makes the following argument: "What a policy means for implementing agents is constituted in the interaction of their existing cognitive structures (including knowledge, beliefs, and attitudes), their situation, and policy signals" (p. 388).

In sum, research on policy interpretation and implementation and the impact cognition has on policy interpretation and implementation has evolved from the idea that implementers either ignore or modify the wishes of policymakers due to a lack of skill or will. Instead, research today suggests the human sensemaking process plays the larger role in how policies are interpreted ultimately play out in practice. Specifically, an individual's knowledge, beliefs, context, and attitude impact policy implementation and differences in policy implementation occur when individuals make sense of a policy reform and draw connections between these new ideas and their existing understandings and knowledge (Spillane et al., 2002).

Part II: Teacher Evaluation Policy Implementation

Schools began to focus some form of attention on evaluating teachers in the early part of the 20th century. The move towards evaluating teachers was due in part to the growing belief of scholars and practitioners that it was necessary to determine teachers' impacts on student learning, including how teachers teach students to be successful and productive citizens (Cubberley, 1929). Cubberley's (1929) work detailed evaluation of teachers should include how teachers delivered instruction and how they managed student behaviors. Despite this early attempt to evaluate teacher effectiveness the evaluation of teachers was often reduced to completing a mere checklist of teacher responsibilities, such as teacher attendance, timeliness, and professionalism (Darling-Hammond, Wise, & Pease, 1983; Wise, Darling-Hammond, McLaughlin, & Bernstein, 1985). The use of a simple checklist was due in part to the growing responsibilities and demands of school administrators and these types of evaluation remained a common practice of evaluating teachers for decades.

Because of the widespread evidence on the importance of teacher quality for student outcomes (Chetty et al., 2014; Darling-Hammond, 2000) the past twenty years or so has seen a push to systematically evaluate teachers in an effort to provide better information on what makes a quality teacher. Recent research examining the implementation of teacher evaluation policies and systems has looked at many specific areas including how principals and teachers initially respond to and accept a new policies (Milanowski & Heneman, 2001), how principals communicate feedback to teachers (Kimball, 2003), the impact of teacher evaluations on principals' human capital decisions (Goldring et al., 2015; Grissom, 2011), and the relationship of teacher evaluation systems to student achievement (Chetty et al., 2014; Derrington, 2013; Donaldson, 2009; Rigby 2015). Each of these studies suggests individuals' cognition,

experience, and context, impacts how these policies and systems are interpreted and enacted in local systems of practice. More broadly, research has suggested individuals and schools struggle to implement teacher evaluation policies due to the number of interrelated components (Kennedy, 2010), which is one explanation as to why teacher evaluation policy implementation remains a challenge. Many scholars believe implementing teacher evaluation policies is becoming even more complex, as Derrington and Campbell (2013) note, “For principals as collaborative instructional leaders, new accountability-driven evaluation policies are affecting the relationships of those principals with their teachers, their sense of what being an instructional leader means, and their capacity to handle the complexities of operating and leading a school” (p. 239).

Given the complex nature of evolving teacher evaluation policies scholars have increasingly turned to studying how individual cognition and sensemaking and organizational context impact how these systems play out in practice. For example, Halverson and Clifford (2006) use a distributed cognitive theory model to show how local context shapes practice finding that cognitive systems of teacher evaluations are more complicated than envisioned by policy designers. The authors note:

Even practitioners perceived as successful implementers of standards-based teacher evaluation practices need to navigate trade-offs as they adjust the demands of the new policy artifacts to the needs of their existing contexts (Halverson et al., 2004; Kimball, 2003; Milanowski & Heneman, 2001). The tendency of teacher evaluation practices to run headlong into the traditions of local practice provides a prime opportunity to study how practitioners make sense of the new in terms of the old (p. 581).

Since the publication of this article almost 10 years ago, teacher evaluation systems have become more complex and more high-stakes, strengthening the argument for more research to better understand the implementation process. Another study that examined cognition and teacher evaluation policy implementation was conducted by Halverson et al. (2004) who found local implementers vary the implementation of teacher evaluation policies and this variation is shaped by principals' individual roles, contexts, and their specific artifact they are using.

Halverson et al. (2004) conclude:

Principal sensemaking seemed to be primarily a function of principal self-perception of their role as a leader and the knowledge and skills they bring to that role; prior evaluation practices in the school and district; and school context factors such as teacher morale and existing challenges facing the school (e.g., student population risk factors, external accountability pressures) (p. 39).

Other work by Rigby (2015) examines how cognition affects teacher evaluation policy implementation and finds first-year principals receive a variety of messages from colleagues, supervisors, and teachers about how to conduct teacher evaluations. This work finds that as principals build their professional identity they come to understand similar policies differently than their colleagues and highlights the variations in implementation of the principals asked to conduct teacher evaluations. In sum, researchers are continuing to sharpen their focus on how individual cognition impacts how policies and systems enter systems of practice. This is particularly true of teacher evaluation policies and systems, which have received unprecedented attention from scholars, policymakers, and practitioners alike. However, despite this escalation of research, researchers and policymakers agree teacher evaluation systems continue to fall short of their intended goals, including identifying high-quality teachers, distinguishing teacher

performance, and improving teacher performance through feedback and support (Kennedy, 2010; US Department of Education; Weisberg et al., 2009). This lack of progress towards providing better information on teacher quality and better identifying high-quality teachers appears to be the case for this dissertation's context as well, as in 2015, 97% of teachers in Michigan were rated as effective or highly effective and of the 96,000 teachers in the state, only 19 have been dismissed due to poor evaluations over the past five years (Michigan Department of Education, 2015). These statistics are concerning given Michigan's overall low student achievement on state assessments (Chetty et al., 2014; Darling-Hammond, 2000; Michigan Department of Education, 2015).

Because there is still much to understand about the factors that influence how teacher evaluation systems play out in practice, this study will fill an important gap in the current literature by providing insights as to how certain policy implementer characteristics influence policy interpretation and implementation. The high-stakes nature of teacher evaluation policies make it important to understand individual cognition and behavior more so than lower-stakes policies because so much is attached to these policy outcomes with respect to future teacher employment and ultimately student learning and achievement. While other policies are more transient, teacher evaluation policies are here to stay (in some form). How the people charged with making sense of and implementing these policies is of unique and great importance and must be better understood.

Part III: Principal Cognition and Policy Implementation

The third stream of research that guides this work focuses on one of the most important actors charged with implementing teacher evaluation policies and systems--school principals. Although early research suggested principals lacked the power and influence to change school

and teacher practices (Bidwell, 2001), more recent research suggests principals play a key role in how initiatives and reforms play out in practice (Coburn, 2005; Donaldson & Papay, 2014). How principals think about and implement education reforms is of particular importance as research suggests principals are second only to teachers as the educational resource who can most positively impact student outcomes, such as learning, increased attendance, and increased graduation rates (Leithwood, Seashore-Louis, Anderson, & Wahlstrom, 2004; Seashore-Louis, Wahlstrom, Leithwood, & Anderson, 2010).

The type of sensemaking in which an individual engages has great implications for how policies and systems permeate through an organization (Coburn, 2005; Weick 1995). This is particularly true with people who are in positions of leadership as leaders impact how other individuals within an organization receive, think about, and are forced to act upon a policy reform (Coburn, 2005; Ganon-Shilon & Schechter, 2016; Spillane et al., 2002). As Ganon-Shilon and Schechter (2016) note:

Leaders play an important role in shaping what and how teachers learn about educational change and reform, so school principals and middle leaders, particularly, influence teachers' sense-making both directly and indirectly. Directly, they influence what teachers find themselves making sense of, by facilitating access to some reform messages rather than others. Providing teachers with interpretive frameworks and ways of understanding reform demands, formal leaders enable the educational staff to adopt strategies that develop and construct their understanding of the reform's intent. School leaders also influence teachers' sense-making indirectly as they participate with the teachers in a collective learning process through formal meetings and informal conversations (p. 6).

Weick and Sutcliffe (2007) argue school leaders play an important role in ensuring everyone within a school can make sense of their responsibilities and as a result school leaders approach to sensemaking directly impacts how policies play out in practice (Ganon-Shilon & Schechter, 2016). For example, a study from Spillane et al. (2002) found novice principals typically prioritized establishing legitimacy with their peers and staff before trying to implement new policies and reforms and in doing so took on a form of collective sensemaking, in an effort to make teachers feel included in the policy implementation process.

The type of sensemaking in which a principal engages is particularly important to note when examining the implementation of a policy as important as teacher evaluation policies. Teachers and administrators both understand the importance of these policies, but each look at the goals, uses, and purposes of the same policy quite differently. For example, teachers look at these policies individualistically, as their careers are in large part dependent on successful evaluations (Ganon-Shilon & Schechter, 2016). On the other hand, administrators think of the teacher evaluation process holistically, with the overall success of their school constantly in the forefront of their thinking (Ganon-Shilon & Schechter, 2016). In short, while teachers and other school staff are left to their own devices of how to assign meaning to a particular policy reform, principals play a large role in guiding teacher sensemaking. Because of this, how principals chose to navigate their own sensemaking process has great implications for how teachers assign meaning to their own evaluations and ultimately how teachers are evaluated. As Spillane et al., (2002) write, “While teachers often encounter district and state accountability mechanisms through media reports, policy directives, and union newsletters (among other sources), their evolving perceptions and understanding of these policies are likely to be mediated through

participation in their school community” (p. 732). This community includes, importantly, school principals.

Distinguishing between individuals who make sense of information, policies, and reforms individually and those who engage in collective sensemaking is an important nuance to understand how policies look in practice. Although undoubtedly some of the characteristics of these two groups of individuals overlap during the sensemaking process, such as drawing on prior knowledge and experiences and current context, there is a distinction of how these groups of people think about policy implementation. Even within the sub-group of collective sensemaking one can hypothesize that there seems to be deliberate collective sensemaking, where a school principal collaborates to make sense of a policy, discuss how the policy will be implemented, etc., and informal collective sensemaking that is more along the lines of social context/network sensemaking. The later could still be very individualistic and seems to fall more along the lines of an individual influence like cognitive schema. For example, if I am a principal and have seven teacher friends, then both my cognitive schemas and my individual discussions with these seven teacher friends about teacher evaluation policy will influence how I think about the policy. This is very different then getting together seven teachers in an organization and collectively making sense of the policy. In short, how a principal engages in sensemaking likely influences how policies play out in practice.

The Co-Evolution of Teacher Evaluation Policies and the Role of Principals

One specific responsibility of the school leader is to evaluate the teaching staff, which principals have done in some form for the past century. Early research showed principals played a more hands-off role during evaluations as principals rarely evaluated classroom instruction and instead completed a checklist of teacher responsibilities, such as if a teacher showed up to work

on time (Darling-Hammond et al., 1983; Wise et al., 1985). As time progressed and principals began to observe teacher classroom instruction as part of a teacher's evaluation they often did so haphazardly, using protocols that were not supported by theory or research (Porter, Youngs, and Odden, 2001). However, as teacher evaluation systems transitioned into more high-stakes policies, principals began to take a more active role in the evaluation process and over the past several decades principals have been asked to become competent evaluators of classroom instruction and provide meaningful and critical feedback to teachers, taking on the dual role of coach and evaluator (Duke & Stiggins, 1990; Duke & Stiggins, 1986).

As their role in the teacher evaluation process became more active, the amount of time principals devoted to evaluating their staff increased drastically teachers (Halverson et al., 2004). Increasingly principals are tasked not only with observing more teacher classroom instruction, but meeting with these teachers outside of instructional time to discuss their instruction and progress. Additionally, the observation rubrics and evaluation forms principals must complete have become complex and time consuming to complete as principals must document evidence to support their claims (Halverson et al., 2004). Finally, in most circumstances, principals evaluate all teachers in their school. Because of all of these increased time demands some research has found in an effort to streamline evaluations and efficiently move through this process, principals scale back aspects of the policy, such as how long they observe teachers and the type and amount of feedback they provide teachers (Halverson et al., 2004).

Most recently, principals have been tasked with taking on the role of an instructional leader, where the principal is charged with supporting teacher instruction and is held accountable, along with teachers, for student learning (Blasé, Blasé, & Phillips, 2010; Smylie, 2010). Now, more than ever, principals are expected to be “educational experts” and understand

what good teaching and learning looks like (Blasé et al., 2010; Halverson et al., 2004). Principals are not only charged with running the school and managing aspects of a complex organization, but new teacher evaluation policy requirements ask principals to understand the ins and outs of teaching and learning. For example, principals must evaluate things such as classroom management, student engagement, and strong lesson delivery from teachers. The level of expertise expected of a principal is becoming increasingly complex each time teacher evaluation policies evolve.

Currently, as most schools use rigorous teacher evaluation systems, which typically include a student achievement based component, as well as an observational component with a detailed and structured observation rubric, the role of the principal in the evaluation process is much more defined than in previous years (Goldring et al., 2015; Steinberg & Donaldson, 2016). For example, in most situations principals are given specific directions of how and when to observe teachers, how to score teachers, and how to provide feedback to teachers (Goldring et al., 2015).

In sum, the role of the principal in teacher evaluations has changed drastically as the teacher evaluation policy landscape has evolved. Now, in most cases part of a principals' own evaluation includes student performance data and how they evaluate their staff. As a result, principals are incentivized to take a more active role in the evaluation and development of the teachers in their building. Given the research that suggests principals make many school-based decisions, including human capital decisions, based on teacher evaluation scores (Goldring et al., 2015; Jacob, 2011), recently in the field of education research much attention has been given to how principal cognition their interpretation and implementation of these important policies.

There is an increasing amount of research coming out revealing how school principals are implementing evolving teacher evaluation policies and ultimately evaluating the teachers in their building. For example, a majority of a teacher's evaluation is based on principal observations (Steinberg & Donaldson, 2016), so research has begun to document how principals evaluate teachers with more detailed and structured observational rubrics. The research overwhelming finds principals assign teachers high evaluation ratings. One explanation for this is any judgement of teaching is inherently subjective and allows the observer much leeway as to how to actually evaluate instruction (Donaldson & Papay, 2014). As a result, researchers have found principals find it difficult to separate teacher instruction and everything else they know about a teacher (for example, a teacher's contribution outside of the classroom), from their evaluation of that teacher (Donaldson, 2013; Papay & Johnson, 2012). As Donaldson and Papay (2014) note, "Although having clear standards, using highly qualified and well-trained evaluators, and focusing on evidence can help remove much of the subjective bias in observation measures, separating the personal from the professional can be difficult" (p. 2).

Additional research suggests teacher effectiveness varies substantially, yet principals' evaluations of teacher fail to differentiate this effectiveness (Grissom & Loeb, forthcoming). Research also suggests that principals tend to rate teachers more harshly in low-stakes evaluations compared to high-stakes evaluations, like those used for human capital decisions (Grissom & Loeb, forthcoming). During high-stakes evaluations, such as a teacher's official evaluation score used for many important teacher career defining decisions, such as tenure and retention decisions, principals overwhelming rate teachers highly, which may explain in part why there is so little variation documented in teacher performance.

The Role of Principal Cognition During Teacher Evaluations

Weatherley and Lipsky (1977) stressed the importance of “street-level bureaucrats” – the individuals who impact how policies are ultimately implemented. These individuals and their cognition, including their beliefs, skill, will, resources, time, context, and capacity, impact how policies look in practice. Although principals are experiencing more clarity and structure around how they are to evaluate their teaching staff, principal cognition still greatly impacts how these policies play out in practice (Coburn, 2005; Halverson et al., 2004; Spillane et al., 2002). Principals have the potential to drive school improvement through policy implementation more than most other individual actors. As Spillane and Kenney (2012) write,

While federal, state, and local government policy makers have gone to considerable lengths over the past several decades to target their policies at the technical core of schooling – specifying what teachers should teach, at times how they should teach, and acceptable levels of mastery for students – their initiatives, which represent a considerable shift in the policy environment of schools, ultimately depend on school administration for their successful implementation (p. 546).

There is a mounting pile of evidence suggesting principal cognition is mediated through individual background characteristics and local context (Coburn, 2005; Hallinger & Heck, 1996; Rigby, 2015; Spillane, Halverson, & Diamond, 2004; Spillane et al., 2002). Prior literature on principals’ cognition and sensemaking of policies provides two broad strands of findings including (1) principals’ prior experiences greatly influence how they understand and make sense of new policies (Harris, Ingle, & Rutledge, 2014; Jacob & Lefgren, 2008; Nelson, Sassi, & Grant, 2001); and (2) principals implement policies based on what they believe is in the best interest of their local school and context (Coburn, 2005; Cohen & Hill, 2001; Koyama, 2014;

Matsumura & Wang, 2014; Spillane et al., 2002). Additional research focusing on school leader cognition looks at how these leaders interpret, make sense of, and communicate policy messages they receive in their local context, finding principals receive and deliver the same policy messages differently and this impacts how policies are implemented in their local contexts (Anagnostopoulos & Rutledge, 2007; Coburn, 2005; Rigby, 2015). Several studies have looked at how cognition impacts teacher evaluation policy and system implementation, concluding principals navigate trade-offs and adjust and negotiate the demands of evaluating teachers in their building, based on their prior knowledge and personal context (Halverson & Clifford 2006; Halverson et al., 2004; Rigby, 2015).

Experience and Context: Why and How They Matter for Policy Implementation

Research shows as teachers gain experience they become more effective at raising student achievement, increasing student attendance, and managing their classroom (Clotfelter, Ladd & Vigdor, 2006; Papay & Kraft, 2014; Rockoff, 2004). Evidence also suggests principals become more effective at their jobs as they gain experience and this is true particularly in their first three years as a school leader (Clark, Martorell, & Rockoff, 2009). Research also suggests the length of a principals' tenure at one school, no matter how long they have served as a principal, impacts their role. This line of research suggests it takes five years for a principal to secure relationships with staff, improve staff effectiveness, fully implement policies and practices, and make significant education improvements (Coburn, 2001; Seashore-Louis et al., 2010).

Research has long documented that prior knowledge aids learning by enabling the learner to make connections and thus deepen their understanding (Harris et al., 2014; Jacob & Lefgren, 2008; Nelson et al., 2001). This line of thinking is true for principals and their work. For

example, some research suggests teachers' and administrators' prior knowledge and experience influence their ideas about changing instructional practice (Cohen & Barnes, 1993; Halverson & Clifford, 2006). Research specifically on principals' suggests principals' experience greatly influences how they understand and make sense of new policies and reforms (Harris et al., 2014; Jacob & Lefgren, 2008; Nelson et al., 2001). This line of research suggests school leaders build mental models that shape what they think about when receiving new information and how these individuals perceive this information (Halverson et al., 2004). As principals continue to gain experience these models shape what individuals notice when encountering new reforms and policies which impacts how principals interpret information, accept or reject new ideas or information, and how principals think about and ultimately enact policies and reforms in their systems of practice (Cohen & Barnes, 1993; Halverson et al., 2004).

Although prior knowledge and experience may be the single most important variable that impacts individual cognition, another important variable is current situational context. Specifically, the amount of outside pressure applied on individuals who are charged with interpreting and implementing new and evolving reforms greatly impacts how individuals think about this process (Grissom, 2011; Hill & Barth, 2004). For example, research suggests when individuals are introduced to and attempt to implement a new policy, individual actors' behavior changes when outside pressures enter their environment, which impacts how these individuals attempt to implement policies and reforms (Grissom, 2011; Hill & Barth, 2004).

There is a growing amount of research that examines specifically how external pressure, from federal, state, and local levels, impacts how principals conduct their work and more specifically, how they make sense of and implement policies (Booher-Jennings, 2005; Coburn, 2005; Halverson & Clifford, 2006; Matsumura & Wang, 2014; Rigby 2015). Specifically,

research shows external pressure impacts how principals reallocate instructional time (Diamond & Spillane, 2004; Firestone et al., 2004), use school-based resources (Dee, Jacob, & Schwartz, 2013), staff subject areas (Booher-Jennings, 2005), and even alter school lunches (Figlio & Winicki, 2003). Although much of this prior research has focused on attempts to raise student achievement, less work documents how external pressure impacts how principals work with and evaluate their teaching staff (an approach that has the potential to raise student achievement, given the wide-spread belief in the importance of teacher quality).

Although schools and school leaders have always been responsible for outside policy implementation, the passage of NCLB in 2001 increased accountability and external pressure placed on schools. As Koyama (2014) writes, “Principals are mediators between external accountabilities, like those of NCLB, and school practices” (p. 283). Outside pressure matters for school principals, particularly new principals who are under pressure of increased accountability (Spillane & Lee, 2014). As Halverson et al. (2004) write,

The sense we make of new information is also shaped by our social and situational context (Greeno, 1998). Organizations and institutions routinize existing models through policies, programs, and traditions. Thus, the intended effects of innovations are not necessarily altered by the malice or laziness of implementers, but instead by the best efforts of local actors seeking to satisfy conflicting goals (Spillane, Reiser, & Reimer 2002, Fischhoff 1975; March & Simon, 1958). Actors make sense of new practices within their existing social and situational context, and often adjust the meaning of the new in terms of their established context of meaning (p. 4-5).

Research that has examined how outside pressures impact how principals make sense of and implement policies finds principals mediate external district and state accountability policies

and demands in strategic ways, including “gaming the system” and implementing policies in ways consistent with local values and beliefs (Spillane & Keeney, p. 18). Additionally, research finds that principals respond to external pressures by restructuring the formal routines of their local instructional programs (Koyama, 2014). Work by Seashore-Louis and Robinson (2012) finds when external pressures and policies align with the values and perspectives of principals they will internalize these policies and attempt to implement them faithfully, but when external policies and demands do not align with a school leader’s vision they are less likely to make this effort (p. 42-43).

Federal level pressure has increased immensely over the past several decades and particularly since the passage of NCLB in 2001. Specifically, schools have faced an increase of federal pressure from things including what should be taught, how much of it should be taught, how long it should be taught, and teacher quality and evaluation (Spillane & Kenney, 2012). Research suggests that federal level pressure can impact how principals make sense of and implement policies including what teachers teach (typically focusing on math and reading while deemphasizing other subjects) and how long they teach these subjects as well as increasing the amount of time and resources devoted towards preparing for tests (Booher-Jennings, 2005; Diamond & Spillane, 2004; Spillane & Keeney, 2012).

Pressure from the state level also impacts how principals make sense of and implement school policies. Particularly since the passage of the Elementary and Secondary Education Act (ESEA) in 1965 and NCLB in 2001, states have experienced an increase of responsibilities and resources which expanded state level control of schools (Spillane & Kenney, 2012). Although the federal government can ask things of state and local schools they ultimately depend on state and local governments to develop and implement policies that are in line with federal

requirements (Spillane & Kenney, 2012). In most circumstances state level pressure comes from annual testing, school ratings, and teacher evaluations. Research shows state level pressure impacts principal decision making with things such as hiring and firing of teachers (Cohen-Vogel, 2011). The majority of pressure school leaders face is from the state level and this pressure shapes implementation of state level policy (Booher-Jennings, 2005; Coburn, 2005; Matsumura & Wang, 2014).

Finally, local level pressure also has the potential to shape how principals implement policies. Even as federal and state level pressure has increased, there has not been a decline in local level policy making (Spillane & Keeney, 2012). Local level pressure is the most pressing type of pressure administrators face, as they constantly have to meet with district administrators to show they are in compliance with district level goals and policies. Much of the prior research on how local level pressure impacts how principals make sense of and implement policies has focused on how administrators respond to outside pressure to raise student test scores, finding principals become more “data-driven” and make strategic decisions about who to teach and where to devote resources to specific students (Booher-Jennings, 2005). Booher-Jennings (2005) found that local school administrators responded to institutional pressure by emphasizing a singular measure of accountability (student achievement on test scores). Additionally, Booher-Jennings (2005) finds that schools make intentional decisions about their resources to help students on the “bubble” including providing these students more one on one or small group time, providing after school programs to these students, moving special area teachers (e.g. music, art, and gym) to teach test preparation activities, and providing these students with access to summer school (p. 241-242).

Gap in the Literature

Although the amount of research on policy implementation is growing rapidly this study aims to fill two specific gaps that currently exist. First, the research above suggests individual cognition greatly impacts how policies play out in practice. However, research currently lacks a more predictive form of how cognition impacts how individuals make sense of teacher evaluation policies. For example, do school principals with high-levels of experience make sense of and ultimately implement teacher evaluation policies differently than their less experienced peers? If so, what does this look like and what does this say about how principals evaluate teachers? Although the research on teacher evaluation policies and how principals are interacting with these policies is growing, there is a gap in the literature as to how exactly principals with different experience levels and facing different accountability pressures think about and ultimately enact teacher evaluation policies. My dissertation will fill this gap in the literature by addressing what factors impact principals' sensemaking of these policies and systems and answer what this ultimately means for how these policies and systems look in practice. In the end my goal is to be able to state hypotheses about how principals with certain characteristics think about and ultimately implement teacher evaluation policies.

The second gap in the literature this study aims to fill is a lack of focus on the sensemaking of school principals and how this sensemaking impacts policy implementation. Research on educational policy implementation has focused heavily on teachers and how a teacher's practice can impact policy implementation. This line of work sees teachers as the individuals who bear the most responsibility for what happens in their individual classrooms (e.g. if students are learning). As such, policy has often focused on teacher accountability mechanisms. However, we know very little about how principals' cognitive schemas and the

context within which they work impacts their understanding and implementation of policies; particularly teacher evaluation policies. Where teachers' understanding and willingness to implement a policy may impact an individual classroom, a school leader's cognition and sensemaking impacts how policies enter and diffuse throughout the entire school building (Coburn, 2005; Derrington, 2013). Prior research has shown even the most diligent school leader and policy implementers adjust their sensemaking based on their local context and the meaning they give to a policy (Coburn, 2005; Halverson et al., 2004; Rigby, 2015). Given the uncertain nature of policy implementation and the ever expanding policy interpretation opportunities handed to school leaders, additional research should begin to understand how school leader characteristics affect their interpretation of policies and ultimately how this interpretation impacts policy implementation. Because education policies are implemented very differently in different contexts with different individuals, it is important to move beyond preconceived notions of policy implementation and begin looking at specific answers to questions of how and why policy implementation is executed in certain contexts with certain types of people.

Chapter 3: Framing the Research

The purpose of this chapter is to describe cognitive schemas and sensemaking theory as well as to explain why these two sub-categories of cognitive theory are appropriate and useful lenses through which to view how principals think about and ultimately implement teacher evaluation policies and systems. Part one of chapter three examines cognitive schemas, an idea which derives from cognitive science and which I define as the pattern of how individuals think about collecting, organizing, and processing information (Piaget & Inhelder, 1958). This review will examine how cognitive research uses cognitive schemas generally and within the field of education specifically. Part two of this chapter focuses on sensemaking theory, including distinguishing between individual and collective sensemaking and making the argument for why sensemaking theory is best suited to help guide this work. One's cognition, including their cognitive schema(s), impacts their subsequent sensemaking of a task or event. In this way, one's cognitive schema(s) and an individual's sensemaking work together and form the basis of the framing of this work.

