

HOW TEACHER EVALUATION SHAPES CONCEPTIONS OF GOOD PRACTICE:
POLICY INTENTION, IMPLEMENTATION & INTERPRETATION

Laura Gutmann

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in
partial fulfillment of the requirements for the degree of Doctorate of Education in the School
of Education.

Chapel Hill
2014

Approved by:

Jocelyn Glazier

Kevin Bastian

Julie Justice

George Noblit

Sharon Ritchie

© 2014
Laura Gutmann
ALL RIGHTS RESERVED

ABSTRACT

Laura Gutmann: How Teacher Evaluation Shapes Conceptions Of Good Practice: Policy Intention, Implementation & Interpretation
(Under the direction of Jocelyn Glazier)

During the 2010 Race to the Top (RttT) grant competition, the U.S. Department of Education encouraged states to link teacher performance ratings to evidence of student achievement and hold teachers accountable for meeting performance standards. As a result, winning RttT states like North Carolina implemented more rigorous teacher evaluation policies that emphasize documenting proof of instructional effectiveness. Although prior research has focused on the efficacy of particular evaluation mechanisms, there is less information about how evaluation impacts teachers on the ground level. This study asked, “How, and to what extent, does teacher evaluation policy influence teachers’ conceptions of good practice and professional identity development?”

Professional identity development is complex and multi-faceted. Teachers’ conceptions of their role are influenced by factors like prior schooling experiences, personal beliefs, and teacher education. Contextual factors within schools, such as leadership and collaboration with colleagues, also play a role in shaping teachers’ approaches to the classroom. As a result, the effects of implementing policies like teacher evaluation must be understood in relation to other key drivers of practice.

First, this study used discourse analysis to examine federal and state level policy rhetoric that established the intended purpose of updating teacher evaluation measures. Next,

this rhetoric was compared to narrative accounts of evaluation experiences from North Carolina PreK-3 teachers. Overall, evaluation only somewhat mattered to them, as it often failed to live up to its full potential. Flaws in implementation and little sustained connection between evaluation ratings and improvement strategies contributed to evaluation's lack of influence. In addition, aspects of their teaching that had greater value to teachers were largely based on sustained, meaningful relationships with other members of their educational community. If performance feedback was not based on a foundation of trust and respect, teachers found it to be less effective.

These findings point towards considering how evaluation is actually being carried out in schools. In addition, they highlight the importance of developing relationships with teachers and better connecting evaluation feedback to ongoing professional identity development.

ACKNOWLEDGEMENTS

Thank you to my patient and dedicated advisor, Jocelyn Glazier. Thank you to the faculty who guided me along the way, especially my committee members Kevin Bastian, Julie Justice, George Noblit, and Sharon Ritchie. Thank you to colleagues and mentors who allowed me to be part of their work. Thank you to my fellow graduate students for their encouragement and inspiration.

Thank you to the teachers who participated in this study for sharing their story with such insight and thoughtfulness. I admire what you do each day and feel truly fortunate to have gotten a glimpse into your professional lives.

Most of all, thank you to my family and friends who are like family. Thank you to Harold, who truly is the best husband in the world, and my sweet Abbie Rose. You not only made this process possible, you made it worthwhile. I feel so lucky to have had you along for the ride.

TABLE OF CONTENTS

	Page
LIST OF FIGURES.....	ix
Chapter	
1 TRACKING THE TRAJECTORY OF TEACHER EVALUATION POLICY.....	1
Framing an Approach to Examining Evaluation Practices.....	1
What Does Teacher Evaluation Look Like?.....	4
Major Types of Evaluation Structures.....	5
How Did More Rigorous Evaluation Become Popularized?.....	12
Early Evaluation Practices.....	12
Growing Demands to Measure Teacher Quality.....	19
Recent Policy Trends that Shaped Evaluation Efforts.....	23
From Local to National: The Effects of Increased Federal Influence.....	25
Studying Teachers' Evaluation Experiences and Analyzing Evaluation's Relative Influence.....	29
2 UNDERSTANDING THE ROLE OF POLICY WITHIN THE BROADER SCOPE OF TEACHERS' PROFESSIONAL IDENTITY DEVELOPMENT.....	32
How Professional Identities Are Formed.....	33
Identity as a Socially Constructed Concept.....	33

	What Specifically Contributes to a Teacher’s Professional Identity Development?.....	37
	Teacher Evaluation Policy’s Place in Teachers’ Lives.....	58
3	A TWO-PART QUALITATIVE EXPLORATION OF THE EFFECTS OF TEACHER EVALUATION ON TEACHERS’ CONCEPTIONS OF GOOD PRACTICE AND PROFESSIONAL IDENTITY DEVELOPMENT.....	64
	The Need for Additional Qualitative Research About Evaluation Policy.....	64
	Analysis of State and Federal Policy Documents.....	66
	Teacher Case Studies.....	67
	Research Methodology: Discourse Analysis.....	68
	Details About Part One: Documentation of Policy Environments That Define Teaching Roles.....	73
	Details About Part Two: Building Rich Case Examples to Illuminate Evaluation Experiences.....	78
4	PART ONE: TRACING THE PATH OF TEACHER EVALUATION DISCOURSE AND EXAMINING POLICY RHETORIC ABOUT EVALUATION’S INTENDED PURPOSE.....	98
	Federal and State-Level Policy Makers Assign High Importance to Rigorous Evaluation of Teacher Performance.....	98
	Setting the Stage on the Federal Level.....	98
	Federal Messages About Role of Teaching During RttT Competition.....	101
	North Carolina’s Response to Feedback About Their Evaluation Plan.....	111
	The Further Development of North Carolina’s Current Teacher Evaluation Process.....	113

5	PART TWO: ANALYZING TRENDS WITHIN TEACHER EVALUATION EXPERIENCES.....	129
	“Who Has a Relationship with a Rubric”? : Current Teacher Evaluation Policy Does Not “Really” Matter to Teachers.....	129
	Teachers’ Impressions of the Official North Carolina Teacher Evaluation Process.....	129
6	IT’S ALL RELATIONAL: IMPLICATIONS FOR THE FIELD.....	192
	Teachers’ Overall Impressions of Evaluation’s Influence.....	193
	Major Themes From Across Teachers’ Evaluation Experiences.....	195
	The Overarching Importance of Relationships.....	196
	Evidence from Prior Research that Relationships Influence Teaching Practice and Policy Implementation.....	201
	Evaluation as Filtered Through School Context.....	203
	Evidence from Prior Research that Context Influences Teaching Practice and Policy Implementation.....	205
	Comparing Major Trends in Teachers’ Experiences to Evaluation Rhetoric.....	206
	Considerations for Policymakers and Practitioners.....	211
	Limitations to the Findings.....	220
	Future Directions for Research.....	221
	APPENDICES.....	226
	REFERENCES.....	252

LIST OF FIGURES

Figure		Page
1.1	Examples of factors that may contribute to a teacher's professional identity development.....	2
1.2	Sample evaluation metrics.....	15
2.1	Examples of factors that may contribute to a teacher's professional identity development.....	38

CHAPTER ONE

TRACKING THE TRAJECTORY OF EVALUATION POLICY

Although the forces that shape teachers' views of their job responsibilities and drive their instructional goals are varied and complex, evaluation policies that define teaching success have emerged as potential influences on pedagogical behavior (Berryhill, Linney & Fromewick, 2009; Olsen & Sexton, 2009; Day, Flores & Viana, 2007; Lasky, 2005). The advent of dramatically more explicit and rigorous performance guidelines provides a distinct opportunity to examine, *"How, and to what extent, does teacher evaluation policy influence teachers' conceptions of good practice and professional identity development?"* As policymakers weigh the efficacy of policies and procedures that seek to produce data on teacher effectiveness, information about their role in teachers' lives can demonstrate how the standard expectations outlined within evaluation metrics have translated into the workplace.

Framing an Approach to Examining Evaluation Practices

This study's goal was to map the connections between evaluation policy intention, implementation realities, and teachers' professional identity development within a set of cases in North Carolina. In doing so, it compared the potential for teacher evaluation standards and performance feedback to affect pedagogy with teachers' own interpretations of evaluation's purpose, meaning, and influence. In addition, since contextual factors within schools can drive implementation differences, it was vital to consider how school setting affected how evaluation goals were framed and processed. For instance, depending on the context, evaluation could be perceived as an opportunity for collaborative problem solving and

improvement, an administrative requirement with limited true impact on practice, or overly prescriptive and punitive. Regardless, the climate within local schools can contribute to how evaluation policies are presented, executed, and interpreted.

The following graphic captures this relationship between self, school, and state, which, in turn, characterizes major components of evaluation interactions for teachers. In combination, these factors can determine what evaluation will look like and what its implementation will mean for teachers who interpret and then, to varying degrees, internalize its influence (*Figure 1.1*):

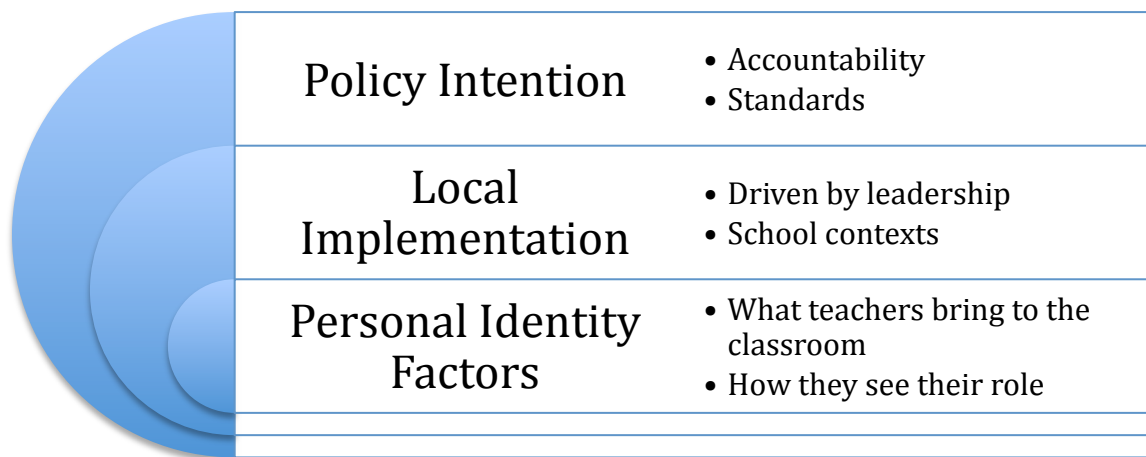


Figure 1.1. Examples of factors that may contribute to a teacher's professional identity development, in this case, in response to evaluation policy guidelines.

These major categories of influence contain sub factors like the relative importance of evaluation compared to other mandates, the nature of evaluation training provided to schools, the utility of evaluation tools, and the consequences or rewards attached to performance (Spillane, Reiser & Reimer, 2002; Cohen & Spillane, 1992). Similarly, school leadership, observation consistency, the availability of mentoring, student population, grade level, and school setting can each shape the nature of evaluation practices in the field (Ingersoll & Strong, 2011; Hope & Pigford, 2002; Leithwood, Steinback & Jantzi, 2002; Fullan, 2001;

Leithwood & Montgomery, 1982). Components of teachers' own identities, which include person-based characteristics such as prior preparation and experience, further determine teachers' outlook towards pedagogical success and subsequent practice decisions. These individualized factors can impact the relative influence of evaluation policy on practice, as evaluation enters a crowded marketplace of previously established ideas about teaching (Beijaard, Verloop, & Vermunt, 2000; Gee, 2000; Zeichner & Gore, 1990; Feiman-Nemser, 1983; Lortie, 1975).