Researchers define cognitive schemas as specific knowledge structures individuals use to make sense of information and interpret this information in their environment (Piaget & Inhelder, 1958; Spillane et al., 2006). Sensemaking theorists believe past experiences and prior knowledge shape an individual's learning and acknowledge that learning occurs through our social and situational context (Greeno, 1998; Weick, 1995). In this way, these two cognitive frameworks intersect and are useful when trying to explain the phenomenon of how individuals attempt to process, understand, interpret, and implement policies and systems. Research that studies individuals' cognitive schemas and uses sensemaking theory to explain policy and system implementation suggests that even if individual actors receive the same message regarding how a policy should be implemented, these individuals will construct different interpretations of this

message based on what they already know and believe (Grider, 1993; Halverson et al., 2004; Spillane et al., 2006; Weick, 1995). This past work suggests that studying policy implementation through the lens of sensemaking can explain how principals make sense of policies and help explain how principals enact and make decisions around implementing policies and systems.

Cognitive Schemas

The origins of cognitive theory research date back to the late 1800s and the work of William James who believed human thinking consisted of non-repetitive thoughts that continually evolved as new information and experiences entered their thinking (James, 1890). Research using cognitive theory grew in the early part of the 20th century with the work Wilhelm Wundt who found that human experiences consist of measurable mental functions including awareness, perception, and reaction (Wundt, 1902). Wundt found that as individuals have more experiences, the range of ways they make sense of these experiences increases, because these individuals have more information to draw on when attempting to make sense of these experiences (Wundt, 1902). Research focusing on how individual's receive, interpret, and act upon information continued to grow with the work of John Dewey who argued understanding one's cognition was imperative to understanding the actions of that individual (Dewey, 1938). As cognitive theory continued to gain prominence, the idea of cognitive schemas developed. This work largely began in the 1930s with Frederic Bartlett, who introduced the concept of the cognitive schema (Bartlett, 1958). Bartlett conducted research focusing on how individuals interpret, remember, and make sense of full and incomplete information. As Grider (1993) writes:

Another of Bartlett's classic experiments involved the relaying of a story from person to person. When the story reached the tenth individual it had virtually become an entirely

different tale from the original version. The people had unknowingly changed segments of the story to fit their expectations (i.e. existing schemas) (Bell-Gredler, 1986). Bartlett's findings helped to develop and ungird the key cognitive concepts of perception and mental processing (p. 9).

Finally, Jean Piaget contributed greatly to the field of cognitive theory with his work on how individuals collect and organize information (Piaget, 1964). Like Bartlett, Piaget emphasized the importance of schemas in cognitive development, arguing cognitive schemas help individuals create mental representations and when linked together help one understand the world and respond to situations (Piaget, 1964). Both Bartlett and Piaget found that as individuals learn, they create cognitive schemas that help them code, process, understand, and respond to information (Bartlett, 1958; Piaget, 1964). In general, cognitive theorists, including Bartlett and Piaget, believe individuals use cognitive schemas to sort information in long-term memory, as well to understand new information that enters their system of thinking (Bartlett, 1958; Grider, 1993; Piaget, 1964).

Currently, research using cognitive theories to study individual and organizational behaviors is expanding. This expansion is due in part to the interest in examining how individuals make sense of outside policies and interventions entering their organizations and systems of practice. Early research on how individual cognitive schemas impact how policies and interventions play out in organizations suggests that when an individual encodes new information, their already existing cognitive schema mediates how new information is received, organized and processed (Bartlett, 1958; Dewey, 1938; Piaget, 1964). Essentially, one uses previous knowledge to interpret new ideas. In this process, individuals sometimes even change new information to fit his or her existing cognitive schema (Grinder, 1993; Piaget, 1964). This is,

in part, why new policies or programs often are implemented in different ways even when local actors are working earnestly to faithfully implement the policy or program.

Like many organizations, schools and school districts are experiencing an influx of new policies entering their systems of practice, resulting in individuals at some level having to make sense of this information (Honig, 2006; Spillane & Kenney, 2012). As a result, research has begun to examine how individual organizational members' cognitive schemas impact how policies and outside interventions play out in schools. In educational research, Spillane et al. (2006) defined cognitive schemas as "specific knowledge structures that link together related concepts used to make sense of the world and to make predictions" (p. 49). For example, an experienced school leader will draw on his or her developed schema when attempting to make sense of good classroom instruction. This school leader's past experiences influence what this leader expects to see in the classroom and ultimately impacts how they interpret or understand what is happening in the classroom. Individuals have varying cognitive schemas which is why focusing on this aspect of cognition is an appropriate way to address questions of policy and system implementation. Even if individual actors all receive the same message regarding a reform, individuals will construct different interpretations of this message based on what they already know and believe (Grider, 1993; Spillane et al., 2006).

This process is not unidirectional, however. Just as existing schemas shape how new information is processed, so too does new information shape existing schemas. When confronted with new ideas that challenge preexisting understanding, individuals update their schema to reflect this new knowledge. In a seminal study in educational research that examines how a leader's cognitive schema impacts policy implementation Halverson et al. (2004) note:

Our cognitive models, however, are not rigid structures that determine what we notice and name. Rather, our models interact with our perceptions and experience in an iterative process through which new experiences can come to shape our existing models. In organizations, new policies and programs can provide this jolt to existing practice, encouraging practitioners to reframe their practice in terms of the new expectations. The ways that practitioners make sense of new initiatives in terms of pre-existing models make the implementation of new, complex programs a far from linear and predictable process (p. 5).

To be clear, each individual may have a unique schema. However, this is not to say that generalizations are impossible. In particular, there are some common factors that can be used to hypothesize about how new information is likely to be interpreted by individuals who share certain key characteristics. For example, the extent of one's prior experience has been shown to be a key determinant shaping cognition. Experience shapes how individuals learn and enact new policies in multiple ways. First, prior knowledge and experience shapes what individuals notice when conducting a process (Weick, 1995). In this way, we might predict that more experienced school principals will have an easier time conducting teacher evaluations because they have experience doing this in the past. Even if the policy they are implementing is different, their past experiences can shape things such as how they communicate information, what they notice during observations of teachers' instruction, and how they might navigate the process of evaluating teachers. This line of research suggests school leaders build mental models based on past experiences and these models impact what they notice and how they enact new versions of old policies (Halverson et al., 2004).

However, other research suggests individuals who are familiar with the task at hand will

implement new tasks in old, familiar way. Therefore individuals with more experience might be less likely to implement new policies faithfully, instead transforming a new policy into iterations which they are familiar with and that makes sense to them. It is well established that as principals gain experience these mental models shape what individuals notice when encountering new reforms and policies which impacts how principals interpret information, accept or reject new ideas or information, and how principals think about and ultimately enact policies and reforms in their systems of practice (Cohen & Barnes, 1993; Halverson et al., 2004). Therefore, individual experience is likely to have some impact on individual sensemaking and specific to this study, how principals think about and ultimately evaluate the teachers in their building.

Second, the degree of pressure one feels when trying to learn something new shapes how our preexisting schemes shape new information (Grissom, 2011; Hill & Barth, 2004). As was mentioned in the previous chapter, research suggests principals make sense of and implement policies as they attempt to mediate external district and state accountability policies and demands and they do so in strategic ways, including “gaming the system” and implementing policies in ways consistent with local values and beliefs (Spillane & Keeney, p. 18). Therefore, it is logical to predict that the amount of pressure facing school leaders may lead to some predictable ways individual principals think about and ultimately evaluate teachers. For example, we might imagine principals with large amounts of accountability pressure from the state level may be more likely to implement a new teacher evaluation policy with fidelity than their peers who work in environments with fewer pressures.

In sum, cognitive schemas are key to consider when studying questions of policy and system implementation in order to understand how individuals make sense of new and existing information and situations. Using the lens of one’s cognitive schemas moves past the assumption

that “sensemaking happens” and instead has the potential to examine questions related to why sensemaking happens and how individual characteristics affect the policy implementation process (Halverson et al., 2004; Spillane et al., 2006).

Sensemaking Theory

There is a growing body of literature in education that uses sensemaking theory, a specific type of cognitive theory, to address questions of how people and organizations interpret and implement policies and reforms (Coburn, 2005; Halverson et al., 2004; Rigby, 2015; Spillane et al., 2002). The goal of much of the prior research using sensemaking theory has been to attempt to explain how individual and organizational sensemaking impacts how policies look in practice. Specifically, research that uses sensemaking theory has examined how individuals come to understand and enact policies and how this process is influenced by prior knowledge, the social context within which they work, and the nature of their connections to the policy or reform message (Coburn, 2005; Cohen & Hill, 2001; Spillane et al., 2002).

Sensemaking theorists believe past experiences and prior knowledge shape individual and collective learning and this learning occurs through our social and situational context (Greeno, 1998; Weick, 1995). Sensemaking theory seeks to understand how people process, understand, and respond to change (Halverson et al., 2004; Spillane et al., 2002; Weick, 1995) and attempts to explain how and why social learning occurs (Weick et al., 2005). When there is a mismatch between what an individual expects and what an individual experiences, individuals are left to assign meaning to what has happened (Ganon-Shilon & Schechter, 2016) and sensemaking helps rationalize these experiences (Weick, 1995). As Ganon-Shilon and Schechter (2016) note:

Structuring the unknown through sense-making enables individuals to act in ways that make sense. It involves coming up with a map of a shifting world as well as testing this

map with others through data collection, conversation, and action. Individuals, then, actively construct meaning by relating new information to preexisting cognitive frameworks labeled by scholars as working knowledge, cognitive frames, enactments or cognitive maps. (p. 4).

Sensemaking theory is particularly useful when attempting to answer questions of how individual actors' attempts to reconcile conflicting policy demands and implement policies and systems. For example, we might imagine school principals are faced with conflicting demands of how to evaluate teachers. Should the principal use the teacher evaluation policy as a means to hold teachers accountable for their performance, rank these teachers, and either award or dismiss these teachers based on these ratings? Or should principals use teacher evaluations as a means of support and feedback in an effort to help teachers improve their instructional practice? Principals are faced with these scenarios and must decide how they will think about the teacher evaluation process. The multiple paths one may take while making sense of a new and evolving policy is one reason why sensemaking theory provides another critical lens to analyze these data.

In short, where one's cognitive schemas impact how they think about implementing teacher evaluation policies, sensemaking theory is useful to study the factors that actually impact how this thinking plays out in practice. For example, a principal will come into an observation of teacher instruction with prior knowledge that will impact his or her focus (imagine a former math teacher who is focused on the instructional strategies of a math teacher). However, what differentiates this existing schema from the sensemaking frame is sensemaking theory includes the factors which influence the actions this principal will take during the observation. Therefore, we can hypothesize that a principal with a preexisting cognitive schema that prioritizes clear and concise mathematical instruction will focus on this during the observation of a teacher, but how

this principal ultimately rates this teacher is influenced by other contextual factors, which impacts the sense this principal will make, regardless of their cognitive schema. Weick (1995) argues there is a strong reflexive component of sensemaking that is particularly useful as people are navigating their way through a process, making sense of this process, and then updating their sensemaking as they make further sense of the ongoing process (p. 15). Weick et al. (2005) write:

Explicit efforts at sensemaking tend to occur when the current state of the world is perceived to be different from the expected state of the world, or when there is no obvious way to engage the world. In such circumstances there is a shift from the experience of immersion in projects to a sense that the flow of action has become unintelligible in some way. To make sense of the disruption, people look first for reasons that will enable them to resume the interrupted activity and stay in action (p. 409).

This description of sensemaking fits squarely into this study's focus on policy and system implementation in schools. Individuals within schools, in this case school principals, expect to experience a certain event or go through a certain process when evaluating teachers. For example, principals with high-levels of experience have evaluated teachers in so form for their entire careers as a principal. Therefore, these past experiences shape how principals think about this process, including how they observe teacher instruction, how they communicate with teachers about the evaluation process, and how to use the results of evaluations. Principals with low-levels of experience also have some understanding of how teacher evaluations might look in practice as the vast majority have recently left the classroom where they were evaluated as a teacher (this is the case for the six principals in this study with low-levels of experience). What these principals expect to see during the evaluation process comes from their existing schema.

However, new teacher evaluation policies are disrupting what principals expect to see by introducing new ideas, concepts, routines, and expectations of what teacher evaluations should look like. This creates an opportunity for principals to make sense of this disruption.

Individual vs. Collective Sensemaking

Within the theory of sensemaking there are two general schools of thought regarding how individuals make sense of information. One is individual sensemaking where individuals make sense of unfamiliar situations on their own by relying on their own personal experiences, beliefs, and values in an effort to bring clarity to an uncertain situation (Ganon-Shilon & Schechter, 2016; Klein, Moon, & Hoffman, 2006). These individuals typically create mental models based on their previous and current cognition in an effort to explain uncertainties in their environment (Ganon-Shilon & Schechter, 2016). The individual vein of sensemaking theory comes from the broader cognition literature described above including how individuals' cognitive schemas influence how individuals make sense of unclear or ambiguous situations (Bingham & Kahl, 2013; Fiss & Zajac, 2006; Maitlis & Christianson, 2014). Individual sensemaking suggests individuals make sense of situations individually based on their personal cognitive schemas, which are constantly evolving as they receive new and updated information (Maitlis & Christianson, 2014).

The other vein of sensemaking theory is collective sensemaking. Collective sensemaking is rooted in studies of social interaction, which argue sensemaking occurs between individuals rather than within one individual (Maitlis & Christianson, 2014). Individuals who engage in collective sensemaking rely not only on their individual thoughts, beliefs, and experiences, but also on the thoughts, beliefs, and experiences of other individuals within their environment. This results in a shared social process of sensemaking (Ganon-Shilon & Schechter, 2016; Weick et

al., 2005). Scholars of collective sensemaking believe in a co-constructed sensemaking process between the people within an organization (Maitlis & Christianson, 2014). In recent research, sensemaking is becoming more widely acknowledged as a social process. Some scholars argue even when individuals act on and interpret information by themselves this individual sensemaking is most often embedded in a social context where individuals thoughts, feelings, and behaviors are influenced by other people within their social context (Maitlis and Christianson, 2014; Weick et al., 2005).

As Maitlis and Christianson (2014) write:

When sensemaking is seen as taking place within individuals, then collective meaning making occurs as individuals advocate for a particular view and engage in influence tactics to shape others' understandings. In contrast, when sensemaking is regarded as unfolding between individuals, intersubjective meaning is constructed through a more mutually co-constituted process, as members jointly engage with an issue and build their understanding of it together (p. 78).

In short, while individual sensemaking occurs in one's head, collective sensemaking occurs among multiple people in an organization or an environment. Based on these definitions one might expect people to make sense of information, events, and processes differently based on the type of sensemaking with which they engaged. For example, do people who engage in individual sensemaking make sense of information in a similar or different way than those who engage in collective sensemaking? For those who engage in collective sensemaking, is how they make sense of a situation different based on who they collectively make sense with—such as collective sensemakers who make sense in structured professional developments with their staff

versus collective sensemakers who draw on their informal networks with their close friends or certain teachers/groups?

The type of sensemaking with which an individual engages has great implications for how policies permeate through an organization (Coburn, 2005; Weick 1995). This is particularly true for people who are in positions of leadership as leaders impact how other individuals within an organization receive, think about, and are forced to act upon a policy reform (Coburn, 2005; Ganon-Shilon & Schechter, 2016; Spillane et al., 2002). As Ganon-Shilon and Schechter (2016) note:

Leaders play an important role in shaping what and how teachers learn about educational change and reform, so school principals and middle leaders, particularly, influence teachers' sense-making both directly and indirectly. Directly, they influence what teachers find themselves making sense of, by facilitating access to some reform messages rather than others. School leaders also influence teachers' sense-making indirectly as they participate with the teachers in a collective learning process through formal meetings and informal conversations (p. 6).

Weick and Sutcliffe (2007) argue school leaders play an important role in ensuring everyone within a school can make sense of their responsibilities and as a result school leaders approach to sensemaking directly impacts how policies play out in practice (Ganon-Shilon & Schechter, 2016). For example, a study from Spillane et al. (2002) found novice principals typically prioritized establishing legitimacy with their peers and staff before trying to implement new policies and reforms and in doing so took on a form of collective sensemaking, in an effort to make teachers feel included in the policy implementation process.

The type of sensemaking in which a principal engages is particularly important to note when examining the implementation of a policy as important as teacher evaluation policies. Teachers and administrators both understand the importance of these policies, but each look at the goals, uses, and purposes of the same policy quite differently. For example, teachers look at these policies individualistically, as their careers are in large part dependent on successful evaluations (Ganon-Shilon & Schechter, 2016). On the other hand, administrators think of the teacher evaluation process holistically, with the overall success of their school constantly in the forefront of their thinking (Ganon-Shilon & Schechter, 2016). In short, while teachers and other school staff are left to their own devices of how to assign meaning to a particular policy reform, principals play a large role in guiding teacher sensemaking. Because of this, how principals chose to navigate their own sensemaking process has great implications for how teachers assign meaning to their own evaluations and ultimately how teachers are evaluated. As Spillane et al., (2002) write, “While teachers often encounter district and state accountability mechanisms through media reports, policy directives, and union newsletters (among other sources), their evolving perceptions and understanding of these policies are likely to be mediated through participation in their school community” (p. 732). This community includes, importantly, school principals.

Distinguishing between individuals who make sense of information, policies, and reforms individually and those who engage in collective sensemaking is an important nuance to understand how policies look in practice. While undoubtedly some of the characteristics of these two groups of individuals overlap during the sensemaking process, such as drawing on their prior knowledge and experiences and current context, there is a distinction of how these groups of people think about policy implementation. Even within the sub-group of collective sensemaking

one can hypothesize that there seems to be deliberate collective sensemaking, where a school principal collaborates to make sense of a policy, discuss how the policy will be implemented, etc., and informal collective sensemaking that is more along the lines of social context/network sensemaking. The latter could still be very individualistic and seems to fall more along the lines of an individual influence like cognitive schema. For example, if a principal has seven teacher friends, then individual discussions with these seven teacher friends about teacher evaluation policy has the potential to influence how this principal thinks about the policy. This is very different than getting together seven teachers in an organization and collectively making sense of the policy. In short, how a principal engages in sensemaking likely influences how policies and systems play out in practice.

The Usefulness of Sensemaking

Sensemaking theory is a useful approach to study questions of how principals implement teacher evaluation policies for several reasons. First, Weick (1995) describes sensemaking as distinct from other approaches, such as social action theory, instructional leadership theory, principal-agent theory, and organizational/institutional theory, because sensemaking is most useful when looking at a sustained activity or an ongoing process (p. 13). Principals are constantly participating in the “sustained activity” of evaluating teachers throughout the school year. They are doing so both formally, through their district’s teacher evaluation process and informally through conversations, walkthroughs, and other ways they collect data on teachers. This process and their sensemaking of this process is ongoing and likely changes as they become more familiar with teacher evaluation policies and have experience conducting more evaluations. Consequently, the way a principal approaches his or her first or second teacher observation will likely differ from his or her fifteenth or sixteenth teacher observation. Observations of the same

teacher at different points of the school year may also vary. For example, a teacher will likely experience a different evaluation process from his or her principal in September when compared to the process that ensues in May as principals are likely to understand these policies better, or at least differently, as the year progresses.

Additionally, Weick (1995) argues there is a strong reflexive component of sensemaking that is particularly useful as people are navigating their way through a process, making sense of this process, and then updating their sensemaking as they make further sense of the ongoing process (p. 15). This is useful in asking questions about how principals evaluate teachers, because principals likely do not completely understand new policies the first time they come in contact with them. Instead, principals gain additional, new, or different knowledge as they become more familiar with these policies and these newfound insights impact their sensemaking process. The ongoing nature of teacher evaluations and how principals make sense of these policies fits squarely into the sensemaking theory framework. As the roles and responsibilities of principals continue to change as teacher evaluation policies evolve, sensemaking theory serves as a useful framework to look at the ever-changing expectations of principals and how principals interpret their new roles. For example, because teacher evaluations are high-stakes, what are the consequences for teachers as principals learn to use these systems? How much space is there in these policies for principals learning and how does this impact teachers who are evaluated by principals early in the school year or the first time the principal uses this system compared to teachers who are evaluated by the same principal who has gained experience using these systems?

Sensemaking theory is also a useful approach to study how outside interventions change an existing model (Halverson et al., 2004; Weick, 1995). Research suggests one important factor

of principal interpretation of new teacher evaluation policies is how and what principals understand from prior teacher evaluation policies (Halverson et al., 2004). Therefore, as new teacher evaluation policies and evaluative requirements permeate the walls of schools, sensemaking theory provides a useful lens to study how principals interact with the changes. As opposed to other theories that look more closely at the impact that an entirely new policy has on an organization, sensemaking theory helps explore what is likely to occur when new iterations to an existing policy enter a system of practice. Almost all school principals were previously evaluating teachers and as a result they already have an idea of what teacher evaluation looks like. Therefore, as new policies aim to change systems of teacher evaluation, sensemaking theory will serve as a useful approach to better understand how changes to previously existing policies impact how these policies play out in different contexts.

Finally, sensemaking theory is a particularly useful approaching when looking at texts, written language, and artifacts (Weick, 1995). The most closely aligned study to principals' sensemaking of teacher evaluations comes from Halverson et al. (2004) who found that the potential effectiveness and usefulness of an educational artifact, such as a new teacher evaluation system, is dependent on how principals filter their understandings through pre-existing knowledge and structures (p. 38). As Halverson et al. (2004) write:

Affordances are an actor's perception of the ways the artifact can be used in practice. The actual use of a complex artifact, such as a teacher evaluation policy, depends not only on the features built into the design of the artifact, but also on affordances of artifact use perceived by actors (p. 6).

These affordances, such as whether principals look for student engagement, classroom management, or strong lesson delivery, are likely to be widely different depending on the person

who is interacting with the artifact (Halverson et al., 2004). Therefore, a sensemaking approach is uniquely suited to examine how various individuals perceive the same document and how these documents play out in different contexts. This is particularly useful when studying how principals use teacher evaluations, as most recent teacher evaluation reform has focused on creating a less subjective, more standardized way to evaluate teachers and sensemaking theorists argue this is unlikely to occur.

In sum, sensemaking theory provides a unique lens to study how principals interpret and implement teacher evaluation policies. Past studies confirm principals have an important place in policy implementation and how principals make sense of policies impacts not only their implementation but also all people with whom these policies come in contact. Although a growing body of literature has provided data on many important questions around this topic, there is a lack of scholarship documenting how principals with specific cognitive schemas make sense of new and evolving teacher evaluation policies.

Chapter 4: Research Design and Methodology

The purpose of this chapter is to describe the research design of this dissertation and provide rationale for a case study method. Additionally, I describe the context of this study and provide details and rationale explaining why Michigan is a timely state to examine how principals make sense of and implement teacher evaluation policies and systems. I then introduce the participants of this study and explain my sampling strategy, my data collection methods and sources, and my approach to data analysis. The chapter concludes by describing how I established validity for the results of this work.

Research Design and Research Questions

The goal of this dissertation is to better understand how principals' experience and external pressure impact how principals implement evolving teacher evaluation policies and systems. Specifically, this study answers the following:

- (1) How do principals' cognitive schemas (i.e., highly developed background knowledge due to experience) influence how they come to understand and implement teacher evaluation policies and systems;
- (2) What role does external context (i.e. high-pressure vs. low-pressure environments) play in shaping principal learning and enactment of teacher evaluations policies and systems; and
- (3) In what ways, if any, do principals' experience and external pressure interact during the implementation process?

To assist in answering these questions, I relied on decades of policy implementation research, specifically focusing on teacher evaluation policy implementation. Grounding my

analysis in previous research helped me construct an analytic framework. The data analysis that follows assisted me in describing and explaining the data collected for this study.

Rationale for a Case Study

This dissertation took on the design of a case study, as case studies have proven to a good design to understand how multiple variables interact in an environment (Derrington, 2013; Halverson & Clifford, 2006; Miles, Huberman, & Saldana, 2014). For example, variables such as environmental context, principal and teacher knowledge and skills, local, state, and federal accountability measures, and pressures from the district office or parents all come together and interact within the school environment. These variables create a complex environment in which a case study research approach has the potential to better understand the impact of these variables on organizational processes and policy implementation. Additionally, a qualitative research design is best suited to help answer these research questions because good qualitative research does the following: (1) takes place in natural settings in an attempt to make sense of or interpret a phenomenon (Denzin & Lincoln, 2003); (2) is grounded in the lived experiences of people (Marshall & Rossman, 1999); and (3) asks questions about how one variable interacts with another variable and why these variables act the way they do (Maxwell, 2005). Finally, according to Yin (2013), case studies are a preferred approach to answering “how” and “why” questions regarding a particular phenomenon.

In its most general terms, a case study analysis takes on the form of in-depth data collection in an effort to compare a similar phenomenon across different contexts (Patton, 2014). The goal of case study research is to collect comprehensive, systematic, and in-depth information about each case of interest (Patton, 2014). Three major types of case studies are commonly used for social science research; (1) exploratory case studies (used to help the researcher develop an

idea or project); (2) descriptive case studies (used to help the researcher describe causal relationships within a phenomenon); and (3) explanatory case studies (used to help the researcher understand what influences behavior in a case) (Berg, 2007). This study takes on the design of an explanatory multi-case study in an effort to provide answers to my research questions and in an attempt to better understand how and why certain individual traits and characteristics affect policy and system implementation.

Research should meet three conditions in order to conduct a reliable explanatory case study: (1) the research must seek to explain how or why a phenomenon occurs; (2) the research must examine a contemporary phenomenon; and (3) the researcher(s) must have no control over the phenomenon (Yin, 2013). My study meets each of these conditions. The individual cases in this research represent K-8 school principals throughout the state of Michigan tasked with implementing teacher evaluation policies and systems. The principals in their specific context and their cognitive schemas bound the larger multi-case study and principals' thinking and enactment of teacher evaluation systems and policies is the primary unit of analysis of this work.

Specifically, I began by developing a theory of what factors might influence how school principals make sense of and ultimately implement teacher evaluation policies and systems. From this theory, I selected individual cases that fit the criteria of the theory (more information on my sampling later in this chapter). After designing all data collection protocols, I began conducting individual case studies before writing individual case study reports. Finally, I analyzed each of these individual reports.

Study Context: Educator Evaluations in Michigan

Michigan's effort at reforming teacher evaluation laws throughout the state began in 2009 when the state first applied for Race to the Top (RTTT) funding. In an effort to make their

application more competitive Michigan began to make changes encouraged by RTTT, including making student growth a significant part of a teacher's evaluation (Keesler & Howe, 2015). Michigan did not receive RTTT funding in 2009 or again when they applied in 2010, however the passed legislation set in motion new teacher evaluation systems. In 2010 Michigan did receive an NCLB waiver and condition of receiving this waiver they were required to rework their teacher evaluation system (Keesler & Howe, 2015). Although the 2010 teacher evaluation legislation had certain expectations of districts, such as making student growth a significant part of teachers' evaluations, the legislation still gave individual districts a lot of autonomy when determining how to evaluate teachers in their district.

A larger shift occurred in Michigan's teacher evaluation landscape in 2011, which increased the probationary period of beginning teachers from four years to five years and legislated that an untenured teacher, if rated effective or highly effective, could not be removed from his or her current teaching placement solely based on seniority (Michigan Department of Education, 2015). These changes aimed to improve the teacher workforce throughout the state by keeping the best teachers in classrooms. Additionally, the legislation said the state would put in place a teacher evaluation system beginning in the 2013-14 school year. In order to assist in creating a statewide system the Governor of Michigan created the Michigan Council for Educator Effectiveness (MCEE). MCEE consisted of educational researchers, educational experts, school principals, and members of the Michigan Department of Education (MDE) charged with developing a fair, rigorous, and transparent state-wide system for evaluating teachers and administrators. Together, these educational experts spent 18 months reviewing the most recent research across the country and globe regarding the most effective and fair way to evaluate teachers. In July of 2013, MCEE released its final proposal to overhaul Michigan's

teacher evaluation system. Based on these recommendations, House Bills 5223 and 5224 were written and originally scheduled go into effect during the 2013-14 school year. However, despite initial bipartisan support, HB5223 and HB5224, stalled for more than two years. The House and Senate could not reconcile several areas of contention, including what percentage of a teacher's evaluation should be tied to student test scores and what tests should be used to determine student achievement. As the legislation continued to stall, disagreements over teacher evaluation rubrics surfaced, with the Senate suggesting the type of rubric used to evaluate teacher should be a decision made by Local Education Agencies (LEAs), rather than limiting the rubric to one of the four originally recommended by MCEE.

Table 4.1.

Timeline of Educator Evaluation Changes in Michigan Since 2009

Year	Event	Brief Summary
2009-10	RTTT Applications NCLB Waiver	Michigan applied for but did not receive RTTT funding in 2009 and 2010. The state received an NCLB waiver in 2010.
2011	Public Act 101	Revised teacher tenure laws, established new requirements for teacher evaluations, new limits on collective bargaining.
2011	Public Act 102	Establishes MCEE to reform Michigan's educator evaluation system
2013	MCEE submits final recommendations	MCEE recommends four teacher evaluation frameworks, using student assessment data in teacher evaluations.
2013-14	HB5223/5224	Bills drafted based on MCEE recommendations. Stalled in legislation until 2015.
2015	Senate Bill 103	New requirements for teacher evaluations, including 40% use of student assessment data by 2018-19.

During this time of stalled legislation, Michigan continued have an ineffective system of distinguishing between teacher effectiveness. For example, since the reform of tenure laws in 2011, of the almost 100,000 teachers in the state, only 19 were dismissed due to poor evaluations (Michigan Department of Education, 2015). Additionally, teachers in Michigan continued to be rated overwhelming effective or highly effectively; 97% of teachers in the state meet this criteria

(Michigan Department of Education, 2015). In 2015, Senate Bill 103, a new attempt at teacher evaluation reform, was proposed and passed the Senate. After some changes, the House agreed to approve SB103 and more than four years after Michigan began the process of overhauling the state's teacher evaluation policies, SB103 passed, changing the evaluation measures of teachers and administrators. Beginning in 2018-19, 40% of a teacher's evaluation will be based on student achievement data. Additionally, in most circumstances, multiple observations of teachers will occur annually. These changes are consistent with what many educational researchers and experts consider "smart" teacher evaluation policy. For example, these groups agree that using multiple measures to evaluate teachers, such as observations and student assessment data, weighting these measures evenly (i.e. 50% student assessment data and 50% observational data), and observing teachers multiple times is the best way to evaluate teachers (Darling-Hammond 2012; MET Project, 2013).

Many policymakers and educational leaders believe these new evaluation laws have the potential to improve student achievement in the state by providing current teachers with effective feedback to help them improve their practice and by identifying and keeping effective teachers in the workforce. However, critics argue the large amount of discretion given to LEAs will bring into question how these individual entities will implement these new policies. These critics argue that while this new legislation guides districts and advises school districts on best practices, it lacks the legislative authority to truly impact how districts evaluate teachers.

Given the tumultuous nature of Michigan's teacher evaluation policy reform effort, Michigan is a timely case to study. At the time the data in this study were collected, the participants were learning new teacher evaluation systems and as a result their sensemaking will help shed light onto how principals, in general, might be navigating these new, complex systems.

This study has broader implications as other states continue to rework teacher and principal accountability systems in an effort to meet criteria set forth in the Elementary and Secondary Education Act (ESEA) and its subsequent revisions. In many states principals have had to negotiate and navigate impending teacher evaluation policy changes, changing evaluation rubrics, student growth models, and other evaluation logistics. It is plausible that principals in other states are experiencing similar teacher evaluation reforms and may have similar thoughts and beliefs as principals in Michigan.

Participants and Sampling Strategy

For this dissertation I targeted 12 public elementary school principals and 12 public school teachers. I targeted three principals who have minimal experience and face high outside pressure, three principals who have extensive experience and face high outside pressure, three principals that who have minimal experience and face low outside pressure, and three principals who extensive experience and face high outside pressure (See Table 4.2). After securing these principals, each principal asked for a volunteer teacher that we could observe during the evaluation process and that I could interview near the end of the data collection. I selected these 12 participants using criteria-based sampling. The 12 participants met my criteria of experience and current context.

Table 4.2.

Principal Participant Sample

Principal	1	2	3	4	5	6	7	8	9	10	11	12
High Experience / Low Pressure	X	X	X									
High Experience / High Pressure				X	X	X						
Low Experience / Low Pressure							X	X	X			
Low Experience / High Pressure										X	X	X

According to Marshall and Rossman (2006) when finding participants for a qualitative study it is important to consider; (1) if entry is possible; (2) if there is a high probability that a rich mix of the processes, people, programs, interactions, and structures of interest is present; (3) if the researcher is likely to be able to build trusting relations with the participants in the study; (4) if the study can be conducted and reported ethically; and (5) if data quality and credibility of the study can be reasonably assured (p. 62).

After completing a list of my ideal target of principal criteria I began to reach out to principals who met the before mentioned experience and pressure criteria. I solicited principal participation by phone and email. In the end, I was able to achieve my goal of 12 principals, three from each of the aforementioned categories (see Table 4.3. for complete participant background information). I designed this sampling scheme to capture variation with the different principals from each criteria. Although this type of embedded design was not able to capture all important variables in each context, the design was useful to provide insights of the different

perspectives offered by these principals (McLaughlin & Talbert, 2001). The goal of this type of sampling is not to make generalizable statements about all principals with similar characteristics, but instead to begin hypothesis and theory building about principals with these type of characteristics and how these characteristics may impact policy implementation.

Table 4.3.

Principal Background Information (TPS or Charter, Principal Experience)

Principal	TPS/ Charter	School Rating	Years as Principal	Years at Current School	Years as Teacher	Level of Education
Mr. Bania	TPS	Yellow	10	10	6-10	M.A.
Ms. Goldstein	TPS	Yellow	10	4	6-10	M.A.
Dr. Wexler	Charter	Lime	9	3	6-10	Ed.D
Mr. Bookman	TPS	Red	10	1	6-10	M.A.
Ms. Hamilton	TPS	Red	10+	3	6-10	M.A.
Ms. Cohen	Charter	Red	10+	3	6-10	M.A.
Mr. Jarmel	TPS	Red	1	1	6-10	M.A.
Ms. Robbins	TPS	Red	3	3	10+	M.A.
Mr. Ramon	TPS	Red	3	1	10+	M.A.
Ms. Steinman	TPS	Yellow	1	1	10+	M.A.
Ms. Chang	TPS	Lime	4	4	10+	M.A.
Mr. Sherman	Charter	Lime	4	4	1-5	M.A.

*Michigan's 2014 Accountability Report Card Ratings: Green (highest), 85% or greater of possible points; Lime, between 70-84% of possible points; Yellow, 60-69% of possible points; Orange, 50-59% of possible points; and Red (lowest), Less than 50% of possible points.

Data Collection

According to Yin (2013) case studies typically draw information from sources including interviews, direct observations, participant observations, documentation, archival records, and artifacts. In this study I rely on four sources of information; (1) principal questionnaires; (2) interviews with principals and teachers; (3) observations of principals conducting evaluations of teacher instruction and observations of pre and post teacher evaluation conferences with principals and teachers; and (4) artifacts, including district teacher evaluation policies, teacher evaluation observation rubrics, principal observation notes of teacher instruction, and final

teacher evaluation ratings. My research questions were best addressed by these types of data and this type of data collection is consistent with other work in the field that has tried to better understand how school leaders make sense of and implement school policies (Coburn, 2005; Derrington, 2013; Koyama, 2014; Rigby 2015). Additionally, collecting these type of data allowed me to validate my data and findings through data triangulation by showing that diverse data collection methods confirm the findings (Miles et al., 2014). In this study my goal was to learn how principals in different environments with different experience levels make sense of and implement teacher evaluation policies and generate hypotheses of how and why these type of characteristics affect policy implementation. The coding for this study came from the four sources mentioned above.

I administered a questionnaire to all principal participants at the beginning of data collection. (I began data collected in February of 2016 and collected the final data in January of 2017). The first part of the questionnaire asks participants about their relevant work experience, years serving as a principal (and teacher), level of education, how long they had served as a principal in their current school and additional school-context questions. The second part of the questionnaire asked principals a variety of questions about their school's teacher evaluation system and policy. I used the questionnaire as a screening process to get background data on the principals to ensure the principals met the aforementioned criteria. Additionally, I use the questionnaire to generate some of the interview questions. Finally, I used data source triangulation, specifically the questionnaire, interviews, and observations, to strengthen the validity of my findings. For example, I compared and contrasted the answers principals gave on part two of the questionnaire to the answers they gave to me during interviews and to what I

observed the principals doing in practice. A complete version of the questionnaire is found in Appendix A.

I interviewed the principals in this study three times each between February of 2016 and January of 2017. I conducted the interviews in one-on-one settings and focused on the principals' experiences using and perceptions of teacher evaluation policies. I audio-recorded all interviews and I took notes during the conversation. Each interview lasted between 30 and 60 minutes. The interviews took place three times during the data collection – once at the beginning of the collection, again during the middle of data collection, and then near the end of data collection. I conducted the interviews in an effort to triangulate the data sources and strengthen the findings of this work. The purpose of the three interviews was to examine how principals make sense of teacher evaluation policies. The first interview focused on principals' understanding and knowledge of the design of their current teacher evaluation system and principals' beliefs about these systems. The second interview focused on principals' experience implementing these systems. The final interview focused on reflecting on the observation of the teacher we co-observed. The full principal interview protocols are found in Appendices B, C, and D.

I conducted teacher interviews (with available teachers) near the end of data collection in the spring, summer, and fall of 2016. I conducted the interviews in a one-on-one setting and focus on the teachers' experiences with and perceptions of teacher evaluation policies. Each interview lasted between 30 and 45 minutes. I audio recorded the interviews and I took notes during the conversation. I conducted the interviews in an effort to get the teachers' perspective on how their experiences with teacher evaluations, including how teachers perceive principals implement teacher evaluations and how their work is impacted by the teacher evaluation process. The semi-structured interviews focused on three main areas: (1) teachers' understandings and

knowledge of the design of their current teacher evaluation system; (2) how they feel their principal implemented this system; (3) and how their practice is impacted by these evaluations. A complete teacher interview protocol is found in Appendix E.

I observed each principal conducting a teacher observation that was used for a teacher's final evaluation score. I collected these observations in the spring of 2016 and fall of 2016, as principals conducted official evaluations of their teachers. Each observation lasted between 30 and 60 minutes. Additionally, when available, I observed principals at the required teacher evaluation pre and post conferences. As I observed, I took field notes, completing them immediately following each observation to ensure accuracy. I shared the notes I took with both the principal and teacher to ensure I accurately represented their thinking and conversations during each observation. The purpose of the observations were to better understand how principals observe teachers in practice and how principals and teachers communicate about the evaluation and evaluation process. Additionally, as was previously mentioned, it is important in qualitative work to observe people in their natural environments (Yin, 2013). A complete field notes template for the observation is found in Appendix F.

I collected district- and school-based teacher evaluation documents as provided by principals. These documents included district-wide and/or school specific teacher evaluation policies, observation and conference protocols, and other documents principals used while conducting teacher evaluations. Additionally, if given permission by principals and teachers, I collected final teacher evaluation scores and principal observation notes (see Table 4.4 for full data collection details). The purpose of collecting these documents was to better understand what principals were asked to do by their district and school and to understand how principals were making sense of what they were being asked to do. Finally, the purpose of collecting these

documents was to better understand what principals would be looking for during teacher observations and to better understand the type of feedback principals were giving teachers, including how principals actually rated individual teachers. I collected this information in an effort to see what these principals noticed during teacher observations and if and when this information was addressed during teacher evaluation post-conferences and in principal feedback to teachers.

Table 4.4.

Principal Data Collected

Principal	Quest.	Interview #1	Interview #2	Interview #3	Observe	Post- Conf	Teacher Interview
Mr. Bania	X	X	X	X	X	N/A	X
Ms. Goldstein	X	X	X	X	X	X	X
Dr. Wexler	X	X	X	N/A	N/A	N/A	N/A
Mr. Bookman	X	X	X	X	X	X	X
Ms. Hamilton	X	X	X	X	X	X	X
Ms. Cohen	X	X	X	X	X	X	X
Mr. Jarmel	X	X	X	X	X	X	X
Ms. Robbins	X	X	X	X	X	X	X
Mr. Ramon	X	X	X	X	X	X	N/A
Ms. Steinman	X	X	X	X	X	N/A	X
Ms. Chang	X	X	X	X	X	X	N/A
Mr. Sherman	X	X	X	X	X	X	X

Data Analysis

I used the questionnaire as a screening process to get background data on the principals to ensure the principals met the aforementioned criteria and to generate some of the interview questions used in this study (Miles et al., 2014). Using Atlas.ti software I first analyzed all of the questionnaires, comparing participant responses to school-based questions about the participants' beliefs, thoughts, and knowledge of their current teacher evaluation system. Additionally, I coded the background data gathered from the participants in an effort to look for themes, commonalities and differences between participants with similar and different characteristics.

The characteristics I coded include; 1) the age of the participant; 2) the participants' level of education; 3) years of experience as a principal; 4) years of experience as a principal at their current school; and 5) the number of years each participant spent as a classroom teacher prior to becoming a principal.

After coding all of the data from the questionnaire I coded individual participant interviews. I waited until I had collected fifty percent of the interviews to begin coding these data. Then I randomly selected three of these interviews to begin the coding process. These interviews were Mr. Sherman Interview #2, Dr. Wexler Interview #1, and Ms. Hamilton Interview #1. I began the coding process by looking for overarching themes within the data. In qualitative research themes are more general terms, phrases, or sentences which encapsulate larger groups of more specific codes (Miles et al., 2014). Once I documented these themes I began a generating specific codes, which relate to these overarching themes, but are more specific data points and generally include the language of the participants (Miles et al., 2014). I developed the codes inductively and as themes emerge from the coding process I grouped together by theme (Miles et al., 2014). After developing these codes I coded each of these interviews a second time, noting any discrepancies. Once I developed the initial codes, I began coding the additional interviews and added these codes to my code book. The initial themes that emerged were: 1) communication; 2) data use; 3) principal and teacher prior knowledge and experience; 4) relationships; and 5) the teacher evaluation system/policy. From these themes I developed a larger codebook. For example, under the theme of communication, specific codes included: 1) how principals communicate information about the teacher evaluation process; 2) how principals give feedback/scores to teachers; 3) how principals and teachers address discrepancies/disagreements during the evaluation process; and 4) how principals communicate

new/changing teacher evaluation policies/systems to their staff. I coded all of this data using Atlas.ti software to analyze and interpret patterns, trends, commonalities, and links among the participants (Miles et al., 2014).

I then reviewed all codes, looking for common excerpts that highlighted similar themes and ideas. I then checked the validity of the coding process by recoding the data for a second time. I noted any discrepancies and these discrepancies were addressed in order to refine and justify assertions and to look for possible other alternative interpretations of the data (Guba & Lincoln, 1994; Miles et al., 2014). After I completed the coding process, I compared quotations to the original interview text, making sure these data were taken in context and accurately represented what the participants attempted to articulate.

I completed the same process of coding for all of the observations collected during data collection. Specifically, I randomly sampled three observations to begin the process of developing codes. The observations I selected were; 1) Ms. Cohen's post-conference; 2) Ms. Goldstein's pre-conference observation; and 3) Mr. Bania's post-conference observation. Finally, I coded all other documents and data including principal observation notes, principal final evaluation ratings of teachers, district documents (i.e. observation rubrics, etc.) using Atlas.ti software. I analyzed these documents individually, looking software and analyzed to interpret patterns, trends, commonalities, and links among the participants (Miles et al., 2014).

After completing the coding process outlined above for all collected data I ran frequency checks using Atlas.ti software to further ensure the themes and codes which I developed accurately represented the overall tone, scope, and information presented in the data. Additionally, after individually coding all interview data I trained a colleague in how I coded these data, including providing my colleague my code book and explaining my thinking about

how I coded these data. I then provided a random sample of one interview to this colleague who coded the interview on her own. I then compared my colleague's coding to my own to look for discrepancies and instances where we coding some or all of the data differently. In the end, we had 81 percent agreement on this sample. I then provided this colleague one additional interview and one field notes observation from my sample of data. Again, we compared the results of my colleagues coding to my own and had an 82 percent agreement after coding each of these two documents. Finally, I provided my colleague an additional 10 principal interviews, one teacher interview, and three field notes observations. Upon completion we compared all of my original coding and my colleague's coding. In the end, we had 80 percent agreement on all of the coding.

Establishing Validity

According to Miles et al. (2014) the first step to establishing validity is to thoroughly prepare for the research. The researcher should have some familiarity with the setting and phenomena under study, strong conceptual interest, multidisciplinary approach, and good investigative skills (Miles et al., 2014). I have five years' worth of experience working in public schools, first as a teacher and then on the administrative side as an instructional coach. Additionally, my coursework at Michigan State University, specifically my work in EAD991A (Teachers and Teaching in an Era of High-Stakes Testing), and TE931 (Introduction to Qualitative Research Methods), has helped me hone my qualitative research skills and prepared me for work on this important topic. Finally, my practicum, which focused on how principals make sense of teacher evaluation policies, helped prepare me for the role of the researcher. For my practicum, I developed skills in the following areas: (1) developing interview protocols; (2) interview participants; (3) coding qualitative research data; and (4) analyzing and writing a complete academic paper using qualitative data.

Maxwell (2005) defines validity as, “the correctness or credibility of a description, conclusion, explanation, interpretation, or other sort of account” (p. 106). In qualitative research there are several threats to validity including; (1) researcher bias; (2) reactivity; and (3) manipulation of the data (Maxwell, 2005; Miles et al., 2014). Researcher bias has the potential to influence the data the researcher identifies as important and/or the conclusions drawn from the data. Reactivity can take place when the simple act of conducting research changes the behavior of the participants in the study. Finally, data manipulation may occur when the research tries to find data that fits his or her existing theory or hypothesis. To combat these potential threats to validity, the researcher must thoroughly explore and explain his or her biases and how these biases will be dealt with throughout the duration of the study (Maxwell, 2005). I addressed my potential biases by ensuring all participants are allowed to read transcripts of recorded information and notes and they were afforded an opportunity to address any discrepancies that they feel do not accurately portray what they were trying to say or do. Specifically, to establish validity for all interview and observation data I left room to ask participants about any comments they make, making sure I clarified their statements before drawing any conclusions. Additionally, I contacted all participants to clarify any questions that arose during the transcribing and coding of the data. I also solicited critical feedback from colleagues throughout the data collection and writing process. Additionally, I constantly acknowledged how my past experiences may have impacted data collection and writing and I made every effort to remain neutral by asking non-leading questions, asking for clarifying comments, and collecting and using the data completely and in context.

Finally, Lincoln (1995) identified eight standards for evaluating the quality of qualitative research: (1) standards set in the inquiry community; (2) positionality; (3) community; (4)

participant voice; (5) critical subjectivity; (6) reciprocity; (7) respect; and (8) sharing privileges. Throughout the research design, data collection and analysis, and writing, I attempted to meet each of these eight standards of quality by reviewing similar studies and dissertations that used a similar research design approach. I reviewed this study early in the dissertation design process, even dating back to the drafting of my dissertation proposal. Throughout this entire process I referred back to these studies in an effort to make sound methodological decisions and research design choices throughout this work. Given the sensitive nature of observing teachers while they were being observed and learning about their teacher evaluation scores in the post-conference, I made efforts to acknowledge my role in these environments. I was simply an observer be an observer and used the final interviews with both the teacher and the principal as a chance to answer follow-up questions.

Limitations

There are two main limitations to the design and methodology of this study. First, this study is limited by the participants. The participants in this study were not randomly selected. Additionally, the number of participants does not allow me to make generalizable statements about all principals. If I collected data from 12 other principals these principals could have provided different insights and thoughts, resulting in a different interpretation or analysis of the data. In this way, the principals in this study shape the findings by their experiences, thoughts, and beliefs. To address this participant limitation I used criteria-based sampling, soliciting principals from four different subcategories; (1) principals with high-levels of experience in high-pressure environments; (2) principals with high-levels of experience in low-pressure environments; (3) principals with low-levels of experience in high-pressure environments; and (4) principals with low-levels of experience in low-pressure environments. Although this type of

sampling cannot account for all differences, the goal of this work is to begin to hypothesize about how principals with certain characteristics think about and enact teacher evaluation policies. Therefore, this sampling scheme was necessary and appropriate to answer my research questions.

The second limitation is, although principals were observed in their natural environment implementing their teacher evaluation policy, I did not observe each principal multiple times, with a variety of teachers, or during every interaction the principal had attempting to implement the policy. In this way, my presence as a researcher during data collection and the researcher may not have captured exactly how principals were conducting teacher evaluations in all circumstances. To account for this limitation I spoke with teachers when available to see if what I observed was an accurate or consistent representation of how these principals navigated the process of teacher evaluations.

Despite the aforementioned limitations, the data collected in this study provide a great insight as to how the principals in this study navigate teacher evaluation policy implementation. Although not generalizable to the entire principal community, the results and analysis of this work will serve as the basis for hypothesis building and testing in future research. As principals across the United States continue to make sense of evolving teacher evaluation policies, the results of this work have the potential to be explored with different principals in similar contexts in varying locations throughout the United States.

Chapter 5: How Principals' Cognitive Schemas Impact Their Implementation of Teacher Evaluation Systems

“My prior knowledge as an administrator is huge. My first evaluations in my early years of being an administrator were probably based on a lot of feelings. As you grow as an administrator your feelings die.” - Ms. Cohen (10 years of experience)

My research questions and my theoretical framework guide all of the findings in chapters five and six. This chapter answers my first research question: How do principals' cognitive schemas influence their implementation of teacher evaluation systems? The overarching theme that weaves throughout my analysis of these findings is principals' cognitive schemas influence the type of sensemaking in which they engage, impacting how they think about the overall process and purpose of teacher evaluations. In addition to this overarching theme, four dominate subthemes emerged through my analysis of the data using the lens of cognition and specifically sensemaking theory. The first subtheme is principals' cognitive schemas influence the type of leadership in which they engage, which impacts how principals navigate the process of teacher evaluations. The second subtheme suggests principals' cognitive schemas influence how they use previous teacher evaluation information during the teacher evaluation process. The third subtheme suggests principals' cognitive schemas guide individual principal perceptions and beliefs of the accuracy of their current teacher evaluation system (in terms of accurately capturing a teacher's effectiveness). This finding impacts how principals use and at times take liberties with these systems. The final subtheme suggests principals' cognitive schemas affect how principals think about using the results of teacher evaluation scores when hiring new teachers.

Overarching Theme: Individual vs. Collective Sensemaking

After an analysis of these data (again, principal questionnaires, interviews with the principals and teachers and observations of principals' conducting a teacher evaluation), I found that principals in this study with high-levels of experience (nine or more years as a principal) engaged in "individual sensemaking," while principals with low-levels of experience (principals with four or fewer years as a principal) engaged in "collective sensemaking." To review, individual sensemaking is a type of sensemaking generated from an individual's thoughts, experiences, and beliefs. Collective sensemaking occurs among multiple people who are attempting to make sense of a similar or the same situation (in this case, the teacher evaluation process). My findings suggest a principal's cognitive schema influences how he or she comes to understand and implement teacher evaluation policies based on the amount of experience they had as an administrator. As was mentioned previously, research shows experience impacts how principals think about and ultimately implement policies and reforms (Coburn, 2001; Seashore-Louis et al., 2010). The findings from this study support these earlier findings, while providing more nuanced reasons as to how and why principal experience matters for teacher evaluation policy implementation. The four subthemes that follow further explain how principal experience influenced how these principals made sense of teacher evaluation policy implementation as well as provide nuanced information as to the thoughts, actions, and beliefs of these principals.

Subtheme One: Principal Leadership

All of the principals in this study said their leadership style impacted how they thought about and implemented their school's teacher evaluation policy. However, as evidenced by talking with and observing these principals, the leadership style of principals varied. Principals with high-levels of experience primarily engaged in situational leadership. Social science

researchers define situational leadership as a leadership style which requires a rational, adept, individual who understands situations and responds and reacts to situations in ways that are beneficial to the organization (Grint, 2011; McCleskey, 2014). Situational leaders vary the amount of support, direction, and goals they provide individuals, understanding that individuals need different levels of support when attempting to accomplish a task (Gates, Blanchard, & Hersey, 1976). When describing their leadership style some of the principals with high-levels of experience in this study said specifically they were situational leaders, while others demonstrated the characteristics of this leadership style in both their thoughts and actions. For example, Ms. Cohen (10 years of experience) said:

Situational leadership is what I subscribe to. I think it falls in line actually as an educator too. I think every person is a unique individual. They need to be treated that way.

Every teacher needs something different. I have one teacher down the hallway. She's only been teaching like three years. She's amazing. I can say, "I'd like to look at your student data." She's like, "Okay." She comes back with this spreadsheet and it's all tallied and averaged at the bottom and red and color coded. I have another teacher, "I want to look at your data." She brings me this binder with tests in it. Each one needs something different.

When asked how her leadership style impacts how she evaluates the teachers in her building, Ms. Cohen said the way that she conducts observations of teachers and directs conferences with her teachers varies based on the individual. For example, Ms. Cohen explained she knew the strengths and weaknesses of the teachers in her building and therefore she was able to tailor their evaluations situationally. She referenced one teacher who she knew struggled with reading instruction in particular and because she knew of this struggle, Ms. Cohen focused her

attention on reading instruction during the observation process. The teacher admittedly thought her reading instruction needed support and while both Ms. Cohen and this teacher believed the teacher had strengths in other areas, reading instruction was an area of concern. Because this teacher struggled with reading instruction Ms. Cohen made sure to observe this teacher during a reading lesson; she provided feedback and support based solely on the teacher's reading instruction; and she focused her conversations with this teacher almost entirely on how the teacher could improve as a reading instructor.