While the increased standardization of evaluation seeks to create common messages about expected performance, the elements that shape teachers' behavior go beyond a straightforward response to policy inputs (Zembylas, 2003; Holland, Lachicotte, Skinner & Cain, 1998). As a result, while the potential for assessments of teachers' performance to affect their beliefs, practices, and interpretations of good teaching is worth investigating further, it needs to be explored within the framework of the broader educational ecosystem. Because teaching is situated within particular reform climates, school communities, and personal views of practice, new inputs must get filtered through those contexts (see **Figure 1.1**). The array of factors that contribute to teaching and learning environments add several layers of complexity that often prevent policies like evaluation from making a distinct impact on schools.

In this chapter, I will position North Carolina's teacher evaluation policy relative to socio-historical accounts of how teachers' performance has been measured in the U.S. throughout the lifespan of our public education system. In doing so, I will address questions like: *What are some common mechanisms for evaluating teachers? What is their stated purpose? How have these mechanisms been received? How have prior evaluation*

experiences informed current evaluation practices? This background information will set the scene for later analysis of key performance management standards and procedures that have evolved on the federal, state, and district level.

What Does Teacher Evaluation Look Like?

While various iterations of teacher evaluation based on performance standards have been attempted over the decades, members of the public education system continue to debate what teacher evaluation should be used for, as well as what implementation should look like on the ground level. Evaluation “involves collecting and using information to judge the worth of something” (Darling-Hammond, Wise & Pearce, 1983, p. 290). Teacher evaluation is a tool for examining several different aspects of a teacher’s performance. Its structure therefore depends on its objectives, which may include documenting teaching quality and “helping teachers improve their performance as well as holding them accountable for their work” (Stronge, 2006, p. 1). If it is meant to identify the key components of a teacher’s practice, it may include a rubric designed to rate demonstrated mastery of central teaching behaviors on a developmental continuum. This type of evaluation is focused on real-time educator efforts and driven by observational data, as well as continual follow-up conversation. If it is meant to assess how effective a teacher was in driving student progress, it is focused on compiling data around a teacher’s role in producing learning outcomes. Within such metrics, effectiveness is defined as “a teacher’s ability to improve student learning as measured by student gains on standardized achievement tests” (Little, Goe, & Bell, 2009, p. 1). More recently, states have increased their ability to use evaluation results to rank a teacher’s relative value as a human resource. Because of this, teacher evaluation is often referred to as part of performance or talent management, and associated with notions of

holding teachers responsible for student learning. Today's evaluations typically include some combination of management objectives and relate instructional inputs to achievement outputs in increasingly complex ways (The New Teacher Project, 2010).

Major Types of Evaluation Structures

Value-added models. Value-added models (VAM) that measure gains in achievement while taking students' prior performance into account have become one of the most popular evaluation tools for tracking a teacher's effect on student growth, especially among states that have begun connecting performance ratings to proof of learning gains. By analyzing predicted achievement scores that also incorporate average grade-level performance within a particular school, VAM can provide more precise data about expected academic progress. However, even as researchers continue to tinker with VAM to improve their capabilities, views about their widespread utility are mixed.

Proponents argue that refined VAM can produce vital information that helps administrators identify performance abilities and improve teacher quality. In their view, VAM do not have to exclude other forms of appraisal, and even when emphasized, remain a promising way to hone in on actual student learning and knowledge gained, as opposed to mere delivery of information or effort on the part of the teacher (Hanushek, 1992; Sanders, 2000). Tests of their reliability have in many cases shown them to be sufficiently accurate for use as part of a more comprehensive system (Glazerman, et al., 2011; Kane, McCaffrey, Miller & Staiger, 2013). In addition, analysis of the validity of existing value-added data from states like Tennessee indicates that VAM can give us better assessment data than states have previously been utilizing. Sanders (2000) believes that VAM have immeasurably improved the ability of districts to understand student outcomes and teacher effectiveness,

and steadfastly argues that a “rigorous value-added approach is the fairest, most objective way to hold districts and schools accountable” (p. 335), as well as provide useful diagnostic information. Tapping into test data helps the school system align measures of student learning with grade level standards, and provides a relatively efficient way to calculate progress for thousands of classrooms.

Meanwhile, detractors of VAM worry that highlighting end results, especially given flaws within VAM metrics, creates an unfair focus on achievement scores that may ultimately fail to capture the entirety of a teacher’s efforts. They point out that evidence supporting their utility still includes several caveats, since VAM “are usually based on student outcomes over a very narrow set of domains; they have substantial measurement error; and they are usually only available for a small subset of teachers” (Glazerman, et al, 2011, p. 11). Fears about the growing overuse of VAM led a high-profile group of researchers to co-sign a brief arguing that, “although standardized test scores of students are one piece of information for school leaders to use to make judgments about teacher effectiveness, such scores should be only a part of an overall comprehensive evaluation” (Baker, et al, 2010, p. 2). Their major concerns include a lack of evidence showing that evaluation systems based on VAM would correctly identify the weakest teachers, as well as doubts about any positive effect on teacher motivation that would lead to instructional changes. Because “test scores alone are not sufficiently reliable and valid indicators of teacher effectiveness to be used in high-stakes personnel decisions” (Baker et al, 2010, p. 2), these researchers favor a more holistic approach to tracking performance and assisting teachers in improving their practice.

Since VAM narrowly define what teachers should be held responsible for, programs that tie their results to merit pay, promotion, or tenure are particularly contentious (Darling-Hammond, 2012; Ravitch, 2012; Baker et al., 2010; Ekert & Dabrowski, 2010; Korte, 2002). One common claim is that, “Accountability policies that rely on measures of short-term value added would do an extremely poor job of rewarding the teachers who are best for students' longer-run outcomes” (Rothstein, 2010, p. 177). In part, these challenges are due to differences in classroom composition from year-to-year and the difficulty in accounting for the non-random assignment of students to teachers. According to some, this causes evaluation plans that attach significant consequences to limited data from VAM to have validity issues (Kane & Staiger, 2008). It is also complicated to isolate a teacher’s effect on students from other factors like personal motivation and ability, family influence and prior learning experiences that may play a role in current learning outcomes. If, “Growth measures implicitly assume, without justification, that students who begin at different achievement levels should be expected to gain at the same rate, and that all gains are due solely to the individual teacher to whom scores are attached” (Baker et al., 2010, p. 9), they may be perpetuating an overly narrow sense of who should be held responsible for student progress, as well as the notion that all students should progress equally within the same time period. In classrooms with struggling students, teachers may become especially discouraged by needing to produce scores at the same level as colleagues who have fewer challenges to overcome in order to demonstrate adequate progress (Clotfelter, Ladd & Vigdor, 2010; Baker et al., 2010; Finnegan & Gross, 2007).

Standards-based evaluation systems. Alternatives to VAM generally include cycles of observation, reflection, and support that provide teachers with multiple touch points

with administrators over the course of the year and connect their behavior to pedagogical standards. One of the most prominent voices in this type of evaluation reform is Danielson (2001, 2007, 2010), who provides schools with a framework for providing ongoing feedback about a teacher's performance within areas like professionalism, preparation, instructional ability and classroom environment. These contextual factors provide a starting point for capturing detailed information about a specific teacher and the type of experience they are providing for their students.

Since this approach is more broadly conceived than models based only on achievement outcomes, it provides teachers with a greater voice in the process through mechanisms like discussing observations or documenting professional accomplishments in portfolios based on standards for their grade level. As such, it may do a better job of capturing the entirety of a teacher's efforts and increase buy-in to the system (Adams & Holland, 2002). Danielson (2001) explains if evaluation is instead "conducted in a highly negative environment with low levels of trust" (p. 15), it will be difficult to produce accurate and detailed data that can help teachers identify their specific strengths and weaknesses. After Cincinnati employed an evaluation system based on her recommendations, Milanowski (2004) examined whether or not the resulting observation ratings were in alignment with test score outcomes, and found that they could, in fact, predict future achievement on some level and provide schools with a valid way to gauge performance. Because this type of system allows for on-going discussion between principals and teachers connected to targeted professional development, the quality and amount of information obtained about each teacher might also "make it possible to compare teacher practice effects with effects at other levels and provide evidence for the comparative effectiveness of interventions aimed at affecting

teacher practice versus reducing class size or improving principal leadership” (p. 49-50). By driving continual conversation about how to help teachers improve their practice and meet standards, these evaluation metrics provide a more holistic sense of how to move teaching forward over time.

Difficulties in implementation of standards-based evaluation models. For many districts, however, the resources needed to carry out such an extensive observation plan might not seem worthwhile compared to the relative ease of utilizing standardized exam data, which is more readily tracked, aggregated, and aligned with desired achievement benchmarks (Amrein & Berliner, 2002; Popham, 1999). As a result, although the assumption is that an investment in continuous observation will lead to valuable next steps such as interventions aimed at improving instruction and elevating teachers’ abilities, time-crunched administrators may find it difficult to provide a high level of follow-up support. In addition, although there are moderate correlations between results from standards-based rubrics and value-added models, researchers observed differences in how school leaders implemented evaluations on different grade levels and in varied academic settings (Kimball & Milanowski, 2009). This might be a particular issue in settings where “there is a low level of accountability for accurate evaluation unless a teacher’s job is at stake; evaluators are not required to take follow-up training; and the ratings have little consequence for most teachers” (Kimball & Milanowski, 2009, p. 63). Therefore, although Danielson’s framework tries to provide districts with a template to follow, it may not always be in sync with actual implementation. If administrators want to focus teachers on the behaviors and actions that lead to better quality instruction, they will have to make a deeper commitment to carrying out such approaches with fidelity.

Case examples of comprehensive systems. One well-known approach that tries to integrate observation data with test scores is known as the TAP model, or the Teacher Advancement Program. It aims to create cohesion around a series of measures that intertwine evaluation with coaching and post-observation professional development. TAP has also tried to provide performance incentives like opportunities for teacher leadership roles or additional compensation and bonuses. In this model, value-added measures of student achievement are used to supplement formative components like the feedback received during the series of observations held over the course of each semester. As a result, this model provides us with a strong example of how to effectively blend data and engage whole school communities in the evaluation process. In fact, researchers have found that the highest-performing TAP campuses show school-level effects that highlight the importance of coming together to, “more successfully carry out the site-based collaborative approach to growth and accountability” (Daley & Kim, 2010, p. 39). Across all TAP schools, the blended use of different types of data fosters an environment where standardized test scores and rubric-based information can complement and provide “parallel validation for the other as an accurate measure of a teacher’s instructional quality” (p. 39). A study of a similarly blended model in New York also revealed that subjective evaluations by mentors provided “meaningful information about a teacher’s future success in raising achievement” (Rockoff & Speroni, 2010, p. 249). This data compared favorably with the objective measures that solely aimed to capture test score gains, while enhancing teachers’ ability to use such information to make instructional improvements.