In another example, Ms. Cohen said the teacher that she and I observed during her official evaluation struggled in using student data to make instructional decisions. As a result, the examination of student data was the focus of her observation. Prior to the observation, Ms. Cohen looked at this teacher's lesson plan to see if she provided evidence of how she used data to inform the lesson. During the observation, Ms. Cohen looked around the teacher's classroom to see whether student data were on display. During this teacher's official teacher evaluation post-conference, aside from discussing the teachers' final evaluation rating, the conversation focused exclusively on how this teacher was using student assessment data to guide her instruction and promote student growth. Ms. Cohen led all of her teacher evaluations this way, situationally evaluating teachers based on her personal understanding of where she believed the teachers needed support to become better educators.

Mr. Bookman (10 years of experience) also described his leadership style as situational, particularly when evaluating his teaching staff. He said:

I would describe my leadership style, honestly, as situational. Especially if I'm going to observe and then instruct a teacher, be an instructional coach. I'll be honest with you, it makes some people uneasy because it's like there's this element of unpredictability, and

they want to be able to predict what I'm doing all the time, but I can't even tell them what I might do in a situation because it's like, "Well, tell me more about the situation." I haven't found anything that's black or white in education in my experiences, and I've dealt with them all. I've dealt with the parents. I've dealt with the students. I've dealt with the teachers. There doesn't ever seem to be anything that's black or white.

Mr. Bookman went on to explain he did not believe in a "one size fits all" leadership approach. This philosophy impacted how he approached evaluating the teachers in his building as he believed that all teachers needed something different. For example, Mr. Bookman described one of his teachers as having a particularly challenging class in terms of student behavior. Mr. Bookman said he factored in the challenging nature of this teacher's class by increasing this teacher's score on the professionalism part of the evaluation (although doing so was technically not permitted in his evaluation system). Mr. Bookman constantly referenced how he situationally evaluated teachers throughout the entire year and did not rely only on the official teacher evaluation when assigning final teacher evaluation ratings. For example, after our co-observation of one teacher, Mr. Bookman was quick to note that the lesson we observed was not an accurate reflection of the quality of the teacher. Mr. Bookman addressed what he deemed a subpar lesson with the teacher in the post-conference, but ultimately rated this teacher highly effective, because he knew this one lesson observation, although technically the official observation used for evaluative purposes, was not a true reflection of this teacher's performance.

Ms. Cohen and Mr. Bookman provide two illustrative examples of leadership that are representative of the majority of principals in this study with high-levels of experience. Ms. Goldstein (more than 10 years of experience) said, "My leadership style changes daily based on what's happening in the building. At the end of the day, I have to make the decision of what's

best for our students and our staff.” This sentiment further highlights the finding that principals with high-levels of experience led situationally, which impacted how they evaluated their teachers. Ms. Goldstein went on to explain that her district had trained all principals to evaluate their teachers in a very structured way, in an effort to make sure all teachers were receiving consistent evaluations throughout the district. However, because she thought about each teacher situationally, Ms. Goldstein found evaluating all teachers the same way a difficult task to accomplish. Ms. Goldstein said in her building teacher evaluations might look different for individual teachers and in her mind this variation was fine, because all teachers need something different.

In contrast to their more veteran peers who engaged in situational leadership, principals with low-levels of experience overwhelmingly described their leadership style as relational. Briefly defined, relational leadership “expresses the degree to which a leader shows concern and respect for their followers, looks out for their welfare, and expresses appreciation and support” (Bass, 1990a; 1990b). The characteristics of relational leaders include working with their followers in an effort to achieve a goal and valuing the input and emotional needs of their followers. In this study, principals with low-levels of experience exhibited the characteristics of this definition. Because these principals were new or relatively new in their current role, they all wanted to make sure they established and improved relationships in their building. Wanting to secure positive relationships with their staff impacted how these principals thought about, constructed meaning around, and ultimately implemented teacher evaluation policies. Ms. Steinman (one year of experience) was representative of the larger group of principals with low-levels of experience. She said:

I am very relational in my leadership style. I have tried to consciously be more relational with them (teachers) in the non-evaluation sense because I don't want them to feel like I'm picking on them or targeting them (when performing their official evaluation). That just might be my own insecurity. I might get over that later. It is still just really hard for me to feel like I am giving someone a bad score and then not having a relationship with them. Relationships allow you to have those tough conversations, and if you don't have that relationship, that tough conversation can't occur.

When asked if her leadership style impacted how she thought about teacher evaluations Ms. Steinman said, "Yes, because evaluating a teacher does put a strain on your relationship. Especially the relational part. There is a fine balance there. I really want them to be able to just have their own voice and speak their own truth." Ms. Steinman went on to say that she and her teachers co-developed how teacher evaluations would occur during the school year. Although Ms. Steinman was quick to point out that there were many logistical things that she could not change (such as the documents she needed to provide teachers and the time she spent observing teachers), she thought working with teachers to construct an understanding of how to best use their district's teacher evaluation system would be a good approach and especially an effective approach when trying to secure positive relationships.

Ms. Steinman went on to explain she has seen benefits of her relational leadership style, and, in her opinion, it was important for her staff to see she was working with them, especially on something as important as their evaluation. She said:

I did a survey with the staff so I could get some feedback also. I think that I was reaffirmed in the idea that relationships are strong because almost all of them commented about how much they appreciated being treated like a professional, being allowed to have

a voice, and feeling like I really took time to get to know them on an individual basis. I think that reiterates that they also feel that I'm a relational leader in that way.

Ms. Steinman did not suggest she let her staff take advantage of her when it came to their official evaluation; instead, she explained that the way she and her staff thought about this process was co-constructed. The co-construction of the evaluation process between Ms. Steinman and her teachers resulted in teachers having a say in what was valued during observations. One example of this co-construction is Ms. Steinman said that staff very much valued classroom routines and procedures. As a result, Ms. Steinman paid close attention to this section of the observation rubric during evaluations. She looked for evidence of clear routines and procedures during observations of teacher instruction as well as talked about routines and procedures during conversations with her staff.

Ms. Robbins's (three years of experience) actions provide further evidence that principals with low-levels of experience invested in relational leadership, especially when it came to teacher evaluations. Ms. Robbins explained she spent much of her time thinking about her leadership and how her leadership style impacted evaluations of the teachers in her building. Ms. Robbins constantly mentioned how she wanted to work with teachers so they would be successful during the evaluation process and for working together with her teachers meant communicating with her teachers during the process to make sure they both agreed on what was happening. She said:

I do a lot of thinking before I conference with teachers. I really firmly believe that it's so important to say the right thing the first time. You really can't take things back. You could do a lot of damage. By not phrasing the things the right way, you can shut somebody down and discourage somebody that doesn't need to be discouraged. You

could give false hope. I do think it's really important to make sure that the message that teachers are getting is right, and that you're being as fair as you can to them and that you're not shutting somebody down or ruining a relationship.

For Ms. Robbins, maintaining a strong relationship with the teachers was essential to the overall climate of her school. She noted that strained relationships could compromise things such as communication between her and her staff and the overall climate of her school Ms. Robbins continued:

I'm very positive, very supportive. We try to work as a team. I try to take advantage of the expertise that's in the building and to encourage people that may not be sticking their necks out and showing what they know and sharing their good ideas. I'm a new leader, so sometimes I'm not—I don't always feel sure of myself. I just try to make sure that I'm keeping that—what's best for the kids in mind. That it is important for me to do what's good for teachers too, to make sure they feel taken care of and valued. I feel like our job as teachers is to continue—I'm still calling myself a teacher—it's just to continue to grow ourselves and to improve so that we can meet the needs of the students as they change.

Ms. Robbins went on to say that she was very positive in all her conversations with the teachers in her building when discussing their evaluation. Even if she was giving critical feedback, Ms. Robbins always tried to think back to when she was a teacher and how having these conversations were tough. Ms. Robbins said at times she might not be a critical of teachers as she needs to be, but she thought support and positive affirmation were better approaches than criticism and negative feedback.

Ms. Steinman and Ms. Robbins provide two examples that are representative of principals with low-levels of experience in this study. These principals typically engaged in

relational leadership in an effort to secure positive relationships with their staff. However, this type of leadership style also impacted teacher evaluations as these principals were quick to give teachers the benefit of the doubt and at times avoid difficult conversations around teacher performance because they wanted to secure these relationships. Principals with low-levels of experience also empathized with teachers more often than their more experienced peers as they more recently were in the classroom and had recently went through the teacher evaluation process themselves as former teachers.

Interesting to note is I did not provide and examples or definitions of leadership when asking principals how their leadership impacted their implementation of teacher evaluation policy. The principals in this study knew the terms “situational leadership” and “relational leadership” and used these terms unprompted to define their leadership style. Finally, I think it is important to note while situational leadership and relational leadership are two distinct leadership approaches, these leadership styles are not mutually exclusive. For example, relational leaders show concern for their followers and value their followers’ thoughts, ideas, and opinions. Situational leaders also have these characteristics. However, the ways in which situational and relational leaders approach evaluating teachers is different. As evidenced by the principals in this study, principals who engaged in situational leadership varied how they evaluated teachers, while principals who engaged in relational leadership typically evaluated all teachers similarly.

Nuances. One way in which principals’ cognitive schemas impacted the way in which they thought about and ultimately evaluated teachers was the leadership style to which they subscribed. The principals in this study with high-levels of experience typically completed evaluations situationally, completing each on a case by case basis, while factoring in all they knew about the individual teacher. The principals in this study with low-levels of experience

typically engaged in relationship leadership, which including co-constructing how teacher evaluations looked in practice. Although there was a clear demarcation in leadership style between experience levels, the relationship between leadership style and experience did not hold for all principals. For example, one principal, Mr. Sherman (four years of experience) described his leadership style as situational. Mr. Sherman explained that given the current needs of his teaching staff, he did not think it was beneficial for all of his teachers to be evaluated in the same way. He explained that his teaching staff varied greatly in level of experience and as a result how he evaluated veteran teachers looked much different than how he evaluated pre-tenure teachers. In his mind this was a perfectly legitimate approach to teacher evaluations because given the varying experience levels of his staff, each of his teachers was best supported by an evaluation specific to their current experience level. Dr. Wexler (nine years of experience), said she thought relationships were the single most important factor when leading a school. Dr. Wexler explained her school experienced a low retention rate of teachers and she wanted to change this. Dr. Wexler believed focusing on building strong relationships would help decrease the number of teachers leaving her school and as a result she prioritized relational leadership, particularly during the teacher evaluation process. However, aside from these two examples, five of six principals fit into the aforementioned categories (i.e., high experience principals were situational leaders and low experience principals were relational leaders).

One other nuanced difference in how principals' cognitive schemas are impacted by experience is three principals with high-levels of experience (Mr. Bookman, Ms. Cohen, and Ms. Hamilton) admitted their leadership style changed as they gained more experience. For example, Mr. Bookman said:

Where I've gone wrong in the past, is not having the guts to do it. Not having the guts to have the tough conversations, and when you're having those tough conversations keeping the emotion out of it. It's a matter of fact thing, and it's always in the guise of so that you can be a better teacher, and a better person, and you can be successful. You've got to have some tough conversations, and you got to ask some tough questions, but people respect that a heck of a lot more than they do someone who doesn't address it.

Mr. Bookman went on to say that wanting to develop positive and trusting relationships with his staff impacted how he had these "tough conversations" and how he ultimately evaluated his teachers. However, as he gained experience and became more comfortable in his role as a school leader, he will much more willing to have these tough conversations and critique teacher performance.

Ms. Cohen also reflected on how she evolved as a leader and said:

As a new administrator I did have a hard time. Now I'm pretty cut and dry. I say what I think. I think that's just something you learn as you get older. Those strong-willed teachers in the beginning, some of them are scary. I'm evaluating them and I'm think, oh my gosh, they're going to hate me when we get done. If you want to be a good administrator you need to forget about what people think about you. You'll never make a good administrator if you waffle. My prior knowledge as an administrator is huge. My first evaluations I think, in my early years of being an administrator, were probably based on a lot of feelings. As you grow as an administrator your feelings die and you don't have them anymore. I think you can look at things more objectively, taking the subjectivity out of it as much as possible.

Subtheme Two: Use of Prior Evaluation Data

Another way in which principals' cognitive schemas impacted the way in which they implemented teacher evaluation policy was how these individuals thought about and used prior teacher evaluation data during the teacher evaluation process. Six principals in this study (Mr. Bania, Ms. Goldstein, Dr. Wexler, Mr. Bookman, Ms. Hamilton, and Ms. Cohen) had at least nine years of service as a school principal. The principals in this study with high-levels of experience overwhelming stated they did not rely on previous teacher evaluation data (including prior teacher evaluation ratings and prior student assessment/achievement data) when evaluating teachers in the current year. When asked if he reviewed teachers' previous evaluation data Mr. Bania (10 year of experience) said:

Nope. Fresh each year. I mean, I kind of know where they're at. If someone's highly effective for three years in a row they didn't have to be evaluated, like every other year. It's broken down in here like who those people are. When I go in I know like those are The highly effective teachers that particular year, and these were those that were not highly effective three years in a row. We do look at percentages too as an administrative team to see how many highly effective teachers I had here versus how many effective versus other things.

When I asked why he did not consult any prior information when evaluating teachers in the current year, Mr. Bania said he felt that whatever a teacher had done in the past should not be reflected in their current teacher evaluation. Mr. Bania also said they he wanted to rely on what he saw and heard from teachers in the current school year. He did not want to rely on past teacher performance data and only wanted to evaluate a teacher based on his impressions of that teacher within the current school year, as according to him, evaluations are "year to year."

Mr. Bookman also indicated he did not use previous evaluation information when evaluating teachers in the current year teachers. He said:

With teachers I don't. I want no preconceived notions. You know what I mean? I'm smart enough and experienced enough that I can figure out what kind of teacher they are. I don't want to see any letters that may have happened in their file or any past evaluations that three different evaluators did because that's all arbitrary to me.

Mr. Bookman was new to his current school, although he had been a principal for 10 years. Because he was new to his school Mr. Bookman did not want to have his mindset influenced by the previous administration when evaluating his new staff. Additionally, Mr. Bookman was confident that he would be able to accurately assess a teacher's performance given his prior experience as an administrator.

A third principal with high-levels of experience, Ms. Hamilton (more than 10 years of experience), also stated she never looked at previous teacher evaluation data, including student assessment data or teacher observation ratings. Ms. Hamilton explained:

Not at all. Nope. Each year we have tabula rasa. I have a personal relationship with all of them and I know their peculiarities, I know their strengths, I know their areas of improvement, and I'm helping them with it all.

Six principals in this study (Mr. Jarmel, Ms. Robbins, Mr. Ramon, Ms. Steinman, Ms. Chang, and Mr. Sherman) had four or fewer years of service as a school principal. When compared to their more experienced principal peers, these principals were much more likely to call upon, review, and use prior teacher evaluation data (including prior teacher evaluation ratings, both which they had provided and which other principals had provided, and student assessment data) during the current year teacher evaluation process. Ms. Robbins said:

When I first started as a principal I felt like I kind of needed to know what they had been working on, what the previous principal felt their strengths were and their weaknesses.

When you take over sometimes, you may have somebody that had a personality conflict with the previous person. You may have somebody that was just really chummy with the previous principal that may have been getting something that she wouldn't have given them as a score. I want to get an idea of what happened before I came here, just to see what you're working on and then we're going pick up together and let's see what we do from here.

When asked how reviewing this information impacted, if at all, how she approached evaluating her staff, Ms. Robbins said:

I think it definitely—in some ways, it influenced the conversations that we had. Without saying, “Listen, I know that you and the previous principal were working on this,” I kind of had an idea of what that teacher was really focused on, what somebody had said to them before. It's aware of—that is something that I've wanted to look for as I've been observing. If somebody had mentioned that we really have some very surface-level—we don't have deep questioning happening in this classroom. My antenna's up for that when I go in. In some ways, yeah, it is definitely going to have an impact.

For Ms. Robbins, and other principals with low-levels of experience, it was important to get a complete picture of the teachers they would be evaluating. Ms. Steinman added:

I think it (reviewing previous teacher evaluation information) might help focus me a little bit and then maybe help focus in on what I want to provide feedback to them on. I think in that sense, that'll be nice, to have that knowledge prior.

Ms. Robbins and Ms. Steinman provide two examples of principals with low-levels of

experience who want to make sure they have a complete picture of the teachers they are evaluating. Both Ms. Robbins and Ms. Steinman said this desire to have as much information as possible goes back attempting to establish positive and trusting relationships with their staff. During the post-conference with her teacher Ms. Robbins brought up and discussed the teacher's previous evaluation at length. Ms. Robbins did brought up this information in an effort to show the teacher that she understand what happened in the past and how she would be structuring evaluations moving forward. Additionally, Ms. Robbins brought up specific scores and feedback from the prior evaluation providing her thoughts on these areas and whether she was in agreement with the previous assessment of this teachers' performance.

Another principal with low-levels of experience, Mr. Ramon (three years of experience), said while he initially tried not to review any prior teacher evaluation information he did end up reviewing teachers' previous evaluation data, which he believed helped him feel more justified assigning ratings of teachers. Mr. Ramon said:

I actually tried not to. I know they talk about diminished returns or anything that could potentially impact your evaluation, so I know I initially said I wasn't, and then I did actually go back. I definitely tried to not let it impact the rating I would give the teacher but definitely it was interesting to see some of the feedback. I had some instances where I saw the same evaluation that they got last year was very similar to what they got this year. I think that's where the consistency in the evaluation tool really comes out and shows that if you have proper training, you could potentially see similar variables when you're going in and doing evaluation.

Nuances. Another way in which principals' cognitive schemas impacted how principals thought about and ultimately evaluated the teachers in their building was these principals' beliefs

in the need to consult previous teacher evaluation data when evaluating teachers within the current school year. How principals thought about and ultimately decided whether or not to consult this information impacted these principals' thoughts and behaviors during the evaluation process. For example, overwhelmingly, principals with high-levels of experience said they did not consult prior teacher evaluation information or data when conducting teacher evaluations in the current school year. When asked why they did not want to look at this data, these principals stated that each teacher deserved to start a year fresh and without any previous information influence the thinking of the evaluator. Additionally, principals with high-levels of experience were confident that their prior knowledge and experiences as a principal were enough to judge a teacher's performance in the current year.

The principals in this study with low-levels of experience overwhelmingly stated they looked at previous teacher evaluation information before and during the process of evaluating teachers in the current year. The reasons these principals provided regarding why they wanted to know this information included wanting to know what the previous evaluator had noticed in previous years and getting a more complete picture of individual teachers before providing their own evaluation. Interesting to note, even principals who were not new evaluators and were in their second or third year of evaluating the same teachers in their school looked back at how they had rated teachers in previous years. For example, Mr. Sherman who has been a principal for four years, all at the same school, said he looks at how he rated teachers in prior years because he wants to make sure he is consistent with his approach to evaluations from year to year. None of the principals with high-levels of experience looked back at how they had rated teachers in prior years (at least, not while conducting a teacher's current evaluation).

Although there was a distinction between principals with high-levels of experience and their less experienced peers and how these principals used prior teacher evaluation information, the findings were not unanimous. For example, Ms. Cohen (10 years of experience) said while she does not review a teacher's past evaluation scores before evaluating in the current year, over the summer months she does review this information to see if her teachers' are progressing. She said:

I use student assessment data from the entire year. Yes. I do look at that data. As far as their previous evaluation, not when I'm evaluating. I had a teacher last year that was on probation. I put her in a new spot this year. She's not knocking it out of the park. I wouldn't want that to sway the way I'm thinking I guess. (Over the summer) I do look at it to see are they moving forward, did they go backwards. Then I'll sort of prepare myself because they know what it is. If it went down I need to validate why did it go down. I will let them argue a point. If I don't give them credit for something and they can tell me, "I did do that, Ms. Cohen. This is how I did it." You can tell when someone's making something up unless they're a really good liar. If they are that's going to come out in the end sometime. Everything comes to the surface eventually.

One principal with low-levels of experience, Mr. Jarmel (one year of experience) said he never looks at or review previous teacher evaluation information. Mr. Jarmel believed he should not be influenced by what prior evaluators had written or observed and that wanted to form his own opinions about the teachers in his building. However, Mr. Jarmel provides the only example of a principal with a low-level of experience who expressed the opinion of not wanting to review prior teacher evaluation information before or during evaluating teachers in the current year.

Subtheme Three: Accurate Reflection of Teacher Effectiveness

A third way in which principals' cognitive schemas impacted the way in which they thought about and implemented teacher evaluation policies was individual principals' beliefs on the accuracy of their teacher evaluation system. Principals with high-levels of experience overwhelmingly believed their current teacher evaluation system was an accurate representation of teacher effectiveness. For example, when asked if she believed the final evaluation score a teacher received was an accurate representation of that teachers' effectiveness, Ms. Goldstein said:

Yeah, I do. I truly do, because of the way I do it, because of the dialogue I've had, and the things that I've observed to provide evidence to support why I feel they were where they're at. Yes, I do feel like it's a pretty accurate reflection.

Ms. Goldstein was very confident in her ability to evaluate teacher performance and because of her belief in her ability as an evaluator, Ms. Goldstein felt any system that she used would produce an accurate evaluation of teacher effectiveness. Mr. Bania (10 years of experience) added to this sentiment and said:

Well, I think, yes, the teachers that are effective get marked as effective or highly effective. In the observations I've done and the rubric scores I've given them, it pretty much—when I see the score I'm like, “Yeah, I think that's what I've observed as them as a teacher.

Mr. Bania, much like Ms. Goldstein, was confident that his teacher evaluation system and his teacher evaluation ratings were an accurate representation of teacher effectiveness because the scores matched what he cognitively thought was effective teaching. When asked if he felt his experience as an evaluator aided in his confident in the accuracy of his teacher evaluation system

Mr. Bania said that was a fair statement. He added that he faithfully implemented his teacher evaluation system and the end result was an accurate rating of teacher effectiveness for all of his teachers.

Dr. Wexler (nine years of experience) perhaps best describes the sentiments felt by all principals with high-levels of experience regarding the accuracy of her school's teacher evaluation system. Dr. Wexler believed her teacher evaluation policy and system was very subjective because she believed any evaluation done by human beings has the potential to be subjective. However, because she has been a principal for nine years, Dr. Wexler felt she knew how to eliminate this subjectivity and accurately and fairly evaluate all of the teachers in her building. Dr. Wexler explained through her principal training and her experience observing teacher classroom instruction, she was able to make accurate determinations of teacher quality. In short, Dr. Wexler was confident in her ability to evaluate teachers accurately, regardless of the system she was using.

Overall, because principals with high-levels of experience had confidence in themselves as evaluators they had confidence that final teacher evaluation scores were an accurate representation of teacher effectiveness. Additionally, these principals believed strongly their current teacher evaluation system produced accurate teacher evaluation scores and results. This belief resulted in principals with high-levels of experience typically following these systems with fidelity – at least what these individuals believed to be fidelity to this system.

While their more veteran peers were confident in the accuracy of their teacher evaluation system, an analysis of these data suggests that principals with low-levels of experience do not think their current teacher evaluation system is an accurate reflection of teacher effectiveness.

Perhaps most pointedly when asked if he felt his school's current teacher evaluation was an accurate reflection of teacher effectiveness, Mr. Jarmel said:

No, because I think it's so much more than just a rubric. I know that we've been working hard. I think it's more than just going in for the 40 minutes to sit and do an evaluation on them. If I'm in the classrooms every day, and I see what's going on, that's more to me valuable to the teachers because I can stop right then and offer suggestions and supports. Principals that don't go into the classrooms regularly, I don't see how any evaluation you do could be fair or consistent for teachers.

Mr. Jarmel went on to articulate that his policy did not allow for these extra observations to be counted towards his teacher's evaluations, so while he knew that he was helping teacher he also knew that if a teacher had a bad lesson during an "official" evaluation this bad lesson would be reflected in their evaluation rating and the rating might not be the best reflection of that teacher's effectiveness. If he had control of how to evaluate teacher's Mr. Jarmel "would use many short visits to teacher's classrooms" to evaluate instruction throughout the school year.

Ms. Robbins also was not confident that her district's teacher evaluation system as an accurate representation of teacher effectiveness. She said:

Sometimes I feel like there's pieces on there that are not—there's pieces I'd like to see there that really aren't there. For example, again, coming back to the tone that a teacher uses with a child, if they're respectful with the child or not. I'm not sure that that's really there. I feel like it's really important. It frustrates me sometimes when I do have an issue with a teacher who isn't addressing students with respect. That's something that we're really working on. It's not really there. It's not there explicitly. I feel like I'm having to work it into something where it really—the verbiage isn't there. That gets a little

frustrating sometimes. But it doesn't stop the conversation, because that's an expectation in the building out. We'll work it through it in another way. It will definitely be part of the evaluation, the observation, and it will be something that we're going to discuss every time, because it's an area of focus for that teacher.

For Ms. Robbins, something she valued greatly and thought was a measure of teacher effectiveness was not included in his district's teacher evaluation system. As a result, she lacked confidence in the accuracy of this system.

Ms. Steinman was also not confident that her school's teacher evaluation system captured all that the teachers were doing or that they produced an accurate representation of teacher effectiveness. Ms. Steinman's lack of confidence impacted how she ultimately scored the teachers in her building. Ms. Steinman said:

I'm still very sensitive. I will admit it's very hard for me to give a minimally effective or ineffective or missed opportunity. I still feel that. I don't know if that's good or bad.

Maybe I will always feel that. As a teacher, I always strive to be highly effective. Then you still have to base things on reality and what you're seeing and really trying to use it for growth versus bashing. It's not a tool to be bashed with. It's a fine line. I haven't arrived there yet. I think that I still think more like a teacher than an administrator yet.

That will come later.

Ms. Steinman went on to articulate that partly because of her lack of confidence in the accuracy of her school's teacher evaluation system, she was hesitant to rate teachers critically and she typically defaulted to higher evaluation scores. Giving teachers the benefit of the doubt and defaulting to more favorable ratings was a common sentiment amongst principals with low-levels of experience. Partially because they lacked complete confidence in their evaluation

system and partially because they lacked complete confidence in their abilities as an evaluator, these principals were more likely to rate teachers higher than their more veteran peers would rate their teachers. For example, Ms. Steinman (one year of experience) said:

I think that it has been a transition for me in general to switch my mindset from teacher to administrator. I'm still not there yet. My admin team will tell me frequently that I still think very much like a teacher. I don't know that I think that's bad. It has allowed me to create some very good relationships with my staff this year, which has been great.