Current implementation considerations from across the range of evaluation models. For the last 120 years or so, we have been trying to create systematic ways of

measuring educators, while paying too little attention to the implementation drivers like leadership style and school culture that make these standards meaningful (Darling-Hammond, 2012; Derrington, 2011). This is a major concern for those who believe that evaluations emphasizing teacher outputs and requiring extensive documentation of practice will, “profoundly demoralize teachers, as they realize that they have lost their professional autonomy and will be measured according to precise behaviors and actions that have nothing to do with their own definition of good teaching. Evaluators will come armed with elaborate rubrics identifying precisely what teachers must do and how they must act, if they want to be successful” (Ravitch, 2012, p. 1). The publication of ratings in mainstream media outlets has only amplified the sense of mounting pressure to adhere to a set of expectations that may not be appropriately conceptualized or accurately applied by administrators charged with rating teacher performance. Consequently, the current ethos around evaluation is being perceived as focused on delivering summative ratings rather than fostering ongoing professional development (Danielson & McGreal, 2000; Stronge, 2006).

Given the potential for varied evaluation implementation and interpretations of its purpose to present challenges for both administrators and teachers, states and districts should consider what they could learn from our nation’s prior experiences with performance management. As new models of evaluation are introduced, it is important to recognize that various iterations of teacher evaluation and performance management systems, although different in scope than today’s broadly applied models, have been tested, tweaked, and debated over the course of America’s schooling history. This history gives us a basis for understanding how evaluation has become aligned with broader societal objectives for teacher performance.

How Did More Rigorous Evaluation Become Popularized?

To describe the major shifts in evaluation strategy that created the context for current performance management practices, in this section, I compare widespread efforts to standardize personal accountability for student learning to prior attempts to measure progress that were historically rooted in each community. Tracing the path of management practices within the profession reveals how and why these changes occurred within the policy environment surrounding school systems.

Early Evaluation Practices

Making in-person observations of teachers has been common practice since the late 1800's, when town officials like selectmen and ministers took on the task of inspecting teachers and visiting their classrooms (Borthwick, et. al, 2010; Hazi & Rucinski, 2009). Oversight of school management was highly localized, allowing communities to maintain tight control over personnel decisions. Hiring was often driven by personal relationships with prospective teachers who embodied prevailing ideologies about the type of character that educators should possess. Teachers were expected to serve as role models for students by displaying a proper commitment to their faith and adhering to conservative behavior (Ellett & Teddlie, 2003; Peterson, 1982). In addition, although an inspector could judge the quality of recitations and expect students to display their knowledge, standards for judging these accomplishments were more flexible than today.

Because teachers during this era typically had nothing more than a high school-level education, towns did not expect them to display exceptional content knowledge. Most did not have the benefit of specific teacher training or experience learning pedagogical methodologies, either—they had simply done well enough in their own educational pursuits

to be considered for the job. “Thus, teachers were largely evaluated on their personal characteristics rather than evaluation procedures informed by a knowledge base about effective teaching and learning” (Ellett & Teddlie, 2003, p. 103). Each town’s selectmen “laid down prescriptions regarding methods and materials” (Peterson, 1982, p. 30). While the advent of superintendent positions helped make the educational system more autocratic, supervisors still operated within the boundaries designated by local control structures.

The growth of supervision. After the school system grew past the point of having only one teacher per school, administrative roles were quickly established to manage staff and maintain school culture, beginning in the early 1900’s (Borthwick, et al., 2010; Peterson, 1982). At the turn of the century, schools were poised to follow the industrial model of organization, taking the principles of an era of increasing efficiency and transferring them to educational settings. Heads of schools or districts were seen primarily as business managers rather than instructional leaders. While some leaders wanted to create a cooperative relationship with teachers, others promoted a more authoritarian approach to utilizing scientific principles that allowed them to refine teachers’ instructional methods. At first, less formal “early [evaluation] practice involved entering the classroom as inconspicuously as possible, sitting in the back so as not to disturb the teacher, showing a sympathetic attitude, and taking notes” (Hazi & Rucinski, 2009, p. 32). However, especially in populous urban areas, “the ‘line & staff’ organizational structure of the industrial sector, with its gradations of levels of responsibility was adopted” (Peterson, 1982, p. 3). This meant that schools following this industrial model elevated the role of didactic supervisors while simultaneously reinforcing the idea that teachers should remain compliant and fall in line with established job expectations. Because administrators now handled much of the certification and hiring

process, “the right to control teaching was, therefore, fought out at the supervisory level, where the rigor of science was seen as being preferable to the imposition of idiosyncratic views of constituted improved teaching” (Gitlin & Smyth, 1989, p. 11). Within this model, feedback about teaching was not designed to foster dialogue between administrators and teachers or collaborative problem solving. Rather, it was meant to identify specific elements of instructional efficiency that teachers could then be directed to incorporate into their practice. By depicting teaching as a scientific endeavor, measures of good practice were then seen as best understood by expert supervisors within the field, rather than teachers themselves, who could only speak from their ground-level experience with children. This diminished teachers’ ideas and gave disproportionate power to those above them within the educational hierarchy. Consequently, it was less likely for evaluations of practice to provide teachers with feedback to enhance their professional growth, since assessments of their efforts were primarily seen as a way for supervisors to manage personnel and make decisions about retention (Gitlin & Smyth, 1989; Stronge & Tucker, 2003).

This scientific, research-based approach to evaluating teaching was further inspired by thinkers like Junius L. Meriam, who wrote *Normal School Education and Efficiency in Teaching* in 1905 as he tried to link professional scholarship about ideal pedagogical behaviors to perceived teaching ability. Despite its lack of reliable information, due to variation in school environments and imprecise measures of what strong teaching ability might look like, his work spurred future exploration in this area:

Evaluation prior to 1905 received scant attention as a field for study and development. Evaluations were apparently made on the basis on personal opinions developed by one’s official superiors. Following the publication of Meriam’s landmark study, however, a sustained and growing interest in the development of scientifically accurate methods for the appraisal of teachers’ performance and effectiveness were studied and developed (Peterson, 1982, p. 26).

During this time, there were several attempts made to hone in on the traits that characterized good teaching, which now went beyond personal attributes to include physical, moral, dynamic, administrative, social, projected and achieved efficiencies (see *Figure 1.2*).

Evaluation of these traits was based on observed assessments of how well teachers were able to execute each prescribed aspect of teaching performance (Liu, 2011; Stronge & Tucker, 2003). For example, rating administrative efficiency required evaluating how well teachers were able to tackle the organization-, planning-, and documentation-based components of their responsibilities, such as maintaining student records. However, while such rating scales tried to assign a numerical point value to achievement of required tasks and behaviors, they still left of room for administrator subjectivity and interpretation.

Categories	Number of Points
Physical efficiency	12
Moral nature	14
Dynamic efficiency	24
Administrative efficiency	10
Projected efficiency	6
Achieved efficiency	24
Social efficiency	10
TOTAL	100

Figure 1.2. Sample evaluation metrics adapted from “A Tentative Scheme for the Measurement of Teaching Efficiency” by Edward C. Elliott, originally published in 1910 (Peterson, 1982, p. 13).

As supervisors determined how to apply vague rating scales across a variety of school environments, researchers attempted to make these broad categories of job expectations more

precise and technical. For example, Frederick Taylor contributed to the use of scales to rate teacher effectiveness “based on the assumption that if scientists could study the most effective teachers, descriptors of their behaviors could, in turn, be used to rate, then transform, the ineffective and inefficient ones” (Hazi & Rucinski, 2009, p. 32). During this period, “along with the origin of the standardized testing movement, the teacher-rating instrument (i.e., check sheet, score card, and the like) also grew and became quite popular among administrators” (p. 32). As a result, supervisors tried to employ checklists that would allow them to rate a teacher based on a single visit and then make summative determinations of his or her success in the classroom (Danielson & McGreal, 2000). Unfortunately, the problems that Meriam and Taylor encountered in determining appropriate measures for assessing teaching have not disappeared in today’s schools, as the standardization of teacher evaluation is still challenged by vast differences between both school settings and individual classrooms.

Early resistance to evaluation. From the start, weaknesses within evaluation measures prompted resistance from the field, sparking enrollment by teachers in professional organizations that pushed for a more democratic system and set the stage for future struggles against management (Petersen, 1982). While raising the bar to establish higher standards was initially applauded as a way to elevate the status of the profession, teachers ultimately found themselves at odds with an efficiency movement that sought to manage them much like workers on an assembly line. Teacher advocates such as William Bagley argued that although aspects of teaching could be seen as applied science, the job was also a bit of an artistic endeavor, making it difficult to describe in an instruction manual or assign a rating to performance (Gitlin & Smyth, 1989; Peterson, 1982). Today’s policy leaders still struggle

to set evaluation standards that will satisfy both educators who see teaching as a fluidly constructed and responsive endeavor and reformers that believe stricter measures are necessary, given dismal inequities in student achievement that point to, among other things, a lack of consistent teacher quality.

Initial attempts to refine evaluation methods. While professional associations argued that teaching is inherently ill suited for rigid bottom line expectations (Rowan, 1994), social scientists claimed that schools simply needed a better way to measure teaching efforts and resulting educational outcomes. In the 1920's-1940's, the testing movement began to gain additional ground, as teachers were asked to give out pre- and post-assessments to track student progress over the course of the year. "Researchers were trying to measure the different processes, the products of the schools, and by these measurements not only to standardize, but also to rationalize every step in the procedure" (Peterson, 1982, p. 53). The science of evaluation led to the development of more complex teacher rating systems, despite ongoing debate about how these mechanisms should be used. (Marzano, Frontier, & Livingston, 2011). A central question was: Should teachers be held accountable for producing proof of student learning, or were they merely responsible for the delivery of information?

Within attempts to refine how teachers were being measured, district and school level supervisors began to judge factors like discipline, student interest, respect for the teacher, and methods of presentation, formalizing the inspection process and adding some more specific criteria to the loose structure of previous classroom visits. They also began paying attention to "traits that teachers naturally possessed, such as voice, appearance, emotional stability, warmth, trustworthiness, and enthusiasm. Educators of this era believe that teachers who

possessed these traits were more likely to perform effectively, so they became the centerpiece items in the local teacher evaluation criteria” (Liu, 2011, p. 41). However, there were some practical difficulties in implementing these evaluation systems smoothly, even as the study of supervision and information about student achievement became more common (Danielson & McGreal, 2000; Liu, 2011). Supervisors and researchers were not sure if, “the teacher as a person or his teacher performance should be the focal point of evaluation” (Peterson, 1982, p. 47). These types of ratings continued to contribute to feelings of insecurity among teachers, who worried that the “general impression” method would make observations a highly political endeavor, with favoritism and personal preferences eclipsing recognition of their hard work and numerous struggles.

In addition, even as educators resisted the subjectivity of ad hoc judgments, they remained resistant to their work being reduced to a number on a scale. As principles of scientific management drove an expanded list of expected competencies, quantifying teaching efforts ran counter to the notion that teaching should be more responsive and fluid than a series of precisely calibrated behaviors (Ellett & Teddlie, 2003; Gage, 1972). “Because teaching viewed as an art encompasses elements of personal insight (as well as theoretically grounded professional insight), the teacher as artist is expected to exercise considerable autonomy in the performance of his or her work” (Darling-Hammond, Wise & Pease, 1983, p. 292), flexibly crafting approaches to fit each classroom’s composition. How, then, could evaluations value teachers’ own discretion, yet hold them accountable for certain expected behaviors?