Nuances. Another example of principals' cognitive schemas impacting how these principals evaluated their teachers was the principals' belief of whether or not their teacher evaluation system was an accurate representation of teacher effectiveness. Individual principals' beliefs were clearly divided between experience levels. Principals with high-levels of experience generally took on the mindset of an administrator who believed in the accuracy of their current evaluation system. Principals with low-levels of experience generally questioned the accuracy of their system, typically thinking more from the teachers' perspective.

Although overwhelmingly principals with high-levels of experience said they believed their current teacher evaluation system was an accurate reflection of teacher effectiveness, one principal with high experience, Mr. Bookman, did not believe his system was an accurate reflection of teacher effectiveness. Mr. Bookman did not think his system "account for all that teachers did" and his district's current system was missing certain components. Mr. Bookman explained he was able to provide final teacher evaluation ratings that he felt were accurate, because of his ability to "work around" his teacher evaluation system.

The majority principals in this study with low-levels of experience stated they did not

think their current teacher evaluation system was an accurate reflection of teacher effectiveness. In all four of six principals did not believe in the accuracy of their system, while two principals, Mr. Sherman and Ms. Chang believed their system was an accurate reflection of teacher effectiveness. Mr. Sherman and Ms. Chang thought along the same lines as their more veteran peers. Interesting to note, both Mr. Sherman and Ms. Chang were in their fourth year as a school principals at the time data collection. These two principals were the most experienced principals in the low-experience group.

Subtheme Four: Hiring Decisions

All principals in this study reported their districts used teacher evaluation scores to make hiring and layoff decisions. As a result, how principals rated teachers had the potential to impact these teachers' future employment. The direct association of teacher evaluation ratings and future teacher employment was not lost on principals, who consistently referred back to "the enormity" of these evaluations. However, an analysis of these data show a clear distinction between how principals with high and low experience levels think about considering teacher evaluation data when making hiring decisions. Principals with high-levels of experience generally did not look at or consider a teacher's prior evaluation score when thinking of hiring a new teacher. However, principals with low-levels of experience asked for and looked at prior teacher evaluation data before making hiring decisions. All principals in this study noted information such as the credentials, conversations with former employers, interviews, and how candidates made "data-driven instructional decisions" influenced who they hired. However, the use of previous teacher evaluation scores varied by experience level.

For example, Mr. Bania said:

We always say when we go to hire somebody it is a million dollar decision. We want to make sure that they understand our philosophy and how they answer (interview) questions will depend on whether or not I hire. It's the interview. It's whether I can talk to different community members. We call references who know that person. I don't look at evaluation scores at all.

For Mr. Bania, and all of the principals with high-levels of experience, a teacher's previous evaluation score was not considered a central or important aspect of hiring that teacher. Principals with high-levels of experience were much more likely to say they relied on things such as the interview with the candidate, if a candidate "fit" with their school and philosophy, and if this teacher had the right credentials.

Another example of a principal with high-levels of experience and their lack of use of previous teacher evaluation scores when making hiring decisions comes from Mr. Bookman who explained he too never looks at this information. Mr. Bookman explained that he knew how to select the right teachers for his schools based on interviewing potential candidates and simply talking to them about their teaching beliefs, mindset, and philosophy. Mr. Bookman said he did not put much stock into a teacher's prior evaluation score for a number of reasons, including the relationship that teacher might have had with a previous evaluator (good or bad) and because of the context of the school. Mr. Bookman noted even if a teacher had a previous score of highly effective, that means very little to him because everyone in that school may be been rated highly effective. Mr. Bookman noted as he gained experience as a principal he relied on past evaluation scores less and less when making hiring decisions and currently he does not look at this information at all.

While their more experienced peers tended not to rely on previous teacher evaluation data when hiring teachers, principals with low-levels of experience were much more likely to seek out these data before making hiring decisions. For example, Ms. Chang (four years of experience) said, “We don’t want an ineffective teacher teaching our students. Yes, we do look at previous scores if they’ve been teaching...its part of the puzzle.” For Ms. Chang, it was important to know as much about a teacher as possible for making a hiring decision. Ms. Chang wanted as much information as possible before filling any vacancies she had and therefore she would look at previous teacher evaluation information, mostly for their evaluation score and for any comments and/or feedback provided by the previous evaluator and Ms. Chang would use this information to make her hiring decision.

Mr. Sherman said, “I mean yeah (he does look at and consider prior teacher evaluation scores when hiring a new teacher). We try to get effective teachers. We try to gauge our school and look at our school and say where there’s the biggest need.” Mr. Sherman went on to articulate that he would not consider hiring a teacher unless this teacher’s previous evaluation score was effective or highly effective. Mr. Sherman felt that anything less than a rating of effective reflected poorly on the teacher and therefore he did not want this teaching working in his school.

One final example illustrating that principals with low-levels of experience tended to rely or consider relying on previous teacher evaluation information while evaluating teacher in the current year comes from Ms. Robbins who said, “I would. It hasn’t come up yet, but I wouldn’t accept somebody that was minimally effective if I had a choice.”

Nuances. An analysis of the data reveals several nuances regarding how principals’ cognitive schemas impact how principals consider prior teacher evaluation information when

making hiring decisions. Although this finding is not directly related to teacher evaluation policy implementation, it does speak to what principals think about and value while making hiring decisions based on teacher evaluation information. As evidenced from the analysis of these data, principals with high-levels of experience typically believed in their ability to identify a high-quality teacher, without needing to review that teacher's previous evaluation scores. These principals believe they can look at teacher credentials and most importantly learn about these teachers through interviews and can decide who would be a good fit for their school. However, their less-experienced principal peers almost always looked at prior teacher evaluation scores before hiring any teacher in their building. These principals look at these data in part due to the fact that these less-experienced principals wanted to have a complete picture and perhaps some validation that they are making a strong hiring decision. However, one principal with high-experience, Ms. Hamilton, said her district required her to review and consider this information before hiring a teacher. Ms. Hamilton said while she looked at previous teacher evaluation ratings that she used this information as a "tie-breaker" when two candidates seemed equal (based on interviews and credentials). All principals in this study with low-levels of experience stated they did look at previous teacher evaluation information before making hiring decisions (or at least they would once an opportunity to hire teachers came up, as some new principals had not yet experienced hiring any teachers to their building).

Chapter Summary

The analyses in this chapter demonstrate that principals' cognitive schemas influence how they think about implementing their teacher evaluation system, in part, by the type of sensemaking in which they engaged. Specifically, principals with high-levels of experience typically engaged in individual sensemaking where they made sense of their teacher evaluation

system by themselves with little other support or outside information. The type of sensemaking in which these principals engaged had implications for how these principals thought about, communicated, and ultimately carried out their school's teacher evaluation policies. Principals with high-levels of experience were less likely to use previous teacher evaluation data when evaluating teachers, were less likely to use teacher evaluation information when making hiring decisions, and were more likely to believe their teacher evaluation system was an accurate reflection of teacher effectiveness. All of these characteristics fit into the individual sensemaking framework as the principals engaged in each of these tasks relying on their own sensemaking of the process of evaluation teachers (Ganon-Shilon & Schechter, 2016). These findings support prior literature suggesting principals rely on their prior knowledge when attempting to implement school level policies (Coburn, 2005; Spillane et al., 2002). In short, principals with high-levels of experience draw on their experience as an administrator, because these experiences are what makes the most sense to them.

While their more veteran peers engaged in individual sensemaking principals with less experience typically engaged in collective sensemaking. These principals were more likely to engage in relational leadership and include other teachers in their thought process and discussion of teacher evaluation policy implementation. These principals were also more likely to look at previous teacher evaluation scores when evaluating teachers in the current year, were more likely to look at previous teacher evaluation scores when making hiring decisions, and were less likely to believe their teacher evaluation system was an accurate reflection of teacher effectiveness. Principals with low-levels of experience tended to draw on their experiences as a teacher because these experiences make up a majority of their professional educational experience.

In summary, the principals in this study with high-levels of experience engaged in individual sensemaking, drawing on their own experiences and beliefs about the goals of education, which impacted how these principals thought about the process and purpose of teacher evaluations and how these individuals actually evaluated teachers in their building. Their less experienced peers were more likely to collectively navigate the teacher evaluation process in part because they were more sympathetic to their teachers. It makes sense that individual cognition may change as principals gain experience. However, these findings suggest how teachers are evaluated varies by the amount of experience of the evaluator. This variation may be one explanation why consistent teacher evaluation policy implementation remains a challenge. For example, teachers with an identical skill sets, identical instructional practices, and identical classroom impact could receive a vastly different evaluation rating simply based on the experience level of the principal who does the evaluation. Additionally, based on these findings one might assume that teachers who work in schools with less experienced principals may receive more favorable teacher evaluation ratings than their peers in schools with more experienced principals. A complete discussion of the implications of these findings is found in chapter seven.

Chapter 6: The Role of External Context and Experience in Principal Learning and Implementation of Teacher Evaluation Policies and Systems

“I feel the pressure for the teachers because they want to be highly effective, but it’s really hard to be highly effective when your students are failing. Just getting them to understand that. If we weren’t a priority school, it would be different.” – Mr. Jarmel (high-pressure environment)

This chapter answers my second and third research questions: What role does external context (e.g. high-pressure vs. low-pressure environments) play in shaping principal learning and enactment of teacher evaluations systems and how, if at all, do principal experience and context interact during the policy implementation process? The first section of this chapter answers my second research question. When analyzed through the lens of cognition and specifically sensemaking theory, two important themes emerge from an analysis of the data. First, principals who work in high-pressure environments perceive a pressure to differentiate teacher evaluation ratings among teachers in their building. Second, principals in high-pressure environments did not believe their evaluation system accounted for the challenges their teachers faced (e.g. working in low-income communities, working with transient student populations, and teaching students who enter their classroom several grade levels behind academically). The remainder of this chapter answers the final research question which examines more closely how experience and context interact, if all, during the implementation of teacher evaluation policies and systems. An analysis of the data provides evidence that experience and context do interact and influence principals’ thoughts and actions around teacher evaluation policy implementation in several meaningful ways.

Theme One: Differentiating Teacher Evaluation Ratings

The first theme that emerged from an analysis of the data was principals who work in high-pressure environments perceive a pressure to differentiate teacher evaluation ratings among teachers in their building. In this study, all 12 principals were the sole evaluators of their teaching staff. The school district or charter school authorizer of each of these 12 principals tasked these principals with implementing a formal teacher evaluation policy, including providing specific directives of how and when to observe teacher instruction, how to account for student assessment data in evaluations, and how to use the results of these evaluations for human capital decisions. Despite the formal and prescriptive nature of these policies, an analysis of the data suggests external context played a prominent role in how principals thought about teacher evaluation policy implementation as well as how they ultimately evaluated the teachers in their building.

One way in which external pressure impacted how principals in this study evaluated teachers is how principals thought about and ultimately assigned teacher evaluation ratings. Principals in high-pressure environments were more likely to rate teachers critically than their peers in low-pressure environments. The principals in these high-pressure environments provided several explanations as to why they rated teachers critically. First, some principals felt an added pressure from district administrators to have some form of differentiated ratings among the teachers in their building. Specifically, these principals perceived a pressure limit the number of teachers they rated as effective or highly effective. Second, principals in high-pressure environments reported feeling a pressure from teachers in their building to differentiate ratings amongst teachers because these teachers were aware of the consequences of these scores for their future employment. Although none of the principals in this study said they received directives to

differentiate the ratings they gave their teachers, these principals did suggest that they received some type of message from district administrators, including superintendents, that there should be some distribution of teacher effectiveness ratings. Finally, principals in high-pressure contexts put pressure on themselves to critique teachers' performance because they knew the status of their school (in terms of their state ranking on Michigan's Accountability Scorecard) did not reflect a school where all teachers were effective or highly effective. For example, Mr. Jarmel said:

I know I feel the pressure for the teachers because they want to be highly effective, but it's really hard to be highly effective when your students are failing. Just getting them to understand that. If we weren't a priority school, it would be different.

Mr. Jarmel went on to say because his school's test scores were so low in previous years he would not be able to justify rating all teachers effective or highly effective. Mr. Jarmel went on to explain that simply knowing his school was underperforming on state assessments was enough for him to know some teachers needed to be rated less than effective. In his mind, low student achievement on state assessments equated to less than effective teaching.

Mr. Ramon said while his administration was supportive of how he assigned teacher evaluation ratings, he understood from informal conversations that there was an expectation to differentiate teacher evaluation ratings in the district. Mr. Ramon recalled conversations with his superintendent about the observational rubric and the "high expectations" of the rubric. Mr. Ramon said, "When you look at what highly effective is, those are some really, really high expectations." Mr. Ramon took these conversations with his superintendent to mean that he should look very carefully at the domains of the evaluation rubric to make sure teachers met

these criteria. Mr. Ramon explained he perceived a pressure to make sure if he scored a teacher highly effective, he could point to adequate evidence to validate this rating.

Mr. Ramon also recalled having several contentious conversations with his teaching staff about their final ratings. He said:

We've had some interesting conversations about the rating of ineffective, effective, and highly effective. You know, the conversations among professionals will occur, and you might have a teacher who got highly effective who might tell a teacher who got effective, and they're like, oh, why I didn't get it? You do have a lot of interesting dialog and dynamics in regards to explaining the process. Your employment really can be contingent upon the results (of your evaluation), so I had several teachers who we really debated what their final ranking ended up being. I think I only changed one and it was really, really tough for me because even as I told them, when you look at the specifications that are listed within our model for evaluation, when you look at what highly effective is, those are some really, really high expectations. I potentially have a union action that I will be dealing with in the next couple weeks about an evaluation as well.

For Mr. Ramon, these challenging conversations with his teaching staff occurred each year leading up to and during teacher evaluations. Mr. Ramon felt pressure from his staff to make a clear hierarchy of teacher evaluation scores within his building.

Another principal who worked in a high-pressure context, Ms. Hamilton, told a story of how a teacher in her building came into her office and vigorously debated her evaluation rating of "effective." Ms. Hamilton explained that after about an hour of debate and going back and forth with this teacher on specific rubric scores, this teacher directly told Ms. Hamilton that she knew a fellow teacher, who she considered less effective, received the same evaluation rating. In

this teacher's mind, this rating was not only unfair, but had the potential to impact this teacher's career. This teacher was at risk of losing her job if this district experienced layoffs, because she had fewer years of experience (which was this district's tie-breaker if teachers received the same evaluation rating). Ms. Hamilton explained teachers thinking about their job security was a real concern for many teachers in her district, because her district almost always experienced teacher layoffs. Ms. Hamilton said she was confident she correctly rated each teacher, but she understood why this teacher was arguing for a higher score. Although Ms. Hamilton reported she ultimately did not change this teacher's score, when reflecting on this meeting Ms. Hamilton acknowledged teacher's future employment was something that was always in the back of her mind when she evaluated teachers in the future. She said:

I think it (thinking about comparing final evaluation ratings of teachers) does reflect my reality and maybe my reality isn't somebody else's reality. I use this (the evaluation process) as a tool to grow them. In order to grow them I have to grow myself. I always go back and I look, what did I do? What could I have done different? Sometimes we place the wrong person, to me, in the job.

Ms. Hamilton concluded this story by suggesting that she perceived a pressure from almost all of the teachers in her building to differentiate among their final evaluation score due to the amount of layoffs experienced by her district. Ms. Hamilton explained she had constant conversations with teachers about their evaluation scores as compared to their peers and these conversations led Ms. Hamilton to thoroughly examine and at times reconsider her final evaluation ratings of the teachers in her building.

Unlike their peers working in high-pressure environments, the principals in this study who worked in low-pressure environments did not perceive pressure to rate any set number of

teachers in any category (effective, highly effective, etc.) and said that although it would be unlikely for all teachers to receive a highly effective rating, if that is what they all earned, that is what they would be rated. As Ms. Chang said, “The chips fall where they may. It is what it is.”

Mr. Sherman explained that he did not feel any pressure to assign any specific score to any teacher in his building. He said:

Again, the evaluation is going to be what it is. Even if it's Miss X and Miss X is one of our great teachers. If she starts going down (in terms of her evaluation rating), it's going to be because that's what she was doing in her classroom. That's my mind. Now if somebody wants to engage me then that's fine we can talk. Then if they say something that I believe in, I have changed them (the evaluation score). It could be something like, oh I'm sorry you're right I missed that. If they fight for it and it's right, I will change. If it's not, it stays the same.

Mr. Sherman went on to explain that he let his teachers have some form of conversation with him regarding their final evaluation score, but he was not under any pressure to assign teachers certain evaluation ratings. Therefore, he was comfortable adjusting these scores if the teacher made a compelling case.

Ms. Goldstein also indicated she felt no pressure to differentiate amongst the evaluation ratings she provided the teachers in her building. She said, “Evaluations are everything, but I say that tongue-in-cheek because your evaluation isn't everything. If you're doing the job you were hired to do and doing it well, your evaluation is going to be highly effective.” For Ms. Goldstein, final teacher evaluation ratings were a result of teacher actions, student assessment data, and overall professionalism. Therefore, depending on the results of these teacher actions, all teachers might be rated similarly. Although she said she has never given all teachers highly

effective, Ms. Goldstein said that if all of the teachers in her building met the criteria to be highly effective she would not hesitate to assign all of her teachers a rating of highly effective.

Nuances. One way in which external context impacted how principals thought about and ultimately rated teachers was the pressure perceived by principals when assigning final teacher evaluation ratings. However, principals who worked in high-pressure environments were not the only principals who perceived a pressure from teachers to differentiate evaluation scores. This perceived (and real) pressure did come up in several interviews with principals in low-pressure environments, although much less often. For example, Ms. Steinman said she felt a pressure from teachers to rate them as effective or highly effective. In her opinion she had many strong-willed teachers who believe they were highly effective and these teachers knew how to argue/make a compelling case for themselves. These teachers also seemed to know all of the teachers' evaluation ratings and although they might be okay with an effective, they would argue their score if they felt it was not accurate compared to other teachers in the building. Other principals in low-pressure environments (and the teachers I interviewed) certainly referenced the importance of their final evaluation scores as in almost all instances employment decisions were based on these scores. However, principals in high-pressure contexts felt this pressure much more strongly, in part due to these districts laying off teachers annually. The low-pressure contexts in this study rarely experienced teacher layoffs.

In summary, one way in which external pressure impacted the way in which principals thought about and ultimately rated the teachers in their building was the perceived (and real) pressure administrators felt from teachers and district level superiors. Principals who worked in low-pressure environments reported experiencing some perceived pressure; however, they never experienced pressure from the district level to distribute teacher evaluation ratings and the

pressure these principals felt from teachers was different from the pressure perceived from their high-pressure peers, because of the lack of teacher layoffs generally experienced in low-pressure schools.

Theme Two: What do Teacher Evaluations Measure?

Another way in which external pressure impacted how principals thought about and ultimately implemented teacher evaluation policies was principals' beliefs about the efficacy of these systems in measuring the true performance of the teachers in their building. Principals in high-pressure environments did not believe their teacher evaluation systems accounted for all of the challenges the teachers in their contexts were facing. This belief resulted in principals looking for creative ways to increase a teacher's their final evaluation score. For example, Mr. Bookman said:

It's tough because teachers aren't going do it (be at their best) all the time. They're just not. It's human nature. I'm not looking to lambaste anybody, but I feel like sometimes that's what the evaluation process does. I don't think there's any tool that's going accurately evaluate all the things that teachers are doing. It's that human factor that you just can't evaluate.

Mr. Bookman, along with other principals in high-pressure environments, reported that teachers in their buildings had more challenges than other teachers with whom they have worked in contexts with less pressure. For example, Mr. Bookman told a story of one teacher who had a goal of perfect attendance (or as close as possible) for her whole class for the entire year. Aiming for perfect attendance was an ambitious goal, as many of this teacher's students were chronically absent. However, in the second part of the year, her students rarely, if ever, missed a day of school. Mr. Bookman said the relationships this teacher developed with students and parents and

her efforts to make school enjoyable caused the increase in attendance. Mr. Bookman went on to say that “obviously attendance is associated with learning and other growth”, but there is nothing on his teacher evaluation system that can “reward” teachers for increasing student attendance. As a result Mr. Bookman said he would try and factor in increased student attendance into this teacher’s final evaluation rating in the professionalism part of observational rubric. Although evaluating a teacher in this way may have been stretching what the professionalism part of his system meant (and he knew this – it was supposed to include things such as number of teacher absences and the amount of professional developments a teacher attended and if they were involved in student-related activities beyond teaching, such as coaching a sports team or mentoring other teachers) he gave this specific teacher the highest possible professionalism score because of her accomplishments of increasing student attendance. In his mind, Mr. Bookman believed it was his job as an evaluator to account for all that his teachers did, even if his current evaluation system did not.

Mr. Bookman was not alone in his belief that his school’s current teacher evaluation system did not account for all of the challenges faced by the teachers in his building. Ms. Cohen agreed that her evaluation system did not capture the relational part of teaching. She said:

Here in the urban setting relationship is everything. I have a hard time measuring those soft pieces. Our tool was great at measuring the data and those things that you can see like how is your classroom set up, is your classroom organized, is it functioning so that people can travel from place to place, are your transitions good. All that stuff that you can see is easy to evaluate. The tool was still missing that relationship piece.

Mr. Bookman and Ms. Cohen felt their evaluation system, particularly given their context, did not include crucial pieces that reflected teachers' impact in the classroom, and as a result these principals looked for ways to credit their teachers on other areas of the evaluation.

Mr. Jarmel also thought his system did not do an adequate job capturing the "whole teacher" and everything a teacher was doing on a daily basis. He said:

I mean you have to hold people accountable for their jobs, but I think it's so much more than that. I know we all need to be held accountable for our jobs, but I don't quite know what that is. I think it's more than just going in for the 40 minutes to sit and do an evaluation on them. If I'm in the classrooms every day, and I see what's going on, that's more to me valuable to the teachers because I can stop right then and offer suggestions and supports. Do you want the development of your staff and to have high quality, or is it just a dog-and-pony show that you get a couple of times a year?

Mr. Jarmel explained that he factored in all of the visits he had with teachers throughout the school year. Although technically he was supposed to evaluate teachers twice yearly, in 30-45 minute observations of their instruction, Mr. Jarmel used observation data he collected throughout the school year when assigning his teacher's final evaluation score. In his mind, this data point was more valuable both to him and his staff than two scheduled observations of teacher instruction where the teacher might perform above or below their actual ability.

While principals who worked in high-pressure environments thought their evaluation system did not accurately encapsulate all that a teacher did, principals in low-pressure environments were more likely to report their teacher evaluation system did capture all that their teachers did. For example Ms. Steinman said:

We do the dimension ten (which accounts for everything outside of the classroom). Do they complete things on time? Do they attend work? All of those things. How do they conduct themselves during parent interviews, parent conferences, parent contacts with home? Their attendance based on sick versus leave versus conferences that they've attended, and any disciplinary actions that may have occurred have to go in that, the professional learning number ten area.

Ms. Steinman believed that because of the way her district set up their current teacher evaluation system and used dimension ten of her evaluation rubric, the system did in fact account for everything that the teacher did in the current year, including things both inside and outside of the classroom.

Mr. Bania also suggested his district's teacher evaluation policy accounted for everything teachers in his school did. He said his policy had certain aspects, such as professionalism, that allowed him feel confident in the accuracy of his teachers' final evaluation scores. He said, "The teachers that are effective that get marked as effective or highly effective. I believe we do a good job in this district here. It's something we look at as a whole, not just by building, but as a whole district." Mr. Bania was confident that his district had taken the necessary steps to ensure that their current teacher evaluation system account for everything that the teachers were expected to do as a teacher in their district.

Dr. Wexler makes another argument suggesting that principals in low-pressure contexts felt their teacher evaluation policy was comprehensive and accounted for all that their teachers were asked to do. Dr. Wexler said:

Teacher effectiveness seems to be so much bigger than just a piece of paper. I think it (her school's teacher evaluation policy/system) captures in essence what they do. How

they interact with their kids, raise scores, raise the self-worth of our kids. You can't really measure that, but you can see it in the kids. You can see it in the classrooms and the way that they interact.

Dr. Wexler went on to say that she believed her school's evaluation system did a fine job capturing all that she expects from teachers and as a result, the final evaluation ratings she assigned her teachers was an accurate representation of not only her teachers' instructional effectiveness and their ability to raise student test scores, but also their ability to relate with student and improve student self-esteem and confidence.

Nuances. A second way in which external pressure impacted how principals thought about and rated the teachers in their building was the principal's belief and perception that his or her teacher evaluation did or did not account for all that their teachers were asked to do in their context. While principals in low-pressure environments thought the evaluation system did a fair job capturing all that their teachers did throughout the school year principals in high-pressure environments continually referred to the fact that teacher evaluations did not encapsulate the many challenges faced by their teachers. At times this belief resulted in principals in high-pressure contexts looking to give teachers additional credit, ultimately raising some of these teachers' final evaluation scores.

Principals who worked in high-pressure environments were not exclusively critical of their evaluation system in terms of it capturing all that their teachers did throughout the school year. For example, Ms. Chang (low-pressure context) did not believe her school's current system accurately accounted for all responsibilities of her staff. Ms. Chang was complimentary of her current evaluation system in many places, but other times, specifically in how her district accounted for student growth, she was critical and did not believe how the district measured

growth was an accurate or fair representation for her teachers. Additionally, two principals in high-pressure contexts, Ms. Robbins and Mr. Ramon, felt while their teacher evaluation had limitations, it did account for all their teachers were asked to do. Other principals in low-pressure environments noted some things they would like to see changed or added to their current teacher evaluation system, but these administrators overwhelmingly believed their current system account for most if not all things done by their staff and therefore was an accurate reflection of teacher effectiveness. Their peers in high-pressure environments were much less likely to have a similar mindset. In summary, a second way in which external pressure impacted the way in which principals thought about and ultimately rated the teachers in their building was the belief of principals that their current teacher evaluation system did or did not account for all that their teachers did throughout the school year. This belief resulted in some principals manipulating the final evaluation ratings of teachers.

How do Experience and External Pressure Interact during the Implementation Process?

The final research question that guided this work examines how principal experience and external pressure interact during the process of teacher evaluation policy implementation. Specifically, do principals with certain experience levels and who work in contexts with differing amounts of external pressure think about and implement their school's teacher evaluation system in similar or different way? When analyzed through the lens of cognition and specifically sensemaking theory several themes emerge from each grouping of principals. The thoughts, beliefs and actions of principals with different experience levels and facing different amounts of outside pressure had implications for how these individuals thought about teacher evaluations and ultimately how they rated the teachers in their building.

When compared to their peers in other categories principals with high-levels of

experience in high-pressure contexts believed (1) all teachers should be evaluated annually, regardless of the effectiveness of the teacher; and (2) it was their responsibility to provide teachers with specific directives of how to teach in an effort to improve their instruction and ultimately student learning.

All three principals with high experience in high-pressure environments believed all teachers in their building should be evaluated annually, regardless of the effectiveness of the teacher. These principals believed they needed as much information as possible on these teachers to make sure the teachers were improving their practice and to make sure these teachers were held accountable for their performance. Mr. Bookman said:

The whole idea of tenure going out the window where it's like everybody's on a level playing field. I was happy about, to tell you the truth, because I don't think anybody should just be guaranteed a job. If you're horrible, you shouldn't be guaranteed a job. That's not how it works in the real world. That's not how it works from my chair either. You got to perform, and you got to perform every year. As evaluators, or as evaluators and administrators, we're evaluated every year.

Mr. Bookman and the other principals in this category believed their teachers should be evaluated at least annually in order to hold teachers accountable for their performance. Mr. Bookman's experience conducting teacher evaluations gave him confidence that the data he collected during teacher evaluations would be beneficial for not only him as the principal, but also the for teacher. Ms. Cohen and Ms. Hamilton, the other two principal in this category, also believed more evaluations were typically beneficial for not only teachers, but for themselves as evaluators to make sure teachers were improving. For example, Ms. Hamilton described a teacher who was highly effective and in her words, "absolutely a rock star" one year. However,

the following year, for a variety of reasons, this teacher's performance slipped and she was rated minimally effective. Ms. Hamilton said that if this teacher had not been officially evaluated that year (as is the case with highly effective teachers in some districts) this performance would have gone unchecked, ultimately hurting the students in this teacher's classroom.

The second way in which principals with high-levels of experience in high-pressure contexts differed from their peers in other categories was these principals were most likely to give specific directives of how teachers should teach. While their peers in other categories tended to lean towards giving more support, guidance, or suggestions, these principals believed that they should be telling teachers what to do and these teachers should be following their directives. This belief was in part because of their extensive experience and in part because of the high-pressure context of their school. For example, Ms. Hamilton explained how she meets individually with each teacher prior to the beginning of the school year and they co-develop goals that the teacher will work on throughout the year. Although co-developing goals with teachers was not unique to principals with high-experience in high-pressure contexts, Ms. Hamilton was quick to note that although technically these were co-constructed goals, as the principal, she set the goals, monitored the goals, and made sure by the teacher's evaluation that he or she had made progress towards these goals. When asked what she did if her teacher's may disagree, she said "tough". Ms. Hamilton continued:

It would be less difficult if I didn't take ownership for their growth, but I take ownership for their growth. I know principals that just go in, score it, and they go to sleep at night. If I didn't take ownership for their growth, I would say it'd be a lot easier. I do take ownership for their growth, so I want the best for them and it's up to me to differentiate it for them.

Part of this belief and action of principals in this category could be due to the fact these schools typically employed less experienced teachers and these principals were aware of this fact and wanted to make sure they were giving these teachers specifics about what works in their context. For example, Ms. Cohen said:

Most of my staff has less than three years experience. I'm the first school that they've ever been too. They came straight out of college and they're fine with it (the fact that Ms. Cohen gives them specific directives of how to teach). I think the reason why they're fine with it is because in staff meetings I promote why we're doing this. I give them a mission, a vision as to why are the test scores important. I do it because I feel like I should. I realize with new teachers or people that are just new to your school, you have to say it at least four or five times before it sinks in because they've got so much on their plate.

Ms. Hamilton added, "We have a real attrition problem. The front door is like a turnstile with them (teachers) coming and going. My goal at the end of the day is to have them as successful as they can be."

In sum, the beliefs and actions of principals with high-levels of experience in high-pressure contexts differ from the other principals in this study. Specifically, these principals believe their teachers should be evaluated as often as possible in an effort to provide both the principal and teacher with feedback to improve their practice. This belief impacted how these principals thought about using teacher evaluation information. Additionally, these principals believe in giving their teachers specific directives of how to improve their teaching. This belief and this action impacted how often these principals spent in teachers' classrooms. Although principals could not increase the number of formal evaluations, they could observe teacher

instruction more often and direct the goals their teachers created, the feedback they provided teachers, and what they expected from teachers throughout the school year. These characteristics manifested themselves with these principals much more so than their peers in other categories.

When compared to their peers in other all other categories principals with low-experience in high-pressure environments (1) spent the most time in teachers' classrooms; and (2) provided more support and guidance (in terms of official and unofficial observation feedback) than their peers in similar high-pressure contexts with high-levels of experience. These findings suggest these principals provide constant and mostly supportive feedback to their teaching staff throughout the school year. This finding holds true for the type of feedback these principals provide teachers during the official teacher evaluation process. For example, Mr. Jarmel said:

I'm more of an instructional leader. I work extremely hard, so the teachers see that and respect that. I'm on the front lines with the teachers. I don't expect anything from them that I'm not going to show them how to get there. We need highly effective teachers. Making sure that all professional development surrounded by that differentiated instruction, depth of knowledge, and DEI, which is direct, explicit. Those are the key components because I want them to be highly successful. I want them to be highly qualified.

Mr. Jarmel continued that in order for him to provide the most accurate representation of a teacher's performance, he needed to spend as much time as possible in the classrooms of these teachers. Mr. Jarmel was quick to point out that all observations of teacher instruction he conducted counted towards a teacher's evaluation score, even though only two observations of instruction was used for official purposes. However, if Mr. Jarmel noticed a teacher attending professional development opportunities and taking his feedback and implementing it into

practice, this would be reflected in his teachers' final evaluation scores.

Principals with low-experience levels in high-pressure environments also believed it was not fair to rate their teachers based on one or two 45-minute observations. Like their more experienced peers who also work in high-pressure contexts, principals in this category believed that while their teachers should be evaluated at minimum annually, they also believed their teachers should be evaluated multiple times throughout one academic year. Because of this belief, these principals tended to spend as much time as possible in teachers' classrooms and observing teachers instruction. These principals said they considered all of these informal observations, not just the official observations, in their teacher evaluations. For example, Mr. Ramon said:

One of my goals this year was to visit at least four to five classes every day. I let my staff know what would I look for when I came into the classroom for identifying the goals and objectives to key vocabulary in regards to the subject area that was being covered, and that falls under, once again, informal evaluations. I ended up deciding this year to provide that instantaneous feedback that when I do go in to do regular evaluations.

Mr. Ramon explained that this year long feedback played a role in how he ultimately evaluated teachers. During one day of teacher observations Mr. Ramon and I visited four classrooms in my two-hour visit. He said visiting multiple classrooms was very common for him as he hated staying in his office. One teacher commented "He is in here all the time", providing further evidence that Mr. Ramon prioritized spending time in the classrooms of his teachers. He said that if he saw teachers taking this feedback to heart and making improvements, he would include this effort in their final evaluation. At the same time, if he observed a teacher for their official evaluation and this teacher did not perform as well as Mr. Ramon knew this teacher

could, he would not rate this teacher solely based on one below average performance. He would use all he knew about the teacher throughout the year before making a final evaluation rating.

Principals with low-experience in high-pressure environments were also more likely to provide their teachers with guidance, support, and feedback, when compared to their more experienced peers who worked in similar environments. For example, Ms. Robbins explained she believed it was her best interest to provide her teachers with support and guidance, as many were beginning career teachers and others had many challenges “outside of their control” – which Ms. Robbins described as challenges in the community, such as poverty. Because of this belief, Ms. Robbins thought it best to support her teachers, as opposed to providing them with mandates and directives about what they must do. In her opinion, her teachers were working very hard to ensure the academic success of all students and giving them mandates or deadline or directives would be counterproductive. Instead, Ms. Robbins focused on provided structured, positive feedback and as long as she felt her staff was making a good faith effort to address this feedback, she was content.

In sum, principals who worked in high-pressure environments with low-levels of experience spent the most time in classrooms of all of the principal categories and this time spent in classrooms factored into their thinking about teacher evaluation feedback and ultimately how they rated the teachers in their building. Typically, these constant observations manifested in these three principals providing support and feedback to their staff throughout the school year. Additionally, these principals used these informal observations when calculating a teacher’s final evaluation score. Spending time in teachers’ classrooms is not unique to this category of principals, but these three individuals referenced using all observations throughout the school

year in a teacher's final official evaluation much more frequently than their peers in other categories.

While principals in high-pressure environments said they took a more active and directive role in the classrooms of their teachers, principals in low-pressure environments with high-levels of experience were more likely to provide suggestions, ideas, and support that teachers could consider to improve their practice. This behavior manifested itself during principal/teacher conversations, as well as in the feedback principals provided teachers. For example Ms. Goldstein said:

I feel like, to be an educational leader, you've got to be on the front line with your people. You've got to be in the room. You've got to be hearing about what's working, what's not working, watching the behaviors of students that are making it impossible to teach. You've got to be there. You've got to be supportive in the sense that "I'm there with you, not just sitting behind my desk in an office that's 100 yards from you." That's not effective.

Ms. Goldstein went on to say that when it came to things such as supporting teaching practices that promoted change, innovation, and teachers taking risks, she would never tell her teachers how to teach or not to try something. She explained:

I think everybody should be (allowed to try new things). It helps you understand what you do well and what you need to look at. To the complexity in which we're doing it, are we pushing teachers away from teaching because of it? They walk in, one shot (references how teacher evaluations are currently structured in his district), and that's your deal. If you did good that day, you have a job. If you didn't, well, sorry about your luck. I don't understand how we got to where we're at with evaluations. I don't know, 30,

40, 50 years ago, in education, of why it got to be what it is today. Something must have happened, and now we're reactive to that, instead of being proactive with the teachers that are coming in.

Ms. Goldstein believed that evaluations should be a means of support and a way to let teachers know how they are doing and to help them improve their instructional practice. This belief was a common sentiment from principals with high-levels of experience in low-pressure environments. Principals in this category believed their teachers should be evaluated, but the emphasis of the evaluations should be changed from punitive to supportive.

Dr. Wexler also stressed that she focused on using evaluations as a way to provide support and suggestions to the teachers in her building. She said, "Well, we look at things like supporting change and innovation, how we communicate as a team, how we communicate with each other, how we communicate with our students. It's how do they influence students and others in collaborative ways." Dr. Wexler explained that evaluations were more than a rating system of her teachers. She did believe her teachers should be evaluated and held accountable for her performance, but she believed much more strongly that evaluations should be used as a means of support and feedback for her teachers.

Additionally, principals with high-levels of experience in low-pressure environments believed their teachers were evaluated too often and the total number of teacher evaluations should be reduced for most teachers. For example, Ms. Goldstein said:

If you're doing your job, and you're moving kids, and your classroom is just ticking along, and I have evidence to support the fact that you're innovative, and you're matching our district philosophy, why do I need to evaluate you every year? Why? We're trying to not punish the best of the best by making them do enormous amounts of tedious

paperwork and give them a year off every year. What I'm hearing from them, they don't like that, because then in the year they're off, things change, and so the next year when they are evaluated, they're like, "Oh my gosh. There were so many changes. I don't even know what I'm doing now." As helpful as it was supposed to be, it's not becoming helpful. Where's the common ground there?

Ms. Goldstein thought that unless a teacher was struggling or perhaps new to the district, she should not have to evaluate all teachers in her building. In her opinion the official evaluation process was a waste of time and resources, as she could better spend her time communicating with her staff in other ways, such as informal walkthroughs and shorter, simple and point blank conversations. In short, the principals with high-levels of experience who worked in low-pressure environments were most likely to believe evaluations should provide support for teachers as well as to believe that annual formal evaluations were not necessary.

Finally, when compared to their peers in other categories principals with low-experience in low-pressure environments were likely to co-construct the evaluation process with their teachers and were the most likely to provide teachers the benefit of the doubt and negotiate with teachers during the evaluation process (when providing teachers their final evaluation rating). For example, Ms. Steinman said:

I want to make sure that I'm open and fair and consistent. I release all my walkthroughs. My teachers come and talk to me about them. "What can I do?" or "Why did you do this?" or "Did you notice this?" Sometimes I haven't. I really just try to be fair and open with them so that they know where they're at, there's no surprises. It's not going to be at the end of the year, they're like, "Oh, my goodness! I didn't even know that you didn't think

I was an effective teacher. Ultimately, when I go to complete the evaluation, all of them will come up. I still get to pick the final score.

Ms. Steinman went on to explain that she and her teaching staff had detailed conversations throughout the school year about their projected teaching effectiveness and their projected final evaluation rating. She said she also allowed teachers to voice their concerns about how she ultimately rated these teachers. Ms. Steinman was quick to point out that because she was new in her position she tended to want teachers to get the benefit of the doubt in an effort to secure positive and trusting relationships. Ms. Steinman believed building these relationships would be beneficial in the long-run as she would be able to have more difficult conversations with her staff. However, Ms. Steinman did note that her lack of experience as an administrator and her recent experience as a teacher did lend to negotiable teacher evaluation ratings and to her defaulting to higher teacher evaluation scores with her staff.

Ms. Chang also said she worked very hard to co-construct meaning around her school's teacher evaluation system and ultimately her teacher's evaluation ratings. Ms. Chang said:

I leave some blanks (in the final evaluation rubric). For certain teachers, I can't see everything in the evaluation and I ask for them to bring some evidence. Then I fill out as many boxes as I can from the information that I've gathered in those walkthroughs and observations. I'll leave some blank spots of things that we want to talk about, or if they have evidence to bring or show me, then I want to mark those boxes efficiently.

Ms. Chang went on to say that in her mind she was not negotiating scores, but allowing her teachers to provide evidence and state their case as to why they should receive a higher evaluation score. Ms. Chang said she was fine with this approach to evaluations because as a former teacher she knew an evaluator could not capture everything that was happening in a

classroom. Therefore, if she didn't see something she would not mark this teacher's evaluation score down. Instead she allowed these teachers to present their case for why they should receive a higher score.

The other principal in this category, Mr. Sherman, talked extensively about how he let teachers "argue" for a higher score. Although letting teachers dispute their final evaluation rating did not mean Mr. Sherman changed a teacher's score just because they argued, he encouraged his teachers to have these conversations and fight for themselves. If they made a compelling case, he would change the score, because in his mind, the teacher knew best what they did every day in the classroom and just because Mr. Sherman might not have seen something, as long as the teacher could point to some evidence, he was fine giving teachers the benefit of the doubt and ultimately a higher rating. In short, the three principals in this category were more likely than their peers in other categories to negotiate teacher evaluation ratings with their staff. These conversations were often ongoing throughout the school year and teachers had a chance to influence their final evaluation rating based on these conversations outside of their classrooms.

Chapter Summary

The results of this chapter suggest external context influences how principals think about implementing their teacher evaluation system in several ways. First, principals who work in high-pressure environments believed they had an added pressure to differentiate among teacher evaluation ratings. This perceived (and real) pressure typically caused principals in these contexts to rate teachers more critically than principals in low-pressure environments. Second, principals in high-pressure environments did not believe their evaluation system accounted for all that their teachers were doing and the challenges the teachers in their context were facing.

This belief resulted in these principals looking for ways to include some of these things, even if the policy did not call for it.

The second section of this chapter highlights differences between different categories of principals. In sum, principals with high-levels of experience in high-pressure environments believe their teachers should be evaluated more often and believe it is their job to provide specific directives to teachers in regard to how to improve their teaching practice. Principals with low-experience in high-pressure environments believed it was their job to provide their teachers with support and guidance. Additionally, these principals spent the most time in teachers' classrooms compared to principals in all other categories. Principals with high-levels of experience in low-pressure environments believe their teachers should be evaluated less often and believe in providing suggestions and support to teachers. Finally, principals with low-experience in low-pressure environments were most likely to co-construct how their teacher evaluations looked in practice with the teachers in their building and were most likely to give teachers the benefit of the doubt when assigning teacher evaluation ratings. These findings suggest how teachers varies based on a combination of the pressure faced and experience of the evaluator. These findings have several implications for practice, including teachers who work in different contexts receiving different teacher evaluation ratings solely based on their current work environment. A complete discussion of the implications of these findings is found in chapter seven.

Chapter 7: Discussion, Implications, and Conclusions

Today's education policy conversation includes an increasing amount of scholarship dedicated to principals' evaluation of teachers (see for example, Donaldson & Papay, 2014; Goldring et al., 2015; Rigby, 2015; Steinberg & Donaldson, 2016). This dissertation complements and extends this growing body of literature by providing nuanced evidence of how principals' cognitive schemas impact their implementation of teacher evaluation policy and teacher evaluation systems. Analyzed through the lens cognition and specifically sensemaking theory, the results of this work indicate that principal experience as well as external context impact how principals think about implementing teacher evaluation policies and systems and ultimately how these policies and systems play out in practice. Specifically, this dissertation fills an important theoretical gap in the literature by suggesting that principals with high-levels of experience engage in individual sensemaking when implementing teacher evaluation policies and systems, while principals with low-levels of experience engage in collective sensemaking when implementing these same policies and systems. Additionally, this dissertation fills a gap in the empirical teacher evaluation literature by providing insights as to how principal experience and external context influence teacher evaluation policy and system interpretation and implementation. In this chapter I situate my findings into the broader teacher evaluation policy scholarship landscape. I then discuss the implications of the findings of this work and provide concluding remarks.

The Goals of Teacher Evaluation Policy

As research continues to show a high correlation between teacher quality and positive student outcomes, such as achievement, attendance, and graduation (Aaronson, Barrow, & Sander, 2007; Chetty, Friedman, & Rockoff, 2014; Rockoff, 2004), ensuring all students have

access to high-quality teachers is of critical importance. The pace at which teacher evaluation policies and systems are changing is one indication that governments (nationally and locally), researchers, and practitioners believe carefully and thoughtfully constructed teacher evaluation policies have the potential to realize the goal of high-quality teachers for all students, by identifying high-quality teachers and by providing better information on what makes a quality teacher. Although teacher evaluation policies and systems have changed dramatically in recent years, the goals and purposes of teacher evaluations have changed very little. Early research suggested teacher evaluations were meant to serve the general purposes of teacher improvement or accountability at either the individual level or the organizational level (Wise et al., 1985). Thirty years later, the two schools of thought regarding the purposes and goals of teacher evaluation remain the same. One is as a means of support and improvement for teachers (Kraft & Gilmour, 2015) and the other as a means of accountability in terms of rating teachers and dismissing ineffective teachers (Hanushek & Rivkin, 2010). Steinberg and Donaldson (2016) put it best when they write:

Most new teacher evaluation systems incorporate measures of student achievement and observations of classroom instruction to assess teacher performance (NCTQ 2013; Hallgren, James-Burdumy, and Perez-Johnson 2014). The espoused goal of these new evaluation systems is to more closely tie the work of teachers to improvements in student learning (Darling-Hammond, Wise, and Pease 1983; Murphy, Hallinger, and Heck 2013). There are two approaches to satisfying the system's fundamental goal of improvement in student outcomes: (1) developing teachers' skills to improve student performance, and (2) evaluating teacher effectiveness for accountability purposes related to tenure, rewards, and dismissal (p. 341).

However, despite these seemingly clear purposes and goals of teacher evaluations one question situated in the teacher evaluation discussion is can teacher evaluation policies and systems, as currently constructed, achieve the these aforementioned goals? Put differently, can teacher evaluation serve the dual purpose of providing useful feedback for teachers to help them improve their practice, while holding them accountable for their performance? And, can teacher evaluation policies and systems provide policymakers and practitioners better information on what makes a quality teacher?

The findings from this dissertation suggest teacher evaluation policies and systems, as currently constructed, are not well-suited to accomplish these goals in part because principal cognition and external context greatly impact how principals generate teacher evaluation information. The main theoretical contribution of this dissertation is the type of sensemaking in which a principal engages is dependent upon the experience level of that principal. Specifically, principals with high-levels of experience engage in individual sensemaking (a type of sensemaking that occurs in one's head and relies on personal experiences and knowledge to make sense of a situation or task) while principals with low-levels of experience engage in collective sensemaking (a type of sensemaking that occurs among multiple people in an organization or an environment). The results of this analysis suggest principals who engage in individual sensemaking make sense of evaluating teachers differently than principals who engage in collective sensemaking. For example, principals who engage in individual sensemaking rely primarily on their own definitions of good classroom instruction where principals who engage in collective sensemaking co-construct what good classroom instruction looks like with their informal networks within their school. These findings suggest that teachers will receive evaluations that look quite different simply based on the experience level of the

principal performing the evaluation. Ultimately, these findings suggest the information generated by teacher evaluations will vary by the experience level of the principal, which might make it difficult for policymakers to decipher what information truly shows quality teaching.

Distinguishing between principals who engage in individual versus collective sensemaking is one way this dissertation moves past the assumption that “sensemaking happens”. This dissertation suggests sensemaking happens differently within principals with certain experience levels, which impacts teacher evaluation policy and system implementation. In short, principals with high-levels of experience evaluate teachers differently than their less experienced peers, which brings into question the consistency of the teacher evaluation information generated by school principals. However, this finding provides potentially significant information to policymakers. For example, if high-experience principals consistently generate high-quality teacher evaluation information, policymakers may be able to design a teacher evaluation system that uses principals with high-experience to conduct all teacher evaluations in a district or state and remove low-experience principals from this process.

One of the practical contributions of this dissertation is that, among my participants, principals with low-levels of experience navigate the process of teacher evaluations with different mindsets and priorities than their more experienced peers. For example, principals with low-levels of experience typically find it difficult to critique teacher performance. Instead these principals prioritize cultivating positive relationships with their staff. Principals with low-levels of experience use teacher evaluation systems to achieve the goal of providing teachers feedback and support to help teachers improve their practice, but in most cases they avoid using these systems to hold teachers accountable for their performance (except in extreme cases). Therefore, the information generated by these principals has the potential to be quite different than the

information generated by their more veteran peers and looking at teacher evaluation ratings across principals with varying experience levels might make it difficult for policymakers, researchers, and practitioners to determine the accuracy of this information.

The different backgrounds, knowledge, experiences, and contexts of principal evaluators raises questions about the capability of current teacher evaluation systems accomplishing the goals of providing teachers information to help improve their classroom performance and hold them accountable for their performance as educators, while also providing policymakers better information on what makes a quality teacher. For example, this study's findings build on previous research which suggests principals in high-pressure environments are more likely to use teacher evaluation policies as a way to rank teachers and as a tool to determine who is effective and who is ineffective, while principals in low-pressure environments use the same policies as improvement tools for their teaching staffs (Chingos & West, p. 428; Fuller & Ladd, 2007). Although principals may be working with the best of intentions to both accurately critique teacher performance and provide teachers with actionable feedback to help them improve their practice, principals' cognition often prioritizes one of these goals over the other. For example, Mr. Jarmel provided his teachers with feedback to improve their practice, but in his mind, his school's teacher evaluation system was a way to show which of his teachers did not reflect effective or highly effective teachers. Because the principals in this study worked in very different contexts in terms of the amount of outside pressure they felt from the state of Michigan, as well as their district-level superiors (and even amongst the teachers in their building), how these principals evaluated their teachers looked quite different from school to school. The amount of outside pressure facing schools and principals is one reason why relying on principals to generate better information on teacher quality is a challenge.

Another challenge of realizing the goals commonly associated with current teacher evaluation policies is there is no, or a very limited, consensus on what makes a teacher effective (Donaldson & Papay, 2014; Fenstermacher & Richardson, 2005). Supporting this research, the principals in this study had varying definitions of teacher effectiveness. For example, Mr. Jarmel believed a teacher's effectiveness was measured by high student achievement on state assessments. In contrast, Mr. Bania did not consider state assessments at all when evaluating teacher performance and instead relied on what he saw in the classroom during observations of teacher instruction. Still another principal, Mr. Bookman defined teacher effectiveness largely based on teacher-student interactions and the relationships teachers built with students and parents. What an effective teacher looks like and means to one principal may differ from others, even within the same school district. Therefore, how principals evaluate teachers and what they prioritize during evaluations will look quite different. Because of this lack of consistent definition of teacher quality, it is often difficult for teacher evaluation policies and systems to provide quality information on what characteristics make a quality teacher. The implication of a lack of consensus on a definition of teacher quality is that teachers receive vastly different evaluation scores, based on the cognition of the evaluator. This is not inherently bad, but given the enormous stakes attached to evaluation scores in terms of teacher employment, this may appear unfair to individual teachers and may lead to unintended consequences, such as teachers leaving schools where principals do not score them favorably. In short, the results of this work suggest teacher evaluation policies and systems, as currently constructed, will likely continue to fall short in providing policymakers, researchers, practitioners better information on what makes a quality teacher. Instead, this research suggests the information from evaluations will provide these individuals information on what principals with specific amounts of experience and who

worked in specific contexts think constitutes quality teaching. I address this further in the following implications section.

Principals' Role in Teacher Evaluations

Given that most policymakers, practitioners, and researchers agree on the goals and purposes of teacher evaluations, as difficult as these goals may be to accomplish, the next logical question is, is it reasonable for principals to be the primary people charged with achieving these goals? In almost all cases, school principals are the primary school-based actors tasked with enacting teachers evaluations systems and assigning these important teacher evaluation ratings (this was the case for the 12 participants in this study) (Steinberg & Donaldson, 2016). As a result, how these individuals make sense of implementing these policies will affect how teacher evaluations look in practice. Additionally, school principals' sensemaking will affect the data produced from these evaluations. The principals in this study were tasked with making sense of external demands while balancing the needs of their specific school and the teachers within the school (Ganon-Shilon & Schechter, 2016). Moreover, the principals in this study "make key decisions that determine which reform demands they bring in, which demands they emphasize with the staff, and which they filter out" (Ganon-Shilon & Schechter, 2016, p. 7), a finding that supports other research that examines how principals make sense of policies that enter their systems of practice. Given the widespread research that shows teacher quality is a significant factor that leads to the aforementioned desirable student outcomes and other, longer-term positive outcomes, such as increased labor market opportunities, and increased employment wages (Hanshuek, 2010; Rockoff 2004), the role of principals in identifying quality teachers and helping teachers improve their craft as educators is arguably a principal's most important responsibility. Although principals certainly can impact student outcomes in a variety of ways

(e.g. fostering a positive working environment, establishing strong communication within a building, supporting parental and community engagement, etc.), one of the most direct ways principals can positively impact student school experiences and outcomes is by identifying, hiring, and retaining high-quality teachers (Boyd et al., 2011; Harris, 2010; Ladd, 2011; Leithwood et al., 2008).

However, principals are responsible for a host of other things outside of the realm of teacher evaluations. For example, principals must manage their building, serve as the instructional leader of the school, and communicate with district administration, parents/guardians, the local community, and various other stakeholders. Additionally, principals are expected to take on the dual role of coach and evaluator, providing support to their teaching staff, while being competent evaluators of classroom instruction and student learning, in a multitude of subject areas and grade levels. Principals also must manage the school budget, bus schedules, design and deliver professional development for staff, and deal with issues of student discipline, absences, and safety. Given all that is asked of school principals, is it reasonable for policymakers to expect principals to be able to accomplish the goals associated with teacher evaluations in addition to their myriad of other tasks? Since the RTTT initiative in 2009 changes to teacher evaluation systems have occurred at unprecedented rates. Donaldson and Papay (2014) note, “teacher evaluation is a prime policy lever as a conduit to combine accountability and support,” but are principals the individuals best suited to accomplish these goals? Principals are charged with understanding these changes and policymakers rely on principals for successful implementation of these policies. New teacher evaluation systems task principals with rating teachers accurately and differentiating amongst teacher effectiveness, while also supporting teacher instructional improvement. Teacher evaluation systems are time consuming to implement

and this implementation must be done delicately given the enormous stakes attached to current teacher evaluations policies. Put another way, is it reasonable to expect principals are able to fairly critique teacher performance and at the same time provide teachers the support and feedback to help them improve their craft and provide districts, states, and policymakers more accurate information on teacher quality and effectiveness?