Initial efforts to seek teacher cooperation. When formal evaluations were first introduced across larger school systems in the early to mid-1900’s, some principals worked

with teachers to foster collegial relationships and build a shared desire to make the entire school successful. Simultaneously, the continued rise of unions led to more explicit parameters for observations, reached during collective bargaining agreements that sought to limit some of evaluation's more problematic features (Borthwick, et al., 2010). However, since teachers were ultimately responsible for their own performance, this created a dilemma for advocates of a more community-minded process, especially given the inevitable variation among staff members in terms of their teaching style and success with their students. Even though teachers were encouraged to be active members of cooperative work environments, there was also a greater demand for excellence that made it difficult to avoid focusing on individual assessments. Amid concerns about declining school systems, by the 1970's and 1980's, accountability advocates had begun to shift their focus "from broad issues of finance and program management to specific concerns about the quality of classroom teaching and teachers" (Darling-Hammond, Wise & Pease, 1983), which prompted local education authorities to consider how to enforce more stringent standards. Although there was not widespread use of set evaluation mechanisms, reformers had a renewed interest in developing them.

Growing Demands to Measure Teacher Quality

A shift towards demanding higher quality instruction first came on the heels of Sputnik, which alarmed the public and drove conversation about ensuring that our schools were globally competitive. The context of both "Sputnik and the Cold War focused additional attention on education by raising fears that Soviet students were better educated than American students. The Cold War brought about the desire to find better teachers in order to compete with the Soviets" (Markley, 2004, p. 1). Mass communication and access

to media allowed for widespread discussion of related educational issues such as lags in student performance, which also created new pressures for teachers.

Continued attempts to create evaluation objectivity. Because different iterations of the “science of evaluation” had been complicated to employ on the ground level, the 1960’s and 1970’s brought renewed interest in refining evaluation methods to make them more reliable. Even though it was difficult to evaluate teaching, schools felt that they needed tools that could provide fair assessments of performance while helping strengthen insufficient instructional practices (Marzano, Frontier, & Livingston, 2011). They were also trying to address growing concerns among citizens who wanted better proof of their return on investment in education. Popham (2003) explains that:

The chief reason for what seems to be an explosion of educational testing is that U.S. educational policymakers, bent on making the nation’s educators more accountable, want hard evidence regarding how well public schools are performing. These policymakers, and most of our citizens as well, believe that student test performance should be the ultimate yardstick by which we measure a school’s effectiveness. Naturally, then, teachers are under pressure to raise their students’ test scores. You know the logic: high test scores signify good schooling and low test scores signify bad schooling (p. 4).

Although Popham emphasized the need to create mutual satisfaction with evaluation, this was not always easy to accomplish. His work around the validity of different types of assessment continues to spark debate about the value of summative versus formative measures and how their results can be utilized to provide feedback while reinforcing bottom line expectations.

As taxpayers clamored for ways to ensure that their support of the school system was funding adequate educational conditions, administrators were forced to address their obligation to the public while mediating teachers’ concerns about the evaluation process (Peterson, 1982). George Redfern suggested that administrators should focus on identifying

specific areas in need of improvement and then utilize this knowledge to promote targeted growth for each teacher. He envisioned administrators collaborating with staff throughout the school year and allowing the teacher to, “set specific goals toward which he will work” (p. 85). At the same time, numerous researchers also tried to satisfy the demand to know more about the habits of excellent teachers by looking at patterns of behavior and formulating a better understanding of what their practice looked like, eventually considering what other teachers might be asked to emulate within performance standards.

State-level changes. By the 1980’s, the fear of keeping incompetent teachers in the classroom had led to several state programs and initiatives intended to keep teachers in check while encouraging them to work harder to meet standards. In the wake of the warnings about American schools’ weaknesses presented in *A Nation at Risk*, “renewed and more sophisticated efforts to evaluate teachers were viewed by many politicians and education policy makers as the bottom line in efforts to improve education in the USA” (Ellett & Teddlie, 2003, p. 106). The major problem, however, was that no clear path had been set forward, leaving each state agency to grapple with exactly how to organize productive evaluation systems. Some policymakers and administrators wanted to keep concentrating on identifying visible markers of good teaching that included performance, classroom behavior and learning gains. However, Medley, Coker, and Soar (1984) warned that while information about student experiences in classrooms could be useful, taking the next step towards focusing on final learning outcomes was less reliable. They believed that in order for teachers to be make sure that teachers were, “evaluated as professionals, not as technicians” (Medley, Coker, & Soar, 1984, p. 6), administrators needed to account for factors like differences in student population and provide contextualized critique that would

help practitioners develop their pedagogical knowledge without becoming overly demoralized.

Furthermore, because the “desire for accountability does not always match the capacity to collect the information necessary to reach judgments” (Duke, 1995, p. 5), it seemed as if the work administrators had put into testing varied methods of performance management was headed towards being largely ineffective. Even if teachers could be identified as sub-par, union pressures and the potential for litigation made their dismissal complicated. Innovations like pay-for-performance measures were generally opposed by teacher organizations that wanted the focus to be on professional development. “Among their stated concerns are the qualifications of those charged with making judgments about merit, and the instruments used to collect the data upon which such judgments are placed” (Duke, 1995, p. 7). These types of programs got the most pushback and were the hardest to implement. “While *A Nation at Risk* and other national commission reports called for programs that would pay bonuses or promote teachers whose performance and productivity were clearly outstanding, these turned out to be the programs most strongly resisted in the educational community” (Brandt, 1995, p. 30). As a result, the struggle to strike a balance between motivating teachers to meet rigorous individual expectations and collectively fostering their development continued.

The lasting effects of business-like management models. As today’s schools consider ideal models of personnel management, residual effects of the nation’s history with teacher evaluation include an ongoing inclination to favor documenting end teaching results like student achievement over analyzing the pedagogical choices that may lead to those outcomes (Ellett & Teddlie, 2003). In addition, efforts to link teacher inputs to learning

outcomes still mirror the efficiency-minded values and standardized operating procedures that marked the factory-approach to production (White & Lowenthal, 2009). Reformers who take a hard stance towards teacher expectations, insisting that poverty or other contextual considerations cannot be a justification for poor performance in the classroom, have looked to the business world to provide examples of appraisal systems (Ravitch, 2012). The ideology behind this approach is closely related to a belief in the power of free enterprise and the need for competitive pressures to push underperforming schools beyond the status quo of stagnancy and failure. Desired shifts in school culture therefore include more rigorous tenure and performance requirements, tighter organizational structures, increased management of human resources, and the ability to link rewards and consequences to student results (Ravitch, 2012; White & Lowenthal, 2009). Within this model, evaluation serves as a means for gathering data that carefully tracks progress over time, combined with real incentives to increase instructional capacity. While questions about how to measure a teacher's work create challenges much like those raised during the time of rudimentary efficiency ratings, the use of such metrics is rapidly spreading and evolving.

Recent Policy Trends that Shaped Evaluation Efforts

The magnitude of more recent trends towards uniform conversation around set measures of teacher effectiveness follows groundbreaking federal legislation like No Child Left Behind (NCLB), which ushered in an era of accountability that broadly applied performance standards, instead of leaving the monitoring of student achievement progress up to chance or regional variation (No Child Left Behind Act, 2001). In the early 2000's, data that tracked NCLB adherence was originally intended to hold failing schools collectively responsible for student progress and call attention to drastic achievement gaps. However, the

desire to use it to inform human capital decisions became a natural next step for those who believe that the teacher is the major driver in the classroom, and the primary source of school-based influences on learning (The New Teacher Project, 2010; Goldhaber, 2009; Rivkin, Hanushek & Kain, 2005; Hanushek, 1992). This emphasis on the role of teachers in student success brings us from wanting to ensure that teachers are “highly-qualified” to enter the classroom to demanding that they are also “highly-effective” after being placed there. Because “the difference in student performance in a single academic year from having a good as opposed to a bad teacher can be more than one full year of standardized achievement” (Hanushek, 1992, p. 113), the effects of ongoing exposure to low-quality instruction are seen as exponentially disadvantageous over the course of a student’s schooling (Gordon, Kane & Staiger, 2006). In fact, the impact of teacher quality on academic performance is currently believed to significantly supersede other factors like class size or teacher certification credentials (Rivkin, Hanushek & Kain, 2005).

Within summative evaluation plans, indicators of a teacher’s effect on academic progress through the production of learning *outcomes* may hold greater value than evidence of instructional techniques and effort, or the *means* of facilitating learning within particular school contexts. In fact, recent waivers of school level NCLB requirements were in part granted in exchange for state agreements to ramp up performance management and link each teacher to annual achievement results (U.S. Department of Education, 2012). Therefore, although a focus on individual effectiveness has surfaced in previous iterations of teacher expectations and staff management, it has never before been so precisely aligned with assessments and other end-of-the-year benchmarks or overtly incentivized by federal authorities.

From Local to National: The Effects of Increased Federal Influence

Widespread local adaptation of more rigorous and precise teacher evaluation would not have been possible without the federal government creating conditions tailored to motivate change. During the Obama administration, teacher evaluation measures made a rapid ascent to the top of the education reform agenda, as isolated pilot programs became models for states scrambling to retool rudimentary performance management systems, scale up district innovations, and align themselves with federal priorities. Much of the impetus to make policy changes could be linked to a lagging economy that prompted fierce jockeying for extra dollars from the Department of Education (DOE). Major initiatives like the Department's Race to the Top (RttT) grant competition were explicit about the need to link teachers' jobs to student achievement and make better use of classroom data, while incentivizing action with the appeal of additional funding (U.S. Department of Education, 2010a). Other funding sources, such as foundations that have embraced business-minded approaches to evaluation reform, followed suit, contributing to the buzz around increased teacher accountability and corporate-style review mechanisms.

However, although these shifts seemed like dramatic steps away from the relatively simple feedback mechanisms that preceded them, prototypes had been in the works for quite some time. States would not have been able to implement new evaluation policies to meet competitive grant or waiver requirements as quickly if similar models had not already been tested on a smaller scale. For instance, in North Carolina, districts like Guilford County had already initiated pay-for-performance bonuses intended to draw quality teachers to struggling schools (Grier & Holcombe, 2008), while Charlotte-Mecklenburg Schools had also taken preliminary steps to connect financial consequences to academic achievement outcomes

(Alternative Salary Plans/Charlotte-Mecklenburg Schools, 2011; Charlotte-Mecklenburg Schools, 2009). The main difference is that today's reform leaders have moved towards participating in a national conversation, rather than working in isolation to try out state-specific innovations. For teachers, this means that their role and responsibilities are becoming further standardized as performance incentive programs are no longer confined to meeting niche needs like solving teacher shortages in particular content areas or boosting staff quality in underserved communities. As achievement scores increasingly shape salary structures and staffing decisions across entire states and districts, placing a literal value on effectiveness is the new normal.

Support from powerful national organizations. Although we can still see that “differences in process, standards, instruments, and stakes are the result of local history and the context in which the model evolves” (Borthwick et al., 2010, p. 2), updated evaluation practices have now spread across geographic boundaries. Advocacy groups such as StudentsFirst, which was founded by former D.C. Public Schools Chancellor Michelle Rhee, typify influential supporters of scaling up performance management efforts. The brand of accountability that she and her compatriots champion argues that we must “aggressively pursue reforms to teacher evaluation systems, so that parents and school leaders can reliably distinguish among great, fair, and poor performers and so that teachers can better understand their strengths and areas for growth” (StudentsFirst, 2013). As a result of this groundswell towards greater autonomy for leaders in using student results as part of hiring and firing decisions, administrators and district officials are gaining more leeway in evaluation implementation.

Responses from teacher organizations. As increased evaluation measures seem inevitable, fears about instructional impacts (Darling-Hammond, 2012) have done little to sway a renewed federal influence over state policy decisions. In addition, while unions and teacher associations have long resisted evaluation that seeks to individualize performance or threaten equal distribution of benefits, the traditions of tenure, seniority-based advancement, and limited appraisal are fading (Duncan, 2010). Although teacher associations may still protect rating systems from being unfairly administered, fear of losing a seat at the table has prompted some naysayers to concede to pressure to remain aligned with power players, jump onto the federal funding bandwagon, and at least try to influence the nature of performance management proceedings. Union leaders in New York explained the compromises made during evaluation agreements as, “good for students and fair for teachers,” since they were able to negotiate provisions that supplemented test scores to include additional performance information (NYSUT, 2012). In non-union states such as North Carolina, while teachers have rallied against measures like Charlotte-Mecklenberg’s attempts to enforce universal participation in pay-for-performance plans (Charlotte Observer, 2011), the structure of the profession is even more likely to change, as policymakers are already positioned to take advantage of a rising tide in support of increasing teacher effectiveness.

The effects of national messages about teacher effectiveness on North Carolina. Despite the potentially problematic nature of attempts to measure teacher performance, North Carolina is nonetheless committed to enforcing stricter evaluation standards that use student test data as the primary means of gauging progress. Within the application for the state’s entry in the Race to the Top (RttT) grant competition, leaders needed to demonstrate that they were prepared to put proposed evaluation programming into action (U.S. Department of

Education, 2010a). This required strong partnerships between local entities that forced them to commit to shared goals around performance management. As the Secretary of Education during the U.S. Department of Education's inaugural promotion of the RttT competition, Arne Duncan was so pleased with the effort made by applicants to create new policies that matched his recommendations that he claimed, "Race to the Top has helped advance reform more in the past eighteen months than any other program in the history of the Department of Education" (Johnson, 2010, para. 1). One of his biggest successes was spurring 17 states, including North Carolina, to change "their laws around teacher evaluation to include student achievement" (Duncan, 2010j, p. 5) over the course of just a few short rounds of competition. Because of this, RttT sped up the policy making timeline, simultaneously managing to encourage both innovation and compliance with federal priorities. In addition to highlighting promising pilot programs, it took advantage of the increased control over schools gained since No Child Left Behind (NCLB), which had already primed states to making decisions based on federal stipulations. The increase in student achievement data made available through NCLB also contributed to the possibility of linking teachers to test results, at least on the school-wide level (Podgursky & Springer, 2007).

However, despite the DOE's work to create the conditions necessary for the widespread adaptation of upgraded performance management systems, linking teachers to student achievement has been one of the boldest and trickiest changes for states to execute, even with the help of RttT dollars. This study uses North Carolina as a case example of how federal pressures to make evaluation more rigorous have played out on the state and local level, presenting both implementation challenges and opportunities for professional development. It also maps out how the intended effects of performance management become

filtered through contexts for instruction and individual perceptions of their value. While the route that policy priorities have taken from the DOE to state legislatures has been marked by a remarkable alignment with the evaluation expectations and definitions of teaching success outlined in RttT, the translation of those guidelines may vary on the district and school level.

The current body of research on varied evaluation measures has already helped states like North Carolina determine which type of performance management system might best meet their objectives (Glazerman, et al., 2011; Ekert & Dabrowski, 2010; Harris, Sass & Semykina, 2010; Rothstein, 2010; Koedel & Betts, 2009). Evidence about the most reliable evaluation tools has factored into policymaker decisions, along with input from special interest groups, cost, human resource needs, administrative capacity and community-based preferences. However, there is more work to be done when it comes to examining the effects of new evaluation policies and procedures on teachers' views of the classroom. In the coming years, states like North Carolina will be making unprecedented links between individual teacher effectiveness and student achievement scores – and so it is important to consider if such measures are having the desired level of influence on teachers' perceptions of successful pedagogical behavior. We need more information about what evaluation means to teachers to better understand how teachers interpret policy guidelines and apply them to their daily lives.

Studying Teachers' Evaluation Experiences and Analyzing Evaluation's Relative Influence

This study's examination of the potential for current teacher evaluation policy to influence teachers' conceptions of good practice and professional identity development is situated within the policy environment that shaped current evaluation measures in states like North Carolina. In Chapter One, I set the historical scene that led to the implementation of

more rigorous evaluation measures and placed North Carolina, as this study's case example, within that context.

In Chapter Two, I outline what research already suggests about how teachers' professional identities are formed, including how prominent accountability policies might affect teachers' sense of ideal classroom practice. Questions explored through the review of literature include: *What are some common influences on teachers' professional identity development and conceptions of good practice? How does the role of evaluation policy compare to the role of other prominent influences on teachers? What factors might make policy implementation more or less effective?*

In Chapter Three, I outline the methodology I used to (1) conduct critical discourse analysis of broader teacher evaluation policy themes and to (2) develop narrative case study examples that capture the specifics of local evaluation experiences. First, data analysis of key evaluation policy documents serves to highlight the messages about why evaluation is important, how teachers will be rated, and what good teaching looks like. Second, data analysis of teachers' first-hand experiences serves to identify common interpretations of evaluation's significance, meaning, and influence within schools. This structure also provides an opportunity to compare participants' perceptions of their key job responsibilities with the messages that teachers receive about what they will be held accountable for.

In Chapter Four and Five, I present this study's findings, beginning with major trends in recent evaluation policy, which I use to establish the central themes within evaluation discourse on the federal and state level. Questions explored in Chapter Four include: *How were local norms surrounding teacher accountability formed, and how well do they align with federal evaluation priorities? How has the focus on a teacher's effectiveness shaped*

performance management directions within local North Carolina contexts? What do North Carolina's evaluation policies currently tell us about expectations for teacher performance?

I then illuminate the intersection between evaluation policy intention, the range of potential influences on a teacher's professional identity and practice, and the contextual variation caused by implementing evaluation across different school settings.

In Chapter Five, participants' responses to evaluation implementation, as well as their own interpretations of what it means to be a good teacher, provided me with a ground level perspective of the policy-to-practice relationship. Questions examined in this section include: *How have these evaluation expectations affected teachers' lives? What is evaluation's relative influence on these teachers' practice? How do the realities of school contexts contribute to the relative importance of evaluation implementation? How do teachers define good practice, as compared to evaluation standards? How do North Carolina's evaluation policy expectations compare to a teacher's own ideal practices?*

The concluding chapter draws across the findings to identify key themes that reoccurred throughout the data. Chapter Six further compares policy intention to teachers' own views towards their evaluation experiences. It also links findings to relevant implications for evaluation standardization and responds to the question: Given findings from this study, *what should policymakers, administrators, and other teachers take into account as they continue to teacher evaluation mandates continue to be implemented in North Carolina?*

CHAPTER TWO

UNDERSTANDING THE ROLE OF POLICY WITHIN THE BROADER SCOPE OF TEACHERS' PROFESSIONAL IDENTITY DEVELOPMENT

Because teacher evaluation policy includes performance standards, it provides a unique basis for comparison with a teacher's own beliefs about ideal instructional practices and outcomes. Administrators providing teachers with evaluation feedback have the opportunity to contribute to a teacher's professional growth trajectory by discussing how well a teacher's efforts align with expectations. However, although stricter evaluation has become a prominent feature of recent accountability measures, it is just one of many factors that might determine how teachers decide to utilize curriculum, interact with students, or prioritize job responsibilities; in other words, how they teach (Spillane, Reiser & Reimer, 2002; Aguirre & Speer, 2000; Gee, 2000; Zeichner & Gore, 1990). Therefore, examining the effect of teacher evaluation policy necessitates positioning its influence relative to other forces that impact how teachers conduct themselves professionally. These may include prior experiences, such as training and mentoring, or individualized drivers, like personal values and motivation. Contextual variables, including how implementation plays out in one's school, can also have considerable weight in decisions about how teachers approach the classroom. Given these factors, administrators charged with carrying out teacher evaluation must consider how such policies and procedures fit within the broader scope of how teachers develop a professional identity and perceive their role.

How Professional Identities Are Formed

The concept of professional identity development is often described within research about factors that can drive pedagogical choices and influence how teachers see their role. Beijaard, Verloop, and Vermunt (2000) note that within the field of education, there are mixed definitions of this term, which range from attaining a certain level of professional expertise to acquiring “personal practical knowledge” that becomes part of each teacher’s “story” and shapes beliefs about what teachers “personally desire and experience as good” (p. 109) within their classrooms. This study is conceptualizing *professional identity development* as the process that teachers engage in as they determine what kind of teacher to be. By “interpreting oneself as a certain kind of person and being recognized as such in a given context” (p. 108) for teaching and learning, teachers synthesize the myriad contributions to their identity to form a sense of their role within their school setting. As teacher identities shift, change, and expand throughout one’s career, they also become influenced by, “the workings of historical, institutional, and sociocultural forces” (Gee, 2000, p. 100).

Identity as a Socially Constructed Concept

The construction of a teacher’s professional identity is socially situated within socio-historic parameters for ideal practices and behaviors. For instance, accepted ways of thinking about children’s development, frameworks for transmitting knowledge to students, and institutionally organized beliefs about pedagogy might all contribute to teachers’ views of their job (Spillane, Reiser & Reimer, 2002; Zeichner & Gore, 1990). Popkewitz (1998) explains that “particular sets of norms are privileged through the ‘wisdom of practice’ and concerns about the psychological management of children. It is here that we can consider

how ‘purposes’ are socially constructed through principles generated to enable teacher and students to participate and act in school” (p. 83). Shared beliefs about the purpose of teaching are the accumulated result of living within an entrenched system of societal standards that determines what public schooling is supposed to look like. These universal beliefs explain many commonalities in classroom practice that surface across schools, such as the way the school day is typically designed, dominant styles of transmitting information, and approaches to managing behavior. While different teachers see variations of these norms as they develop within the profession, they still internalize some basic notions of typical classroom structures that carry throughout their career.

Because the teaching role is affected by instructional settings, “these social spaces or ‘fields’ both enable and are characterized by particular types of practices” (Hardy & Lingard, 2008, p. 64), that educators recognize as familiar markers of appropriate behavior and pedagogical skill. Similarly, Berger and Luckmann (1966) emphasized the impact of “group think” as nested within related societal structures, positing that, “man’s self-production is always, and of necessity, a social enterprise. Men together produce a human environment, with the totality of its socio-cultural and psychological formations” (p. 51). By extension, teachers are operating within a larger context for determining appropriate behavior within their profession, structured by the affinities, institutions, and groups that they associate themselves with (Gee, 2000).

Even teachers’ demeanors are expected to reflect standards for proper comportment, as their role becomes “constituted through social interactions, performances, and daily negotiations within a school culture that privileges emotional self-discipline and autonomy (for example, where female elementary school teachers are expected to be ‘caring’ and

‘compassionate’))” (Zembylas, 2003, p. 109). Within typical settings for teaching and learning, “The language practices that teachers use in talking about the profession of teaching both hold the acceptable identities for teachers and carry the important knowledge, skills, practices, and values for teaching” (Battey & Franke, 2008, p. 129). As such, descriptions of optimal teacher characteristics and behaviors, such as independently managing a classroom or cheerfully serving the community, may contain both implicit and explicit messages about what it takes to adequately fulfill the requirements of the teaching position.