If the answer to this question is yes, principals must receive increased and more targeted training and professional development when using these complex systems. For example, increasing evidence suggests ongoing conferences between principals and teachers are crucial to the overall evaluation process because these conferences provide opportunities for teachers to improve their practice and ultimately student achievement (Steinberg & Donaldson, 2016; Steinberg & Sartain, 2015; Taylor & Tyler, 2012). Therefore, principals should receive constant support as to how to structure these conferences, what to include during conversations in these conferences, and how to deliver useful feedback to their teachers. As is suggested in this study, these conferences varied drastically from school to school and in some cases did not occur at all. Principals will likely continue to play an active role in negotiating federal, state, and local policies and initiatives (Ganon-Shilon & Schechter, 2016; Koyama, 2014), but if policymakers could ensure at minimum the essential parts of teacher evaluation policies, in this case conversations around instruction, were consisted between schools, some of this lack of continuity may be abated.

Alternatively, as principals continue to be held increasingly accountable for student performance and the performance of their school, giving principals greater discretion over how they evaluate, hire, and work with their staff is something policymakers and district leaders should consider. If principals are the primary people charged with successfully running a school,

they potentially should have a larger say in how these people are evaluated and if the teachers are valuable assets to their school. Giving principals greater autonomy over how teachers are evaluated may best support the needs of local schools. Some research shows that principals, at least in part, are able to make strong evaluative and human capital decisions if given the right information (Jacob & Lefgren, 2008; Rockoff & Speroni, 2011). Additionally, this work suggests that principals' human capital decisions often correlate with other positive results, such as increased parental and student satisfaction (Jacob & Lefgren, 2005; Rockoff & Speroni, 2011). Therefore, if the goal of teacher evaluation policies is to provide better information on what makes a quality teacher and to provide all students with the best teachers, allowing principals the opportunity to decide what works best for their local context may be a unique approach to teacher evaluations, especially considering in many cases it appears principals are evaluating teachers in various ways already.

If principals are not the best people suited to evaluate teachers then policymakers should consider the use of outside evaluators. Often times, principals already have their minds made up about how they will evaluate a teacher, even before the process begins. Weick (1995) calls this a "decision premise" where an individual, early on in the process of making a judgement, assigns values, beliefs, and meanings to what he or she will be judging (p. 115). In that way, when it comes to the final judgement, these individuals will be able to make sense of what they are seeing. Evaluators having a predetermined mindset about who or what they will evaluate is concerning as Weick (1995) writes, "As facts give way to values, computation gives way to judgement, and sensation is displaced by ideology, all without the member necessarily being any wiser to these shifts" (p. 115).

One of the main findings about all principals in this study, but particularly those with

low-levels of experience, is it is difficult for these individuals to separate all that they know about teachers from the official teacher evaluation. The principals in this study constantly referenced the idea that a teacher evaluation was a snap-shot in time and did not encapsulate all that teachers did during the school year. Additionally, principals noted that if they knew that an observation of a teacher, or a teacher's final student assessment data, did not reflect what the principal believed was the teacher's true impact on student learning, their teacher evaluation system had wiggle room to evaluate them accordingly, which typically meant rating teachers more favorably.

These findings lend some credence to the research that suggests using multiple observers, or observers who know little to nothing about the teachers they are evaluating may provide more reliable assessments of teacher instructional ability (Kane et al., 2013; Donaldson & Papay, 2014). Research suggests because principals have intimate relationships with the teachers they are charged with evaluating, it is virtually impossible for principals to evaluate teachers objectively. Using outside observers or observing teachers using multiple administrators, possibly the school principal and another individual from the district office, has the potential to alleviate this concern. The principals in this study certainly referenced the relational aspect of evaluating teachers as a challenge and therefore considering the use of evaluators that do not have close relationships with teachers is something policymakers and district leaders can and should consider when designing future teacher evaluation policies.

The principals in this study used their own thinking and beliefs to evaluate teachers and based their justification for this choice on the high-stakes nature of teacher evaluation policies. When analyzed through the lens of sensemaking theory, these findings suggest that the ways a principal values or perceives the purposes of teacher evaluations, and the relationships he or she

has with staff, shape how he or she interprets and ultimately implements teacher evaluation policies. One example of principals using their own thinking and beliefs to evaluate teachers relates to teacher evaluation scores being used for human capital decision. Because some of the principals in this study knew their school district was using teacher evaluations for human capital decisions, some principals were less likely to rate teachers critically. In other words, the principals in this study interpreted and implemented teacher evaluations while always thinking of the future employment of their teaching staff. Recent work from Grissom and Loeb (2016) produced similar findings in which principals were more likely to rate teachers higher on high-stakes evaluations than on low-stakes evaluations.

In short, given all that is expected of school principals, I argue that if policymakers want better information on what makes a quality teacher, future teacher evaluation policies should allow principals much greater input and freedom when evaluating the teachers in their building or remove principals from the evaluation process. The suggestion to remove principals from the evaluation process entirely and use outside evaluators is not without limitations. For example, outside evaluators will bring their own cognition to the evaluation process. These individuals will have set expectations and beliefs on what makes a quality teacher and will have a predetermined mindset on what quality teaching and instruction looks like. However, the use of outside evaluators does eliminate the relational aspect of teacher evaluations, which has consistently surfaced as a concern or factor for principals while evaluating the teachers in their building. The first suggestion (and my personal preference) gives principals more say in who teaches in their building and gives principals the power to decide what makes a quality teacher for their specific context. Giving principals greater autonomy regarding how they evaluate teachers is not without flaws, as surely some teachers and policymakers may object to evaluations

that are outright subjective and may look different from teacher to teacher. However, if we operate under the assumption that all principals want what is best for their school and students, allowing principals greater discretion on what constitutes a quality teacher in their specific context may help schools cultivate stronger teaching staffs. Therefore, I suggest principals should have greater professional judgement and say as to how teachers are evaluated in their local context.

Implications: For Policymakers

The results of this dissertation's analysis have implications for both policymakers and practitioners. First, policymakers indicate that a primary reason teacher evaluation policies continue to change is the need to design policies that provide better information and what makes a quality teacher, as well as hold teachers accountable for their performance in the classroom. However, an analysis of the findings of this dissertation suggests principal cognition greatly impacts the consistency and transferability of this information, putting into question how useful this information is for policymakers. This is not to say that the information collected by principals during evaluations is not valuable. Observation information collected by principals may in fact be very valuable, particularly when evaluating a teacher at a specific school in a certain context. However, this type of data collection makes it difficult to make between school teacher comparisons, even between schools in the same district. For example, some principals in this work noted that their teacher evaluation systems had room to adjust final scores if they felt such adjustments were called for (e.g. if a principal felt the outcomes of teacher observations and/or students' final assessment data did not reflect teachers' true impact on student learning). If this approach is used consistently by one principal for all of the teachers in a school, this information may be useful when attempting to determine what type of teacher is most effective

in that specific context. However, if some principals do this within a district and others do not, policymakers cannot rely on this information to decide which teachers are in fact most effective.

Other principals in this study considered factors outside of, but related to, teacher evaluation policies (e.g. future teacher employment) when evaluating teachers. For example, during an official observation, Mr. Bookman and I observed a lesson by a teacher that Mr. Bookman told me did not reflect the effectiveness of this teacher. Mr. Bookman said because this lesson was just one bad 45-minute snapshot, he would not penalize this teacher, even though this observation was technically the official observation used for evaluation purposes. Mr. Bookman ultimately rated this teacher highly-effective, even though he admitted the lesson we observed rated more as minimally effective. This implication being, the information provided by Mr. Bookman on this teacher's performance was not an accurate depiction of what we observed. The observational assessment provided by Mr. Bookman may in fact be a fair representation of this teacher's quality (Mr. Bookman has observed this teacher throughout the school year and said these observations were very high-quality), but if the observation conducted by someone other than who knew this teacher so intimately, the evaluation of this teacher would have looked quite different. The implication here for policymakers being the information generated by teacher evaluation policies is largely dependent upon who does the evaluating. Individuals may suggest using outside observers or multiple observers who do not know the teachers as intimately has the potential to alleviate the concern of subjective evaluations of teacher performance. However, the use of outside evaluators does not remove the ethical question surrounding the ways teacher evaluations will be used, which may be a concern for any evaluator. For example, if an outside evaluator knows the results of a teacher's evaluation will be used for employment decisions by the district and this does not align with the evaluators' personal beliefs, outside evaluators may

still rate teachers favorably. Additionally, it is important to note that outside evaluators, like principals, will bring in their own cognition when evaluating teachers, which has the potential to cause similar disruptions to policy implementation efforts. Therefore, I believe if policymakers truly want accurate information on what makes a quality teacher, policies should be designed that allow for increased professional judgement of individual principals. This approach has the potential to produce more useful information on what makes a quality teacher in certain specific contexts. For example, principals who work in high-pressure environments may evaluate teachers in specific ways, by looking for specific characteristics and teaching skills.

Policymakers will be able to use the information generated by principals who work in these contexts to better predict the type of teacher that will be successful (and remain teaching) in these high-pressure environments. In short, I argue principals should have more professional judgement when assigning teacher evaluation ratings, particularly given how intimately principals know their own school, teachers, and students. Research (including this dissertation) suggests principals do this already anyway and perhaps if policymakers and district leaders provided an opportunity for principals to evaluate teachers based on what principals believe is in the best interest of their local school, there might be a better match between teachers and schools, ultimately increasing the quality and length of tenure of teachers in schools. The implication here is that policymakers must understand that principals with different levels of experience and who work in different contexts need different types of teachers in their building and what constitutes and “effective” teacher in one context may not constitute an “effective” teacher in another context.

Implications: For Practitioners

Like policymakers, practitioners (e.g. school and district leaders) hope new teacher evaluation systems will provide better information on what makes a quality teacher, as well as hold teachers accountable for their classroom performance. Additionally, practitioners hope these new policies and systems create opportunities for principals to provide support and feedback to teachers to help teachers improve their classroom practice, resulting in increased student achievement. Given these stated goals, districts and school systems would be well-served to provide principals more structured and intensive training with how to best use these new evaluation systems. The initial training of new teacher evaluation systems is a crucial element to how principals come to understand and implement these systems. While some states and districts have increased the amount and quality of training principals receive on how to implement evolving teacher evaluation systems, the training of the principals in this study varied drastically and in some cases the principals in this work received no training. For example, Ms. Steinman said:

This year I really—I had one day of training as PD that was actually where they were training teachers, not administrators. I went to the teacher training, so at least I got a little bit of feel for that. The rest of my training really has been on the go, reading by myself, researching online, and then working with my administrative team for consistency. It's been a limited training.

The training received by the principals in this study varied in length and quality, suggesting principals everywhere do not always receive adequate teacher evaluation training and support. Principals nationwide would likely benefit from more in-depth and detailed initial and ongoing teacher evaluation training when adopting new teacher evaluation systems. If more

consistent and ongoing training is provided to principals these individuals may be more likely to use these systems in the ways envisioned by policymakers and district leaders. Policies may still be adapted to local contexts, but strengthening the initial training and support for principals has the potential to provide a more aligned vision between policymakers' intentions and practitioners' implementation efforts. Additionally, this training may help principals feel more confident in the accuracy and fairness of these systems. An important idea of these new evaluation systems was to improve student learning by identifying teachers with strong instructional practices and providing constructive feedback in areas where teachers needed improvement. However, the principals in this study often felt their teacher evaluation policy and system was not a good tool for evaluating teacher performance. Districts and the state would be well-served to provide training and explicit rationale to all principals using these new systems about how these systems will help teachers improve their practice and ultimately benefit the students in their school.

In addition to providing principals with initial support and training on how to best use teacher evaluation systems, districts should provide principals feedback on how they are evaluating teachers, citing specifics about their evaluative process, not that they are just in compliance and completing the required paperwork. As principals become more comfortable implementing these new systems and receive constructive feedback, perhaps they will be more willing to critique teachers' practice and provide a more accurate picture of which teachers are most effective. This feedback should include how principals are observing teachers, how principals communicate and deliver teacher evaluation information to teachers, and the overall process of evaluating teachers. Providing principals with increased and ongoing feedback on how they are implementing teacher evaluation policies and systems has the potential to move

principals towards uniformity across districts and states and reduce the subjectivity of teacher evaluations. Currently, in many districts principals are forced to make sense of how to best use teacher evaluation systems on the fly and with little support and rarely receive feedback on their performance as an evaluator (beyond, “you completed the evaluation”).

Additionally, school districts and states would be well-served to consult principals when creating and implementing new teacher evaluation systems. Principals are perhaps the best school-based actors who can most accurately speak to what should be included in a teacher’s evaluation and how to best navigate this process. Principals can work together to create meaningful teacher evaluation systems that still can allow for some professional judgement of principals on teacher performance. Principal involvement in creating teacher evaluation policies has the potential to alleviate some of the concerns of lack of policy implementation as principals will have greater buy-in as they are in large part responsible for designing these policies.

Finally, aspiring school leaders need to be trained in how to identify teacher quality, use data to make important decisions, and evaluate teacher performance, which begins in their principal preparation program. Principal preparation program directors should focus much of their attention on principal evaluation of teachers, as this is arguably one of the most, if not the most, important aspect of a principal’s job. Providing current and future school principals with a clear understanding of teacher quality (at least, as much as practitioners, scholars, and policymakers know about what makes a quality teacher), or at least the potential to identify quality teachers, has the potential to lead to a collection of better information on teacher quality at the school level.

If school districts decide school principals are not the best people suited to effectively implement teacher evaluation policies and evaluate teachers, districts should consider the use of

outside district evaluators, which is happening in some districts throughout the country (Kane et al., 2013). Using outside observers or multiple observers has the potential to remove the relational aspect of performance evaluations, which remains a concern when trying to cultivate a fair and objective teacher evaluation system. However, individuals charged with evaluation teachers will still bring with them their personal cognition, experiences, beliefs, and lens when evaluating teacher performance. Although some may argue the district can better control training, implementation, and feedback if they use this approach, I would argue in practice the use of outside evaluators will still be largely subjective based on the background and knowledge of the evaluator. Therefore, practitioners should proceed with caution before investing heavily in the use of outside evaluators. A more cautious approach is for researchers to conduct more randomized control trials to compare the reliability of principals' ratings of teachers compared the outside evaluator ratings of teachers.

Limitations and Future Research

There are two main limitations to this dissertation. First, the 12 principals that participated in this work influence the findings. If another 12 principals participated in this study, the results may look different. The small number of participants, who were not selected randomly, does not allow me to make generalizable statements about all principals with similar characteristics. However, the goal of this work was to begin to hypothesize about how principals with certain characteristics think about and enact teacher evaluation policies. Therefore, these findings begin the process of building information that can test this hypothesis. The second limitation is, although principals were observed in their natural environment implementing their teacher evaluation policy and system, I did not observe each principal multiple times, with a variety of teachers, or during every interaction the principal had attempting to implement the

policy. In this way, my presence as a researcher may not have captured exactly how principals were conducting teacher evaluations in all circumstances. However, to account for this limitation I spoke with teachers when available to see if what I observed was an accurate or consistent representation of how these principals navigated the process of teacher evaluations. Additionally, I reviewed documents completed when I was not present, to compare and contrast what principals did while I was present and while I was absent.

The findings from this research answered my three research questions and future research can again examine these research questions by conducting the same type of study with different principals who have the same characteristics. This approach is one that I will take when embarking on future research. Additionally, some of the findings of this work can best researched and tested quantitatively. For example, in Michigan data on how specific principals rate teachers are available. Therefore, future work could test the finding of if principals in high-pressure environments (as defined by this study) do in fact distribute their teacher evaluation ratings more so than their peers who work in low-pressure environments. Additionally, researchers can quantitatively examine if principals with low-levels of experience do in fact rate teachers less critically than their more experienced peers. Examining the results of this work quantitatively will help either support these findings, or disconfirm some of this work, either way testing these hypotheses. Finally, the principals in this study used a variety of teacher evaluation systems. Future work could examine if individuals using certain teacher evaluation systems, such as the Charlotte Danielson Framework for Effective Teaching, are more likely to produce consistent and reliable ratings of teachers. I hypothesize that individual principal cognition will impact any subjective evaluative system, but future work geared towards supporting this hypothesis would provide much needed information to policymakers and practitioners alike.

Conclusions

The goal of this study was to inform both practitioners (e.g. school district leaders) and policymakers on the importance of understanding how individuals with certain characteristics implement teacher evaluation policies. Practitioners need to know how individuals with certain characteristics make sense of evaluating teachers because if they better understand the individuals who are primarily charged with implementing new policies (in this case, school principals), they can directly address these variations and challenges by providing specific professional development, creating benchmarks and check-ins with principals throughout the implementation process and by holding these individuals accountable for their performance as evaluators. Policymakers need to know how individuals with certain characteristics make sense of evaluating teachers because as a policy is designed and develops, understanding how the people with whom this policy interacts make sense of the policy will help policymakers address some of this variation of sensemaking while drafting future legislation. Put simply, policymakers will be better able to anticipate what challenges may occur when practitioners attempt to implement future policies.

The results of this study, coupled with other emerging work (Donaldson & Papay, 2014; Goldring et al., 2015; Grissom & Loeb, 2016), may be one explanation as to why teacher evaluation policy implementation remains a challenge in Michigan and beyond. This study and other research shows principals' cognitive schemas will likely always have some impact on how policies look in practice (even if there is a consensus on the definition of teacher effectiveness and even if all principals receive the same training). The principals in this study were greatly influenced by their experiences and external context resulting in a wide variation of teacher evaluation policy implementation.

Interesting to note, since the reform of teacher tenure laws in 2011, of the 96,000 K-12 teachers in Michigan, only 19 have been dismissed due to poor evaluation scores (Michigan Department of Education, 2016). Additionally, K-12 teachers in Michigan continue to be rated overwhelming effective or highly effective; 97% of teachers in the state met this criteria (Michigan Department of Education, 2016). According to the findings of this study, we can likely attribute these high teacher evaluation ratings and lack of dismissals due to principals scoring teachers higher than would be expected – not necessarily because all of these teachers are effective or highly effective in the classroom.

How principal cognition impacts their implementation of teacher evaluation systems is a double-edged sword. First, principals were able to be nimble and react to local instructional needs, such as tailoring their evaluation systems to focus on teaching attributes that these principals felt were an important measure of teacher effectiveness for their school context. Additionally, the principals were able to use teacher evaluations as a tool for focusing on larger, local priorities. Some research argues policies should be able to be adapted to meet local needs (McLaughlin & Talbert, 1993) and giving local actors more say in how policies look in practice may in fact be a net positive for promoting teacher, student, and school growth. However, the other edge of the sword is that principals in this study did not always address the goals central to the policy aims of teacher evaluation reform. In Michigan and nationally, steps are currently being taken to better standardize the teacher evaluation process calling for more clarity, accountability, and transparency in teacher evaluation systems (Hill & Grossman, 2013; US Department of Education, 2009). However, the results of this work indicate principals use teacher evaluation systems to work towards local goals and priorities and not necessarily towards the goals envisioned by policymakers. This finding is potentially worrisome in the sense that it

shows how a policy can be co-opted and used for reasons outside of the scope of the design of the policy. Although the principals in this study were acting in good faith and doing what they believed was in the best interest of their school and students, the results indicate there is a mismatch between policymakers' intentions and practitioners' implementation.

This dissertation contributes to our understanding of how principals make sense of and implement teacher evaluation policies by modeling the relationship between principal cognitive schemas and teacher evaluation policy implementation. Because past research shows there is considerable variability between how principals implement teacher evaluation policies (Halverson et al., 2004) a more nuanced understanding of how principals with certain characteristics implement these widely popular policies may help district and education policy leaders better support principals and thus ensure more beneficial implementation. Thus, this dissertation contributes to both theory and practice. Specifically, this dissertation contributes to the sensemaking theory literature by suggesting principals with high-levels of experience engage in individual sensemaking, while principals with low-levels of experience engage in collective sensemaking. As was previously stated, the type of sensemaking in which one engages has implications for how teacher evaluation policies look in practice. For practice, this dissertation provides school districts information on how principals with certain experience characteristics and who work in certain contexts are likely to think about evaluating teachers in their building. Practitioners can use this information to better train principals, as well as provide them support as they navigate the process of teacher evaluation policy implementation and anticipate the challenges of implementation.

This study is significant for two reasons. First, in most cases principals are primarily and solely responsible for conducting and implementing teacher evaluation policies. Understanding

why principals think about these high-stakes policies in certain ways and what types of thinking go into the evaluation process is important. Second, there is little evidence to support even the best designed teacher evaluation system will be implemented as intended. Therefore, it will be useful for both school leaders and policymakers to better understand how principals with certain cognitive schemas think about these systems as to predict how they may play out with certain principals. In this way both policymakers and practitioners will be better able to anticipate challenges of policy implementation and better account for these challenges when designing policies and training principals to use evolving teacher evaluation systems.

This study brings together the bodies of literature on cognitive and sensemaking theory with principal and policy implementation with the goal of generating hypotheses of how principals with certain cognitive schemas are likely to implement teacher evaluation policies. This work builds on and extends the idea that school leader sensemaking is influenced by prior knowledge and preexisting understandings impact how they implement policies, particularly teacher evaluation policies (Coburn, 2005). Additionally, this study may help other states outside of Michigan begin to build theory and begin to identify a predictive model of how principals may implement teacher evaluation policies in their context. Many other states are in the process of a teacher evaluation overhaul and stand to learn from the lessons of Michigan. Given the increasing amount of attention, scrutiny, and changing evaluation policies, states across the country will need to better understand what is happening with these policies as they enter their system of practice. Although it is unlikely everyone will implement the same policy identically in all circumstances, past research shows as currently constructed, teacher evaluations are not doing a good job identifying quality teachers (Weisberg et al., 2009) and part of this is because of a lack of fidelity of policy implementation. In the future, states, researchers, and school

leaders should work together to provide better training and support to those charged with implementing these important policies as well consider allowing increased space for the professional judgement of principals when evaluating teachers in their context .

As new teacher evaluation policies continue to permeate the educational landscape, how these reforms play out in different contexts is of extreme importance and should be studied. Given how important school principals are in policy implementation, it is imperative we better understand their thinking about how and why they implement policies or parts of policies in certain ways. This study shows that outside factors such as experience and context do in fact impact principals' implementation of teacher evaluation policies. As more sanctions, money, and overall importance is tied to these policies, researchers should continue to focus on what the people who are charged with enacting these policies deem important and how this interpretation and implementation process affects policies, schools, and students.

APPENDICES

Appendix A

Principal Questionnaire

Thank you for taking the time to complete this questionnaire. As with any part of this study, you can withdraw your consent to participate at any time and you do not have to answer any questions that you do not want to answer. Anything you say will not be connected with your name, the name of your school, or the name of your school district in any publications or presentations. Your responses to this questionnaire will be kept in a locked filing cabinet or on a secure computer. Your identity will be kept using unique ID numbers and will never be released.

Background Information

What is your age range? (Please check one box)

☐ Younger than 30 ☐ 30-40 ☐ 41-50 ☐ 51-60 ☐ Older than 60

What is the highest level of formal education you have completed? (Please check one box)

☐ Bachelor's Degree ☐ Master's Degree ☐ Doctoral Degree

How many years of experience do you have working as a principal? (Please circle)

1 2 3 4 5 6 7 8 9 10 More than 10

How many years of experience do you have working as a principal at your current school?
(Please circle)

1 2 3 4 5 6 7 8 9 10 More than 10

How many years did you spend as a classroom/subject teacher before you became a principal?
(Please check one box)

☐ None ☐ 1-5 ☐ 6-10 ☐ More than 10

Please respond to the following questions by placing and X in the box that most aligns with your feelings/beliefs.

Scale: (1. SD: Strongly Disagree, 2. D: Disagree, 3. SOD: Somewhat disagree, 4. U: Undecided, 5. SOA: Somewhat Agree, 6. A: Agree, 7. SA: Strongly Agree)

	Strongly Dis-agree	Dis-agree	Some-what Disagree	Un-decide d	Some-what Agree	Agree	Strongly Agree
My experience as an administrator impacts how I implement my district's teacher evaluation policy.							
My beliefs on the goals of education impacts how I implement my district's teacher evaluation policy.							
My leadership style impacts how I implement my district's teacher evaluation policy.							
I am implementing my district's teacher evaluation policy as envisioned by policymakers.							

Other principals in my district are implementing the teacher evaluation policy as envisioned by policymakers.							
My district's teacher evaluation policy is an accurate reflection of teacher quality/effectiveness.							
My relationship with my teaching staff impacts how I implement our district's teacher evaluation policy.							
I look at teachers' previous evaluation data (including observation scores and student test scores) before evaluating a teacher.							
I use available resources to support teachers who are struggling.							

If I need clarification on an aspect of my district's teacher evaluation policy I will seek clarification.							
I communicate with my teaching staff as my district's teacher evaluation policy requires.							
I provide feedback to my teaching staff as my district's teacher evaluation policy requires.							
It is challenging to implement my district's teacher evaluation policy.							
I use my district's teacher evaluation policy to make personnel decisions.							

I think about the future of my school when implementing my district's teacher evaluation policy.							
--	--	--	--	--	--	--	--

Appendix B

Principal Interview Protocol 1

Principal ID:

Date:

Thank you for taking the time to be interviewed. As with any part of this study, you may withdraw your consent to participate at any time and you do not have to answer any questions that you do not want to answer. Anything you say will not be connected with your name, the name of your school, or the name of your school district in any publications or presentations. I will audio-record your responses for my use only. First, I'll ask questions about your knowledge and beliefs about your district's teacher evaluation system. Then, I will ask you about your role in implementing your district's teacher evaluation policy. Finally, I will ask you about your experiences planning for and conducting teacher evaluations. Your responses to this interview will be kept in a locked filing cabinet or on a secure computer. Your identity will be kept using unique ID numbers and will never be released.

STATE PARTICIPANT ID NUMBERS, DATE, NAME OF INTERVIEWER, AND "START INTERVIEW" FOR RECORDING DEVICE

Principals' Current Teacher Evaluation System

1. What teacher evaluation framework does your school use?
2. How was that framework chosen?
3. What are the strengths of your current teacher evaluation framework?
4. What are the weaknesses of your current teacher evaluation framework?
5. What would you change?
6. How do you conduct a teacher evaluation?
 - What does the process look like from start to finish?
 - How do you prepare for them?
 - How do you conduct them?
 - How do you communicate the evaluation to teachers?
7. To what extent do you believe your district's teacher evaluation system observational protocol is a valid measurement of teacher effectiveness?
8. How were you trained in using this instrument?
9. Describe all of the factors included in a teacher's evaluation score.
10. What percentage of a teacher's evaluation is based on student growth data?
11. Has the addition of student test scores in the evaluation process impacted how you conduct teacher evaluation observations?
12. What sources of assessment data are used for determining student growth? (e.g. state standardized tests, teacher made assessments, etc.)
13. What factors most impact your ability to implement teacher evaluation policies?
14. How do new teacher evaluation policies affect principal your relationship with your teaching staff?

Principals' Beliefs about Teacher Evaluation Policy

1. Do current teacher evaluation measures help you identify quality teaching? If so, how? If not, why not?
2. Setting aside raising test scores for the moment, what teacher characteristics do you consider most important when evaluating teacher quality?
3. How do you use teacher evaluation scores to make decisions?
4. In your opinion, what should be included in a teacher's evaluation to make his or her score reliable and valid?
5. What do you think is the best indicator of a teacher's effectiveness?
6. What do you think is the best way to accurately determine a teachers' effectiveness in the classroom?
7. Do you think using student assessment data can improve the quality of teachers in your building? The teacher workforce as a whole?
8. In your opinion what are teaching behaviors that accurately represent a quality teacher?
9. How do teachers generally respond to the evaluation process?
10. Are teacher evaluations are "helpful" or "beneficial" to teachers (e.g. the feedback helps them improve?) If yes, how so? If no, why not?
11. Last year teachers were required to be evaluated two times. Do you think this is a fair number? Should it be more or less? Why?
12. What percentage of a teacher's evaluation score should be related to student assessment data (e.g. growth, etc.)? Why this percentage?
13. Do you think being able to effectively judge teacher effectiveness is an important indicator of how you are doing as a principal?
14. Do you think you should be held responsible for the effectiveness of your teaching staff?
15. Do you think you should be held accountable for student achievement/growth of the students in your building?

Appendix C

Principal Interview Protocol 2

Principal ID:

Date:

Thank you for taking the time to be interviewed. As with any part of this study, you can withdraw your consent to participate at any time and you do not have to answer any questions that you do not want to answer. Anything you say will not be connected with your name, the name of your school, or the name of your school district in any publications or presentations. I will audio-record your responses for my use only. First, I'll ask questions about how you provide feedback to your teachers. Then, I will ask you about how you use teacher evaluations to make decisions. Finally, I will ask you about how you think about teacher evaluations in the big picture. Your responses to this interview will be kept in a locked filing cabinet or on a secure computer. Your identity will be kept using unique ID numbers and will never be released.

Principals' Experience Implementing Teacher Evaluation Policies

1. When you observe a teacher, what are some of the indicators that help you distinguish between an effective lesson and an ineffective lesson?
2. Do you take into account outside factors that the evaluation rubric may not account for (e.g. you know the teacher has a challenging class, or maybe is just having an off day) and if so, what are some of these factors?
3. Do you consider outside factors other than teaching (e.g. a teacher who supports the school in other ways, like coaching, after school tutor, etc.) when evaluating teachers?
4. Does increased teacher effort play into evaluations, either consciously or subconsciously? Is it part of the evaluation process and/or should it be?
5. As teacher evaluation policies have changed, have you noticed a change in classroom behaviors from teachers (e.g. teaching to the test?)
6. Can you reflect on a particularly challenging evaluation? What made it challenging and how was it ultimately resolved?

Changing Teacher Evaluation Policies

1. In 2018-19 40% of a teacher's evaluation score will be based on student growth. How fair is this percentage?
2. In this current format standardized test scores would account for only half of the student growth measurement. The other half would be based on local measures or assessments. How fair is this approach?
3. How will the 2018-19 change of "schools are prohibited from assigning a student to an ineffective teacher in the same subject area for two consecutive years" impact your job and/or how you conduct teacher evaluations?
4. How fair is it for you to be personally held accountable for your performance implementing teacher evaluation policies?

Using Teacher Evaluations

1. How would you describe the purpose of teacher evaluations?
 - What are they used for? (hiring, firing, retention, assigning teachers to specific students/classrooms?)
 - What should they be used for?
2. Is your school in a position to successfully implement the current teacher evaluation policy?
 - If so, how?
 - If not, what is missing?
3. Has this system been a success?
 - How do you know/what is your evidence?
4. Has implementation of these policies changed over time? (e.g. was implementation different year one than it is now?)
 - If so, how?
5. How does the current teacher evaluation system help improve student performance?
6. How does the current teacher evaluation system interact with other policy initiatives?
 - Do they conflict?
 - Do they assist?
7. What are the greatest challenges of teacher evaluation policy implementation?
8. How does your current teacher evaluation system allow you, as the school leader, to impact teacher quality and student achievement?
9. To what extent do you believe your district's student growth measurement component is a valid measure of a teacher's effectiveness?
10. How are teacher evaluation scores used in hiring decisions?
 - Firing decisions?
 - Retention decisions?

Appendix D

Principal Interview Protocol 3

Principal ID:

Date:

Thank you for taking the time to be interviewed. As with any part of this study, you may withdraw your consent to participate at any time and you do not have to answer any questions that you do not want to answer. Anything you say will not be connected with your name, the name of your school, or the name of your school district in any publications or presentations. I will audio-record your responses for my use only. During this interview I will ask you questions about the observation we just completed. Your responses to this interview will be kept in a locked filing cabinet or on a secure computer. Your identity will be kept using unique ID numbers and will never be released.

STATE PARTICIPANT ID NUMBERS, DATE, NAME OF INTERVIEWER, AND “START INTERVIEW” FOR RECORDING DEVICE

1. What are your initial thoughts/reflections on the lesson we observed?
2. Was that observation a standard length?
3. How do you navigate your actions during the observation (i.e. typing notes, interacting with students, etc.)?
4. What were the strengths of that lesson?
5. What were some areas of improvement?
6. How do you approach the process of notetaking?
7. Are you thinking about the specifics of what your teacher evaluation policy asks you to do while you are observing the teacher?
8. Does how you observe a teacher change based on that teacher?
9. Are observations an accurate representation of teacher effectiveness?
10. How do you think about providing feedback to teachers that is meaningful and useful?

Appendix E

Teacher Interview Protocol

Thank you for taking the time to be interviewed. As with any part of this study, you can withdraw your consent to participate at any time and you do not have to answer any questions that you do not want to answer. Anything you say will not be connected with your name, the name of your school, or the name of your school district in any publications or presentations. I will audio-record your responses for my use only. I will ask you some questions about your thoughts and experiences with your school's teacher evaluation system. Your responses to this interview will be kept in a locked filing cabinet or on a secure computer. Your identity will be kept using unique ID numbers and will never be released.

Teacher Knowledge and Beliefs of Current Evaluation Policies

1. What are the strengths of your current teacher evaluation framework?
2. What are the weaknesses of your current teacher evaluation framework?
3. What would you change?
4. How would you describe the purpose of teacher evaluations?
5. How should teacher evaluations be used?
6. How have you been trained with your current evaluation system?
7. What percentage of student assessment data should be used in these evaluations?
8. Do you feel your evaluation is an accurate representation of your teaching?
9. What criteria is the most accurate representation of your teaching?
10. How has the feedback you received from evaluations helped you improve your practice?
11. To what extent do you believe your district's student growth measurement component is a valid measure of a teacher's effectiveness?
12. In your opinion, what should be included in a teacher's evaluation to make their score reliable and valid?
13. In 2018-19 40% of a teacher's evaluation score will be based on student growth. Is this a fair percentage?
14. In this current format standardized test scores would account for only half of the student growth measurement. The other half would be based on local measures or assessments. Is this a fair approach?

Teachers' Perceptions of Principal Policy Implementation

1. How well does your principal understand the current teacher evaluation system?
2. In your opinion, is your school implementing the teacher evaluation system in a way consistent with your understanding of the policy?
3. How do you think your principal thinks they should be used?
4. How often to you use advice/feedback given to you by your principal?
5. Does your principal dominate conversations or do they provide you a chance for ample input and a chance to contribute to the conversation?*How Teacher Practice is Impacted by Evaluations*
6. How has your practice been impacted by changes teacher evaluation policies?

7. How does your practice change when you are being observed for a formal teacher evaluation?
8. How does your practice change knowing student assessment data is used as part of your evaluation score?
9. How does your practice change knowing your evaluation scores will be compared to colleagues (both within your school and district wide)?

Appendix F

Observation Protocol

Date:

Time:

Participant(s):

School:

Observations:	Notes:	Questions/Follow Up:

WORKS CITED

WORKS CITED

- Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, 25(1), 95-135.
- Anagnostopoulos, D., & Rutledge, S. (2007). Making sense of school sanctioning policies in urban high schools. *Teachers College Record*, 109(5), 1261-1302.
- Bartlett, F. C. (1958). *Thinking: An experimental and social study*. London: Allen & Unwin.
- Berg, B. L. (2007). *Qualitative research methods for the social sciences* (6th Ed.). San Francisco, CA: Pearson Education.
- Beteille, T., Kalogrides, D., & Loeb, S. (2009). Effective Schools: Managing the recruitment, development, and retention of high quality teachers. CALDER Working Paper 37. Washington, DC: The Urban Institute.
- Bidwell, C. E. (2001). Analyzing schools as organizations: Long-term permanence and short-term change. *Sociology of Education*, 74, 100-114.
- Bingham, C. B., & Kahl, S. J. (2013). The process of schema emergence: Assimilation, deconstruction, unitization and the plurality of analogies. *Academy of Management Journal*, 56(1), 14-34.
- Blasé, R., Blasé, J., & Phillips, D. Y. (2010). *Handbook of school improvement: How high-performing principals create high-performing schools*. Thousand Oaks, CA: Corwin Press.
- Booher-Jennings, J. (2005). Below the bubble: "Educational Triage" and the Texas accountability system. *American Educational Research Journal*, 42(231), 231-268.
- Branch, G. F., Hanushek, E. A. & Rivkin S. G. (2009). Estimating principal effectiveness. CALDER Working Paper 32. Washington, DC: The Urban Institute.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, (104)9, 2633-2679.
- Cubberley, E. (1929). *Public school administration* (3rd ed.). Boston, MA: Houghton Mifflin.
- Clark, D., Martorell, P. & Rockoff, J. E. (2009). School principals and school performance. CALDER Working paper 38. Washington, DC: The Urban Institute.

- Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *The Journal of Human Resources*, 41(4), 778-820.
- Cohen, D. K., & Barnes, C. A. (1993). Pedagogy and policy. In D. Cohen, M. W. McLaughlin, & J. E. Talbert (Eds.), *Teaching for understanding: Challenges for policy and practice*, (pp. 207-239). San Francisco, CA: Jossey-Bass Inc.
- Coburn, C. E. (2001). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis*, 23(2), 145-170.
- Coburn, C. E. (2005). Shaping teacher sensemaking: School leaders and the enactment of reading policy. *Educational Policy*, 19(3), 476-509.
- Cohen, D. K., & Hill, H. (2001). *Learning policy: When state education reform works*. New Haven, CT: Yale University Press.
- Cohen-Vogel, L. (2011). Staffing to the test: Are today's school personnel practices evidence based? *Educational Evaluation and Policy Analysis*, 33(4), 483-505.
- Creswell, J. (2013). *Qualitative inquiry and research design*. Los Angeles, CA: Sage.
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Educational Policy Analysis Archives*, 8(1), 1-44.
- 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.
- Dee, T., Jacob, B. A., & Schwartz, N. (2013). The effects of NCLB on school resources and practices. *Educational Evaluation and Policy Analysis*, 35(2), 252-279.
- Denzin, N. K., & Lincoln Y. S. (2003). *Collecting and interpreting qualitative materials* (2nd ed.) Thousand Oaks, CA: Sage Publications.
- Derrington, M. L., & Campbell, J. W. (2013). The changing conditions of instructional leadership: Principals' perceptions of teacher evaluation accountability measures. In B. Barnett, A. R. Shoho, & A. J. Bowers (Eds.), *School and district leadership in an era of accountability* (pp. 231-251). Charlotte, NC: Information Age Publishing.
- Dewey, J. (1938). *Experience and education*. New York, NY: The Macmillan Company.
- Diamond, J. & Spillane, J.P. (2004). High-stakes accountability in urban elementary schools: Challenging or reproducing inequality? *Teachers College Record*, 106(6), 1145-1176.

- Donaldson, M. L., & Papay, J. (2014). Teacher evaluation for accountability and development. In H. F. Ladd and M. E. Goertz (Eds.), *Handbook of research in education finance and policy*, (pp. 174–193). New York, NY: Routledge.
- Donaldson, M. L., & Papay, J. P. (2014). Teacher evaluation reform: Policy lessons from school principals. *Principal's Research Review*, 9(5), 1-8.
- Donaldson, M. L. (2013, April). How do teachers respond to being evaluated based on their students' achievement? Evidence from New Haven, CT. Paper presented at the annual conference of the American Educational Research Association, San Francisco, CA.
- Donaldson, M. L. (2009). So long, Lake Wobegon? Using teacher evaluation to raise teacher quality. Retrieved from https://www.americanprogress.org/wp-content/uploads/issues/2009/06/pdf/teacher_evaluation.pdf
- Duke, D. L., & Stiggins, R. J. (1990). Beyond minimum competence: Evaluation for professional development. In J. Millman & L. Darling-Hammond (Eds.), *The New Handbook of Teacher Evaluation: Assessing Elementary and Secondary School Teachers* (pp. 116-132). Newbury Park, CA: Corwin Press.
- Duke, D. L., & Stiggins, R. J. (1986). *Teacher evaluation: Five keys to growth*. Washington, D.C.: National Educational Association.
- Elmore, R. F. (1980). Complexity and control: What legislators and administrators can do about implementing public policy. In L. Shulman & G. Sykes (Eds.), *Handbook of teaching and policy* (pp. 342-369). New York, NY: Longman.
- Ganon-Shilon, S., & Schechter, C. (2016). Making sense of school leaders' sense-making. *Educational Management Administration & Leadership*, Published online before print.
- Gates, P. E., Blanchard, K. H., & Hersey, P. (1976). Diagnosing education leadership problems: A situational approach. *Educational Leadership*, 33(5), 348-354.
- Goldring, E., Grissom, J. A., Ruben, M., Neumerski, C. M., Cannata, M., Drake, T., & Schuermann, P. (2015). Make room value added: Principals' human capital decisions and the emergence of teacher observation data. *Educational Researcher*, 44(2), 96-104.
- Greeno, J. G. (1998). Where is teaching? *Issues in Education*, 4(1), 110–119.
- Grider, C. (1993). Foundations of cognitive theory: A concise review. Available at: <http://files.eric.ed.gov/fulltext/ED372324.pdf> (accessed 15 November, 2015).
- Grint, K. (2011). A history of leadership. In A. Bryman, D. Collinson, K. Grint, B. Jackson & M. Uhl-Bien (Eds.), *The SAGE handbook of leadership* (pp. 3-14). Thousand Oaks, CA: Sage.

- Grissom, J. A. (2011). Can good principals keep teachers in disadvantaged schools? Linking principal effectiveness to teacher satisfaction and turnover in hard-to-staff environments. *Teachers College Record*, 113(11), 2552-2585.
- Grissom, J. A., & Loeb, S. (forthcoming). Assessing principals' assessments: Subjective evaluations of teacher effectiveness in low- and high-stakes environments. *Education Finance and Policy*.
- Grissom, J. A. & Loeb, S. (2009). Triangulating principal effectiveness: How perspectives of parents, teachers, and assistant principals identify the central importance of managerial skills. CALDER Working Paper 35. Washington, DC: The Urban Institute.
- Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The handbook of qualitative research* (pp. 105-117). Thousand Oaks, CA: Sage Publications.
- Fenstermacher, G. D. & Richardson, V. (2005). On making determinations of teacher quality. *Teachers College Record*, 107(1), 186-213.
- Figlio, D. N., & Winicki, J. (2005). Food for thought: The effects of school accountability plans on school nutrition. *Journal of Public Economics*, 89(2-3), 381-394.
- Firestone, W., Monfils, L., Schorr, R., Hicks, J., & Martinez, M. C. (2004). Pressure and support. In W. Firestone, L. Monfils & R. Schoor (Eds.), *The ambiguities of teaching to the test. Standards, assessment and educational reform*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Fiss, P. C., & Zajac, E. J. (2006). The symbolic management of strategic change: Sensegiving via framing and decoupling. *The Academy of Management Journal*, 49(6), 1173-1193.
- Hallinger, P., & Heck, R. (1996). Reassessing the principal's role in school effectiveness: A review of empirical research, 1980-1995. *Educational Administration Quarterly*, 32(1), 5-44.
- Halverson, R., Kelley, C., & Kimball, S. (2004). Implementing teacher evaluation systems: How principals make sense of complex artifacts to shape local instructional practice. In C. Miskel & W. Hoy (Eds.), *Theory and research in educational administration* (pp. 66-90). Greenwich, CT: Information Age Press.
- Halverson, R., & Clifford, M. (2006). Evaluation in the wild: A distributed cognitive perspective on teacher assessment. *Educational Administration Quarterly*, 42(4), 578-619.
- Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review*, 100(2), 267-271.
- 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.
- Hess, F. M. (2008). Looking for leadership: Assessing the case of mayoral control of urban school systems. *American Journal of Education*, 114(3), 219-245.
- Hill, H. C. & Barth, M. (2004). NCLB and teacher retention: Who will turn out the lights? *Education and the Law*, 16(2-3), 173-181.
- Honig, M. I. (2006). Complexity and policy implementation: Challenges and opportunities for the field. In M. I. Honig (Ed.), *New directions in education policy implementation: Confronting complexity* (pp. 1-24). Albany, NY: State University of New York Press.
- Honig, M. I., & Hatch, T. C. (2004). Crafting coherence: How schools strategically manage multiple, external demands. *Educational Researcher*, 33(8), 16-30.
- Jacob, B. A. (2011). Do principals fire the worst teachers? *Educational Evaluation and Policy Analysis*, 33(4), 403-434.
- Jacob, B. A., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, 26(1), 101-136.
- James, W. (1890). *The principles of psychology*. New York, NY: Henry Holt & Company.
- Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). Have we identified effective teachers? Validating measures of effective teaching using random assignment. The Bill and Melinda Gates Foundation: Seattle, WA.
- Keesler, V. A., & Howe, C. (2015). Teacher evaluation in Michigan. In J. A. Grissom & P. Youngs (Eds.), *Improving teacher evaluation systems* (pp. 156-168). New York, NY: Teachers College Press.
- Kennedy, M. (2010). Attribution error and the quest for teacher quality. *Educational Researcher*, 39(8), 591-598.
- Kimball, S. M. (2003). Analysis of feedback, enabling conditions and fairness perceptions of teachers in three school districts with new standards-based evaluation systems. *Journal of Personnel Evaluation in Education*, 16(4), 241-269.
- Klein, G., Moon, B., & Hoffman, R. R. (2006). Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent Systems*, 88-92.
- Klein, G., Moon, B., & Hoffman, R. R. (2006). Making sense of sensemaking 1: Alternative perspectives. *IEEE Intelligent Systems*, pp. 22-26.

- Koyama, J. (2014). Principals as bricoleurs: Making sense and making do in an era of accountability. *Educational Administration Quarterly*, 50(2), 279-304.
- Kraft M. A., & Gilmour A. F. Revisiting the widget effect: Teacher evaluation reforms and the distribution of teacher effectiveness. Working Paper.
- Kraft, M. A. & Gilmour, A. F. (2015). Can principals promote teacher development as evaluators? A 21 case study of principals' views and experiences. Brown University Working Paper.
- Leithwood, K., Harris, A., & Hopkins, D. (2008). Seven strong claims about successful school leadership. *School Leadership & Management* 28(1), 27-42.
- Leithwood, K., Seashore-Louis, K. Anderson, S., & Wahlstrom, K. (2004). How leadership influences student learning. New York: The Wallace Foundation. Retrieved from <http://www.wallacefoundation.org/knowledge-center/Pages/How-Leadership-Influences-Student-Learning.aspx>
- Lipsky, M. (1980). *Street-level bureaucracy: Dilemmas of the individual in public services*. New York, NY: Russell Sage Foundation.
- Maitlis, S., & Christianson, M. (2014). Sensemaking in organizations: Taking stock and moving forward. *The Academy of Management Annals*, 8(1), 57-125.
- Marshall, C., & Rossman, G. (1999). *Designing qualitative research* (3rd Ed.). Thousand Oaks, CA: Sage Publications.
- Matsumura, L. C., & Wang, E. (2014). Principals' sensemaking of coaching for ambitious reading instruction in a high-stakes accountability policy environment. *Educational Policy and Analysis Archives*, 22(51), 1-37.
- Maxwell, J. A. (2005). *Qualitative research design: An interactive approach* (2nd Ed.). Thousand Oaks, CA: Sage Publications.
- McCleskey, J. A. (2014). Situational, transformational, and transactional leadership and leadership development. *Journal of Business Studies Quarterly*, 5(4), 117-130.
- McLaughlin, M. W. (1987). Learning from experience: Lessons from policy implementation. *Educational Evaluation and Policy Analysis*, 9, 171-178.
- McLaughlin, M. W., & Talbert, J. E. (2001). *Professional communities and the work of high school teaching*. Chicago, IL: University of Chicago Press.
- MET Project. (2013). *Ensuring Fair and Reliable Measures of Effective Teaching*. Bill and Melinda Gates Foundation.

- Michigan Department of Education. (2015). Educator evaluations. Retrieved from: http://www.michigan.gov/mde/0,4615,7-140-5683_75438---,00.html
- Milanowski, A. T., & Heneman, H. G., III. (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., & Saldana, J. (2014). *Qualitative data analysis: A methods sourcebook (3rd Ed.)*. Thousand Oaks, CA: Sage Publications.
- Murphy, J. T. (1971). Title I of ESEA. The politics of implementing federal education reform. *Harvard Educational Review*, 41(1), 35-63.
- Nelson, B. S., Sassi, A., & Grant, C. M. (2001, April). The role of educational leaders in instructional reform: Striking a balance between cognitive and organizational perspectives. Paper presented at the American Educational Research Association Conference, Seattle, WA.
- Odden, A. (1991). The evolution of education policy implementation. In A. Odden (Ed.), *Education policy implementation* (pp. 1-12). Albany, NY: State University of New York Press.
- Papay, J. P., & Johnson, S. M. (2012). Is PAR a good investment? Understanding the costs and benefits of teacher peer assistance and review programs. *Educational Policy*, 26(5), 696-729.
- Papay J. P. & Kraft M. A. (2014) Forthcoming: Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics*.
- Patton, M. (2014). *Qualitative research and evaluation methods*. (4th Ed.). St. Paul, MN: Sage Publishing.
- Piaget, J. (1964). Cognitive development in children. *Journal of Research in Science Teaching*, 2(3), 176-186.
- Piaget, J., & Inhelder, B. (1958). *Growth of logical thinking: From childhood to adolescence*. London: Routledge & Kegan Paul, PLC.
- Porter, A. C., Youngs, P., & Odden, A. (2001). Advances in teacher assessment and their uses. In V. Richardson (Ed.), *Handbook of research on teaching*, (pp. 259–297). New York, NY: Macmillan.
- Rigby, J. (2015). Principals' sensemaking and enactment of teacher evaluation. *Journal of Educational Administration*, 53(3), 374-392.

- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review*, 94(2), 247-252.
- Seashore-Louis, K., & Robinson, V. M. (2012). External mandates and instructional leadership: School leaders as mediating agents. *Journal of Educational Administration*, 50(5), 629-655.
- Seashore-Louis, K., Wahlstrom, K. L., Leithwood, K., & Anderson, S. E. (2010). Investigating the links to improved student learning. *The Wallace Foundation*. Retrieved on December 10, 2015, from <http://www.wallacefoundation.org/knowledge-center/school-leadership/key-research/Pages/Investigating-the-Links-to-Improved-Student-Learning.aspx>
- Smylie, M. A. (2010). *Continuous school improvement*. Thousand Oaks, CA: Corwin Press.
- Spillane, J. P. (2006). *Distributed leadership*. San Francisco, CA: Jossey-Bass.
- Spillane, J. P. (2000). Cognition and policy implementation: District policymakers and the reform of mathematics education. *Cognition and Instruction*, 18(2), 141-179.
- Spillane, J. P., Diamond, J. B., Burch, P., Hallett, T., Jita, L., & Zoltners, J. (2002). Managing in the middle: School leaders and the enactment of accountability policy. *Educational Policy*, 16(5), 731-762.
- Spillane, J. P., & Kenney, A. (2012). School administration in a changing education sector: The U.S. experience. *Journal of Educational Administration*, 50(5), 541-561.
- Spillane, J. P., & Lee, L. C. (2014). Novice principals sense of ultimate responsibility: Problems of practice transitioning to the principal's office. *Educational Administration Quarterly*, 50(3), 431-465.
- Spillane, J. P., Reiser, B. J., & Gomez, L. M. (2006). Policy implementation and cognition: The role of human, social, and distributed cognition in framing policy implementation. In M. I. Honig (Ed.), *New directions in education policy implementation: Confronting complexity* (pp. 47-63). Albany, NY: State University of New York Press.
- Spillane, J. P., Halverson, R., & Diamond, J. B. (2004). Towards a theory of leadership practice: A distributed perspective. *Journal of Curriculum Studies*, 36(1), 3-34.
- Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research*, 72(3), 387-431.
- 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., & Sartain, L. (2015). Does teacher evaluation improve school performance? Experimental evidence from Chicago's Excellence in Teaching Project. *Education Finance and Policy*, 10(4), 535–572.
- Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review*, 102(7), 3628-3651.
- The Center for Public Education. (2014). Trends in teacher evaluation. Retrieved from <http://www.centerforpubliceducation.org/Main-Menu/Evaluating-performance/>
- US Department of Education. (2009). Race to the Top program executive summary. Retrieved from <http://www2.ed.gov/programs/racetothetop/executive-summary.pdf>
- Weatherly, R., & Lipsky, M. (1977). Street-level bureaucrats and institutional innovation: Implementing special education reform. *Harvard Educational Review*, 47(2), 171–197.
- Weick, K. E. (1995). *Sensemaking in organizations*. Thousand Oaks, CA: Sage Publications.
- Weick, K., & Sutcliffe K. (2007). *Managing the unexpected: Resilient performance in an age of uncertainty*. San Francisco, CA: Jossey Bass.
- Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. New York, NY: The New Teacher Project.
- 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), 60-121.
- Wundt, W. M. (1902). *Principles of physiology psychology*. New York, NY: The Macmillan Company.
- Yin, R. K. (2013). *Case study research* (5th Ed.). Thousand Oaks, CA: Sage Publications.