Bourdieu (1990) portrays these myriad influences on identity as inherent to our composition in ways that we might not even realize:

The conditionings associated with a particular class of conditions of existence produce *habitus*, systems of durable, transposable dispositions, structured structures predisposed to functioning as structuring structures, that is, as principles which generate and organize practices that can be objectively adapted to their outcomes without presupposing a conscious aiming at ends or an express mastery of the operations necessary in order to attain them. (p. 53)

Although educators may not be able to pull apart each of the threads that make up the knot that comprises their own sense of “teacherness”, they will likely have some sense of what appealed to them about their chosen profession and what they hope to accomplish on the job. As they carry those conceptions of ideal practice to their classrooms, their response to their school environment is driven and organized by those *habitus* components. Likewise, since *habitus* is evolving, new experiences in school, such as the introduction of a new policy or program, will become incorporated into previously established notions of what their job should entail.

Spillane, Reiser, and Reimer (2002) describe this process as “the active attempt to bring one’s past organization of knowledge and beliefs to bear in the construction of meaning from present stimuli” (p. 394). Along the way, major organizing factors such as “race, social

class, and gender mediate the socialization process and establish socialization patterns for particular groups of individuals who teach in particular kinds of schools” (Zeichner & Gore, 1990, p. 26). Similarly, Gee (2000) envisions standards for practice as being determined by a combination of personal characteristics, prior experiences, and institutional expectations that drive each teacher’s vision of classroom life. By building bridges between these various inputs and engaging in dialogue about their profession with other members of the educational community, teachers can form a sense of their identity within their workplace. While identities created through societal discourses, “are ultimately rooted in recognition processes” (p. 111) that acknowledge set understandings of how institutions like schools are supposed to work, personal interpretations of these norms ultimately “interrelate in complex and important ways” (p. 101). Wortham (2005) notes that variation is possible even within structures that promote uniform “group think”, since social identification is still “constructed in particular events and local contexts ” (p. 40). From his standpoint, even established groups are flexible enough to allow for unique responses to new situations.

Holland, Lachicotte, Skinner, and Cain (1998) further argue that while our sense of self is based on perceptions of our position within society, “position is not fate” (p. 45). Instead, dynamic interactions between historical precedent and our individual inclinations make us both “social producers and social products” (p. 42). Because the instructional decisions that flow from teachers’ convictions about their job responsibilities are situated within specific settings for practice, “identities and the acts attributed to them are always forming and re-forming in relation to historically specific contexts,” and working in concert with other influences on development in reaction to particular environments (Holland, Lachicotte, Skinner & Cain, 1998, p. 284). Each teacher can then play a role in figuring out

“who they are...in relation to the social types...and in social relationships” (Urrieta, 2007, p. 108) with the other people within their school building. The mutually dependent quality of a teacher’s relationship with her school context leads to this sort of co-constructed reality, where the outcomes of interactions between individuals and their environment cannot be fully explained by institutional norms and traditions. From this perspective, policies like teacher evaluation are but one factor that colors multifaceted layers of influence on a teacher’s sense of self and professional trajectory.

What Specifically Contributes to a Teacher’s Professional Identity Development?

Since the factors that affect teachers’ lives are intertwined, it is difficult to tease out which components most significantly contribute to a teacher’s sense of self, or “teacherness”, and subsequent instructional actions. As policy-driven efforts like performance management standards interact with other major influences on practice, educators may struggle to make sense of a crowded marketplace of competing ideas and mandates (Feiman-Nemser, 1983). The following section outlines major categories of both person- and school-based influences that are likely to work in combination to shape professional identity development and affect policy’s relative impact on teachers (see *Figure 2.1*).

First, I will explore the person-based identity constructs that teachers bring into the classroom, such as their own student experiences, family background, or beliefs about appropriate teaching objectives. I will also explore the effect of structured educational training, such as a traditional teacher education program, which is another common influence on teachers’ early views of their profession. Next, I will describe school-based contributions to professional identity that are connected to particular settings for teaching and learning. Once teachers enter the field, factors like leadership personnel, other colleagues, and the

demographics of the students they serve have the potential to affect their conceptions of good practice. Lastly, I will explain how educational policies might drive teaching behaviors and professional identity development as they interact with person- and school-based influences on teachers' lives. As part of this section, common implementation factors, as well as the particular role of accountability-focused policy measures, will be discussed in more detail.

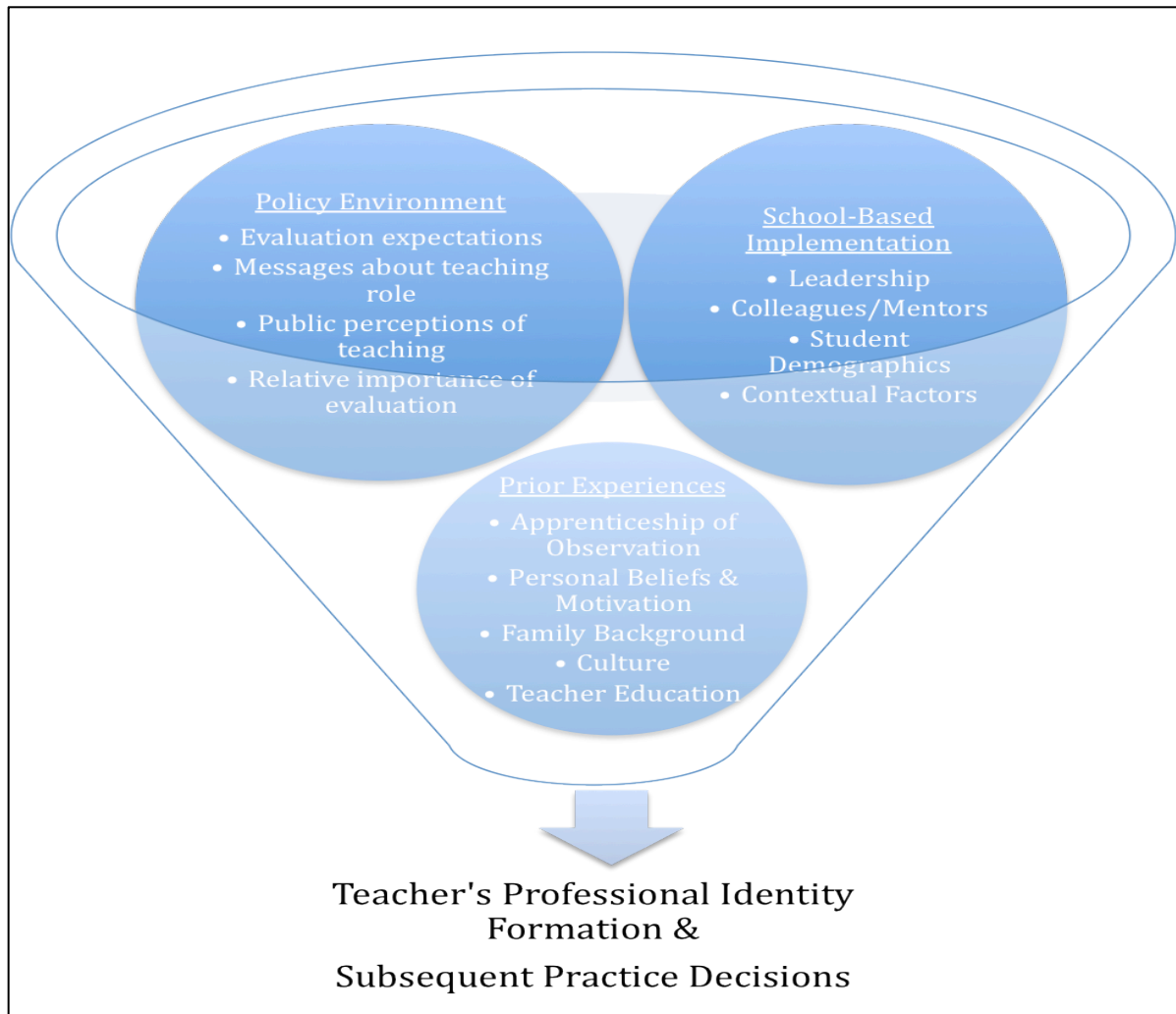


Figure 2. 1. Examples of factors that may contribute to a teacher's professional identity development.

Common person-based factors. In the context of this study, “person-based” factors are broadly defined as both relating to personal experiences, such as exposure to certain

schooling structures as a child, and more intrinsic qualities, influenced by such things as cultural background, social class, and ethnicity. These contributions to a teacher's identity are often deeply rooted, suggesting that other influences could merely become another layer on top of a well-established core. Before new teachers even have the chance to formally construct a formal repertoire of "knowledge about practice" (Cochran-Smith & Lytle, 1990), they have already accumulated a vast amount of experience in school, largely from their time spent as a student (Lortie, 1975). Their instructional ideals may stem from memories of a favorite educator who now serves as a role model, or organizational norms such as typical lesson structures, behaviors, and social dynamics that privileged their own learning. In addition, broader life experiences that go beyond educational background, such as membership in class, gender, ethnic, or other affinity groups, can provide teachers with a basis for shaping their views towards institutional structures like schools (Gee, 2000; Goodson, 1992). Lastly, if teachers participate in pre-service training before entering the classroom, their educational studies may also help form their early views towards curriculum and instruction (Darling-Hammond, 2000).

Effects of prior schooling. The perceived effect of factors like the "apprenticeship of observation" (Lortie, 1975) provides evidence that years of prior schooling, along with family-based views of educational values and norms, are stronger influences on teachers than any other pre-service experience (Feiman-Nemser, 1983). Feiman-Nemser (2001) sees the initial process of learning to teach as characterized by a desire to form "a coherent sense of themselves as professionals by combining parts of their past - including their own experiences in school and in teacher preparation - with pieces of the present" (p. 1029). While educators may come to consider other professional standards and guidelines,

accumulated impressions of teaching from when they were a student still tend to factor heavily into their visions for classroom success. These deeply ingrained norms are typically long-held, “personal values that the person regards as inextricably bound up with her or her existence” (Korthagen, 2004, p. 85). As such, they remain central to navigating daily responsibilities and instructional tasks once teachers are in charge of managing their own classrooms.

Effects of personal values and beliefs. A teacher’s personal values and convictions also have the potential to affect pedagogical choices, even if they are part of a broader belief system that was not originally applied to ideal teaching and learning behaviors. For instance, teachers may have particular convictions about work ethic, opportunity, key knowledge, and the purpose of education. Aguirre and Speer (2000) argue that these “beliefs play a central role in a teacher’s selection and prioritization of goals and actions” (p. 327), and become an apparent feature when teachers consider proposed shifts in pedagogy. Through their examination of changes in teaching approaches, they conclude that beliefs about practice also “shape how teachers perceive and interpret classroom interactions” (p. 330), causing teachers to steer students towards alignment with their own learning objectives and significantly affecting how children experience school. “School systems consist not only in rules and formal structures, but also in beliefs about authority, habits of deference and resistance, and knowledge about how things work. Cultural and social organization intertwine in these systems” (Cohen & Spillane, 1992, p. 31). As a result, underlying views of how schools and classrooms should function often affect teachers’ perspectives towards their practice.

Hargreaves and Fullan (1992) argue that because teacher development is affected by these firm internal convictions, professional growth “involves more than changing teachers’

behavior. It also involves changing the person the teacher is” (p. 7). At times, this has required teachers to reconcile preconceived notions of what school should look like with the individual preferences and learning styles of the diverse range of students in their classroom. As Delpit (2006) points out, teachers’ belief systems may not always match student needs, since “We all carry worlds in our heads, and those worlds are decidedly different. We educators set out to teach, but how can we reach the worlds of others when we don’t even know they exist? ” (p. xxiv). In some cases, teachers who view struggling students from a deficit perspective may even be allowing their preconceived ideas about these students’ abilities to obscure these students’ strengths (Gorski, 2010; Singleton & Linton, 2005; Valencia, 1997). However, in other instances, teachers’ personal inclinations can be a positive force for ensuring that they fully support children’s development. For example, Lasky (2005) studied teachers who felt strongly about maintaining a holistic approach to meeting student needs, because they “all held the belief that their core purpose as a teacher was to teach academic content while also attending to social and emotional elements of their students’ development” (p. 909). As a result, sometimes teachers’ personal inclinations and values can be an important motivational tool, even if they otherwise present challenges to shifting teachers’ established attitudes towards their role.

Effects of teacher education. As pre-service teachers prepare to bring their personal views of teaching – and the world – into varied school contexts, teacher education programs provide another way to foster the advancement of professional knowledge and build foundational skills (Clotfelter, Ladd & Vigdor, 2010; Darling-Hammond, Berry & Thoreson, 2001). Because teacher education can bolster identity formation and strengthen teacher development (Darling-Hammond, 2000), it has the potential to contribute to an educator’s

professional identity in ways that cannot simply be attributed to individual motivation, beliefs or prior talents. Within these training opportunities, pre-service candidates receive key information that makes explicit contributions to the construction of an agreed-upon “set of skills that are fundamental to safe and responsible teaching” (Ball, 2011, para. 6). Gaining subject-matter expertise is one important part of an educator’s professional development, as “the myriad tasks of teaching, such as selecting worthwhile activities, giving helpful explanations, asking productive questions, and evaluating students’ learning, all depend on the teacher’s understanding of what it is that student are to learn” (Ball & McDiarmid, 1990, p. 1). Developing these fundamental attributes of high-quality instruction subsequently enhances decision-making about how to execute the teacher role once in the field.

In addition, preparation to enter the classroom typically includes content that goes beyond subject area or pedagogical mastery. For example, teacher education courses that focus on topics like culture, race, class, special education needs, or social justice issues may allow pre-service candidates to reflect on how their own background might skew their approach to teaching, and subsequently bring greater awareness to their practice. Zeichner, et al. (1998) tell us that such courses, “can help teachers overcome a ‘blame the victim’ and a ‘cultural deficit’ orientation towards students and their families so that they can restructure schooling and classroom processes to be more responsive to a culturally diverse student population” (p. 166). Although teacher education programs may not always do enough to support diverse cultures and viewpoints (Ladson-Billings, 2005), there are models of preparation with a focus on incorporating biographical exploration, discussion of equity, and social justice themes that have succeeded in expanding teachers’ perspectives (Ray, Bowman, & Robbins, 2006). Consequently, the combination of a structured teacher

education program with teachers' prior experiences and beliefs can then provide teachers with a more informed understanding of their role that they can take with them into the classroom.

Common school-based factors. Once teachers enter the field, knowledge from their previous experiences and training is applied to their particular instructional environments and their daily work with students in their schools. Although traditional teacher education programs have the potential to add value beyond what can be gained by learning on the job (Darling-Hammond, 2000), there is some evidence that classroom experience is a key driver of teachers' educational productivity and the subsequent academic progress of their students (Harris & Sass, 2011; Kane, Rockoff & Staiger, 2006). This implies that being within school contexts and interacting with children gives teachers a substantial opportunity to develop their practice, as their school environment significantly shapes the nature of their instruction and their disposition towards their profession (Klassen & Chiu, 2010; Rice, 2010). However, merely being within a school does not guarantee positive effects on teaching that move instructional techniques forward and result in meaningful professional growth. In fact, research has shown that teachers can interpret the same standards for teaching and learning in vastly divergent ways, depending on the preferred strategies for meeting instructional expectations promoted within their school setting (Acheinstein, Ogawa, & Speiglmann, 2004). Given how different schools can be from each other, the nature of their influence on a teacher's professional identity formation and conceptions of good practice depends on contextual characteristics like leadership style, opportunities for collaboration with colleagues, and student demographics.

Effects of leadership. As teachers dissect key elements of their work culture, they have often “described leadership as a factor that influenced access to, and the nature of, learning opportunities” for staff members (Scribner, 1999, p. 253). Since administrators can alternately exacerbate and ease tension between individual pedagogical goals and organizational mandates, they have the power to shape teachers’ perspectives towards the value of various school programs and procedures. Hogg (2001) describes leaders as influential in determining the typical prototype of valued group members, as they provide focus and create cohesion by utilizing their own social attractiveness to set standards for success and define what participation in teaching entails. Even less popular leaders can provide desirable feelings of order, stability, and direction, which counterbalance trade-offs like depersonalization and loss of individual expression. “Because prototypes are relatively consensual, they also furnish moral support and consensual validation for one’s self-concept and attendant cognitions and behaviors” (Hogg & Terry, 2000, p. 124). The standards set by leadership then become, “another particularly important contextual factor which affects the success of teacher development efforts” (Hargreaves & Fullan, 1992, p. 14). While principals may not always be seen by teachers as a positive influence in that regard, “effective principals are able to define priorities focused on the central mission of the school and gain support for these priorities from all stakeholders. Their actions impinge on almost all aspects of the classroom and school that are likely to influence achievement of these priorities” (Leithwood & Montgomery, 1982, p. 335). As a result, researchers widely contend that leadership is among the most important factors in advancing student learning, in part due to the influence they have over teachers’ objectives for their classrooms (Leithwood et al., 2004).

Effects of colleagues and mentors. Aside from looking to administrators for direction, teachers often turn to fellow educators when trying to figure out how to approach their practice and establish themselves within the teaching role. As teachers “try to make sense of what is going on in their classrooms, the explanations and advice they encounter, especially from more experienced colleagues, affect their attitudes” (Feiman-Nemser, 2003, p. 27) and guide the choices they make about curriculum and instruction. However, meaningful interactions with other educators can be hard to initiate, and as such, can become a diminished component of school-based influences on teachers’ lives. Within those constraints:

Norms of politeness and the desire for harmony create additional barriers to productive mentoring interactions. Many beginning teachers are reluctant to reveal problems or ask for help, believing that good teachers work things out for themselves. Mentors may withhold assistance due to the enduring belief that teaching is a highly personalized practice of finding one’s own style (Feiman-Nemser, 2001, p. 1033).

Since teachers are held personally responsible for student learning, established social behaviors can favor seeking independent solutions to problems that are seen as being classroom-specific. Within the dominant profession culture, teachers may be reluctant to disrupt those social paradigms, causing reduced democratic exchanges of new ideas between colleagues.

Even structured mentoring programs, where teachers are formally assigned a more experienced guide, have reported implementation problems that suggest limitations to their effect on professional development and instructional growth. Despite the potential for mentors to positively affect teacher retention and create a much-needed support system (Ingersoll & Strong, 2011; Ingersoll & Kralik, 2004), significant site variation contributes to inconsistent impact. For instance, Vasquez & Urzua (2009) discovered that teachers felt

empowered after engaging in “direct reported speech” with mentors that included, “expressions of certainty, confidence, and assertiveness” (p. 13). By articulating decisions about pedagogy and discussing them with an experienced peer, they gained a sense of control over classroom decisions. However, Wang, Odell and Schwille (2008) conducted a comprehensive review of induction and mentoring program effectiveness, which revealed that, “the quality of influence is dependent on social, cultural, and organizational contexts of schools where such components are situated” (p. 148). Although observing mentors and reviewing lesson plans together can help ground dialogue about teaching in day-to-day practice decisions, opportunities for well contextualized discussion can be rare. Consequently, while these experiences still influence teachers’ overall sense of professional identity and ideal practice, their results are mixed.

On the positive side, in response to the limited ongoing impact of individual mentoring programs, many administrators and teachers have worked to create professional learning communities (PLC’s) aimed at expanding a teacher’s instructional capabilities through collaborative examination of data. These regular opportunities for discussion about students’ progress were established to bring colleagues together to share their expertise, so a school’s staff can “focus on learning rather than teaching, work collaboratively on matters related to learning, and hold itself accountable for the kind of results that fuel continual improvement” (Dufour, 2004, p. 11). In general, opportunities for colleagues to meet together can drive productive discourse about classroom realities, foster inquiry, and empower teachers to engage in meaningful dialogue (Ermeling, 2010; Franzak, 2002; Feiman-Nemser, 2001; Lord, 1994). However, the nature of a PLC’s influence, as well as whether teachers find them to be useful, partially depends on teachers receiving honest but

supportive feedback within an atmosphere in which “ideas have a chance to incubate and develop, trust builds in the group, and participants feel comfortable raising sensitive issues and risking self-revelation” (Cochran-Smith & Lytle, 1999, p. 294). This model connects well with Lord’s (1994) vision of critical collegiality, which includes questioning each other’s practices and making productive suggestions for adjustments as needed.

In contrast, less engaging and inclusive forms of professional development and collaboration may foster negative perceptions towards the value of such opportunities, causing them to be seen as nothing more than “contrived collegiality” (Hargreaves, 1992; Hargreaves & Dawe, 1990). For instance, PLC’s can serve as a way to focus staff on district priorities like examining student data from standardized benchmark assessments to determine how to improve achievement outcomes. While concentrating on these mandated goals can positively impact how teachers plan to carry out broader systemic objectives, the benefits of meeting together may not be fully realized if teachers feel like they have no autonomy over PLC agendas (Hargreaves & Dawe, 1990). These sessions have a greater chance of making an impact on pedagogy when teachers are involved in planning them, since teachers, “are more willing to invest in learning new content if they feel the enhanced professionalism that a commitment strategy affords” (Smith & Rowley, 2005, p. 148). Consequently, opportunities for meeting, collaborating, and learning with colleagues will only result in a net positive gain if teachers have a hand in their development. In addition, more influential iterations of PLC’s remain focused on factors specific to teachers’ daily work contexts, such as the particular needs of their students.

Effects of student demographics and preferences. Although interactions with the other adults in their school building help shape teachers’ professional experiences, teachers

do still spend the majority of the day in their classrooms working with the children under their care. The students that educators are assigned to teach directly contribute to teachers' views of how to approach instruction, since relationships developed within the classroom environment play a part in shaping how learning is facilitated. Experiences interacting with students can particularly enhance the skills of early career teachers who advance beyond their initial tendency "to 'stand and deliver' content" and begin "to grasp and experience the power of listening to and learning from the students" (Merseth, Sommer, & Dickstein, 2008, p. 96). As teaching goals evolve in response to changing student needs, teachers' notions of the best way to fulfill their professional responsibilities and meet learning objectives become a moving target. The daily immediacy of the interplay between a teacher's view of their role and their classroom composition makes it an important part of the identity formation equation.

However, the effect of students on professional identity development depends on how teachers view the role of teacher-student relationships within the process of "becoming teacher" and establishing their pedagogy. If teachers see students as informing their instruction, they may credit them as important influences on their practice. For instance, Lasky (2005) studied teachers who said that, "trusting, respectful relationships with their high school students were considered as a prerequisite for learning to occur" and believed that, "connection with students meant that students would take greater interest in the subject being taught" (p. 907). They saw rapport building as a key part of their job responsibilities that was essential to demonstrating teaching competence. In that scenario, teachers used their intimate knowledge of students as a tool for deciding how to best fulfill instructional standards within policy directives (Lasky, 2005; Sykes, 1990). Conversely, in an example of

obstacles to teachers viewing students as drivers of professional decision-making, Haniford (2009) observes how a teacher “constructed student interest as a product of her instruction, not something students possessed intrinsically,” and speculates that the distance between the teacher’s cultural background and that of her students caused her to, “devalue her students’ lives outside of school, instead of viewing their lives and experiences as resources she could draw upon in her classroom” (p. 994-5). While teachers generally describe their interactions with students as important components of their practice, they may not always value student input highly enough to see it as shaping their core attitudes and beliefs about what to do in the classroom. Therefore, the effect of classroom composition and student preferences on teachers’ views towards carrying out their job responsibilities is varied.

Common policy-based factors. Clearly, teachers bring many influences on their practice into the workplace, which are further shaped by contextual factors within their school setting. Whether these influences have a positive or negative impact, they are key drivers of professional identity development and conceptions of good practice. However, educational policies are another factor with the potential to impact teachers’ workplace experiences and views of their profession. As a central part of the context surrounding public schools, policy can be broadly defined as “a form of structural power” (Marshall & Gestl-Pepin, 2005, p. 4). Institutions with the power to create policies use them to guide the distribution and management of resources and establish systems of organized practices (Collins, 2000). Federal, state, and local government agencies currently implement a myriad of policies that are intended to affect teachers, who are the primary human resource employed within public schools (Loeb & Miller, 2006).

Common examples of teacher-focused policies involve training and licensure requirements, the provision of induction and mentoring programs, professional development initiatives, and salary benefits. Such policies also include provisions about the length of the teacher workday and how teachers are required to spend their time, such as when districts allot certain periods for common planning and professional development. In addition, standards for curriculum and instruction determine how teachers should impart knowledge and skills to their students, both in terms of content and delivery. Assessment of how well teachers deliver that content and foster student learning is monitored through teacher evaluation systems. Although some administrators actively involve teachers in determining policies' role within their school culture, the cumulative impact of proscribed working conditions and performance requirements can make a teacher's job seem highly structured and supervised by top-down management (Cohen, et al., 2009; Labaree, 2011; Sachs, 2001). Teachers' attitudes about policy are therefore largely determined by the nature of policy implementation within their schools. The influence of policy on teacher's overall professional identity development is also affected by the day-to-day enforcement of policy procedures.

The role of policy implementation. Ozga (2000) describes policy implementation as a negotiated process that incorporates input from those who have been charged with carrying out rules and regulations, despite not typically being involved in the initial design of their objectives. Within this "sense-making" process, educators consider how they might react to policy directives (Spillane, Reiser & Reimer, 2002). Since "many reform ideas about teaching, learning, and schooling are very value-laden" (p. 401), messages about policy priorities can prompt educators to weigh their own goals for teaching and respond

accordingly. As a contributing factor within a host of micro-level decisions made by teachers and administrators, policy implementation is driven by a series of interactive, ground level choices about how teachers ought to best approach their responsibilities (Crossley & Vulliamy, 1995; Honig, 2006). Fullan (2001) describes the conditions that affect implementation as a combination of, “the characteristics of the nature of the change, the makeup of the local district, the character of individual schools and teachers, and the existence and form of internal relationships” (p. 93), which produce distinct cultures for reform possibilities. Given their central part in determining these variables, “educational change depends on what teachers do and think” (Fullan, 2001, p. 115), as well as the expectations that administrators set when introducing new policy initiatives. Therefore, the way that policy guidelines end up getting incorporated into teachers’ conceptions of their role may not always match a policy’s original intention.

Examining disconnects between policy and practice at the point of implementation.

Studies of the impact of policy on teachers’ lives reveal gaps between intended policy and actual practice, which are often created during the implementation phase (Bartell, 2001; Cohen, et al., 2009; Elmore & McLaughlin, 1988). While policymakers, legislators, and district personnel tend to focus on how policies drive desired end outcomes, it is also helpful to pinpoint how the process of implementing a policy ultimately contributes to teachers’ perceptions of its utility and value (Spillane, Reiser, & Reimer, 2002). On one hand, education policies affect teachers through their influence on their professional conditions and job status, along with perceptions of their job and attitudes towards those responsibilities. They also affect teaching by influencing teachers’ approaches to interacting with students, conducting lessons, and determining strategies for instruction. However, as teachers interact

with such initiatives, their day-to-day actions may shape what gets carried over from rules and regulations into reality. When teachers decide how to combine their own established teacher identity with the standards set by professional institutions, they drive the translation of policy intention into actual pedagogy (Fullan, 2001).

This interpretive process can alternately prove to be either beneficial or disjointed. Some teachers can smoothly integrate their instructional approaches with performance standards, while others struggle with alignment. Lasky (2005) argues that as teachers consider the realities of policy implementation within their school contexts, “the concept of mediated agency” (p. 900) helps explain how teachers navigate new mandates and determine how to follow their guidelines while still maintaining a certain degree of autonomy over instructional decisions. Ideally, administrators will support teachers in utilizing their expertise to figure out how a policy might best work within their school environment. But, if schools fail to effectively integrate teachers into making choices about policy implementation, the ability for policy to increase the capacity for change becomes limited, resulting in incremental, inconsequential, and fragmented improvements (Cohen & Spillane, 1992). Incorporating educational policy into daily practice is complex, as the difficult task of changing entrenched structural and pedagogical norms makes adjusting teaching actions and producing desired effects less straightforward than policymakers might have envisioned (Honig, 2006; Loeb, et al., 2005).

Because governing structures are often looking for a quick, uniform fix to long-standing, complicated problems, this creates a major barrier to effective policy implementation in the real world of education (Spillane, Reiser & Reimer, 2002; Cohen & Spillane, 1992; Cohen & Ball, 1990). The American model of schooling is a strong cultural

institution, which makes it hard to change typical teacher behaviors that follow deeply rooted patterns of lesson and content delivery:

Changing one's teaching is not like changing one's socks. Teachers construct their practices gradually, out of their experiences as students, their professional education, and their previous encounters with policies designed to change their practice. Teaching is less a set of garments that can be changed at will than a way of knowing, of seeing, and of being. And unlike many practices, teaching must be jointly constructed by both teachers and students. So if teachers are to significantly alter their pedagogy, they must come to terms not only with the practices that they have constructed over decades, but also with their students' practices of learning, and the expectations of teachers entailed therein. (Cohen & Ball, 1990, p. 334-5)

Therefore, it can take time for policies to become a meaningful part of the numerous factors that already contribute to a teacher's professional identity development. The constant election cycle and politician turnover also detracts from the ability to incrementally incorporate new policies into old teaching habits with any kind of methodological consistency. One regime of decision-makers may have different demands from the next. In addition, as policies are further developed, schools may be asked to keep adjusting their practices to incorporate ongoing updates, before they have even figured out how to adequately integrate the original mandates. As a result, assessments of policy outcomes may appear to reveal that implementation was unsuccessful, when new programming simply needed more time to establish itself as the norm.

Under those conditions, the gap between policy and practice may also be due to teachers incorrectly interpreting a policy's original intentions and objectives (Spillane, Reiser, & Reimer, 2002; Spillane, 2004). If there is a lack of communication between policymakers and educators about a policy's underlying values and goals, this disconnect can lead to frustration (Bartell, 2001; Smit, 2005). Teachers often need more professional development and support to increase their ability to effectively implement policies and apply

them to their classroom context (Coburn & Stein, 2006; Cohen, Moffit, & Goldin, 2007; Elmore, 2002). While there are benefits to maintaining a “critical distance” (Keith, 2008) between stakeholders who can then hold each other accountable for policy outcomes, if teachers feel demoralized and confused by the perceived separation between schools and policymakers, they may be less likely to align themselves with a policy’s stated purpose. Because an influential “interplay exists between morale and motivation” (Finnegan & Gross, 2007, p. 624), teachers in struggling schools are particularly susceptible to their practice suffering as a result of burnout or apathy. Or, they may develop potentially undesirable strategies for creating the overall appearance of achievement, like retaining failing students, sending them to special education programs, and eliminating instruction in non-tested areas (Jacob, 2005). These unfavorable outcomes can contribute to a policy’s unfulfilled reform potential within these contexts for teaching and learning.

Disconnects between policy intention and real-life implementation may also be a result of the need to customize broadly conceived policies to fit a particular school environment. Local knowledge is, in fact, a valuable asset that helps teachers better utilize policy structures to benefit their students (Palmer & Snodgrass Rangel, 2011; Smit, 2005; Yan & He, 2012). For instance, “implementation at the classroom level is mediated not only by teachers’ preexisting knowledge and beliefs but also by who the students are and what teachers believe specifically about their students” (Palmer & Snodgrass Rangel, 2011, p. 635). However, in some schools, teachers fear reprisal if they respond to classroom dynamics by adjusting required guidelines to better engage their students, which further limits successful policy implementation within their school setting (Au, 2007; Sullivan,

2006). This phenomenon is most commonly seen when accountability-based policies are introduced that teachers find to be overly restrictive and punitive.

The particular effects of accountability policies on implementation issues. Past experiences with major policy reforms have provided some evidence surrounding the effects of accountability policy on teachers' classrooms, as well as their outlook towards their profession. Teachers may feel that these initiatives restrict their voices, limiting them from participating in dialogue about how to best achieve learning goals and forcing them to produce better learning outcomes under whatever means necessary (Olsen & Sexton, 2009; Day, Flores & Viana, 2007). Policies that seek to shift teacher behavior and improve instruction by enforcing adherence to inflexible procedures can also affect teachers' intrinsic qualities, such as motivation, confidence, and self-worth (Olsen & Sexton, 2009). For instance, accountability sanctions have been shown to limit teachers' identity development through the widespread narrowing of acceptable instructional practices and curricular choices, which almost exclusively links their worth to performance measured by test scores (Barrett, 2009). Political shifts and new pressures threaten to limit teacher practice as part of "a perceptible shift, towards a performance model" (p. 1020), making teachers hesitant to deviate from prescribed guidelines and increasing their frustration.

Berryhill, Linney, and Fromewick (2009) claim that the stress from undesirable and burdensome policy mandates has dramatically changed teachers' outlook towards their jobs for the worse and caused their self-efficacy to decline. As a whole, policies that target teachers and blame them for poor student results can diminish their professional status and subsequent feelings about their career (Labaree, 2011; Loeb, et al., 2005). They can also affect teachers' participation and engagement in professional development activities

(Desimone, Smith, & Phillips, 2007). Sachs (2001) believes that teacher growth is hampered by forced compliance with standards based on rectifying prior failure, as opposed to seeking change through collaborative engagement that builds off teachers' potential. As a result, the role of such measures in creating more difficult workplace conditions threatens job satisfaction, along with teachers' confidence in their abilities (Ravitch, 2012; Berryhill, Linney & Fromewick, 2009; Ma & MacMillan, 1999). These unintended downsides may even prompt dramatic action, such as cheating on standardized tests to either avoid negative consequences of low performance or gain rewards like test score-based teaching bonuses (Amrein & Berliner, 2002).

Within this reform climate, "teachers' work, and moral purpose, forms of autonomy and discretionary decision-making, which have been the traditional keystones of teachers' professionalism, are now being challenged and reframed into forms of audited compliance with results-driven agendas" (Day, Flores & Viana, 2007, p. 251). Barrett (2009) describes how performance pressures motivated by No Child Left Behind (NCLB) achievement goals can change the way that teachers view their instructional objectives. "Under the official pedagogic discourse embodied by NCLB, content is increasingly taught in isolated fragments connected only to standardized examinations in a strongly classified and framed curriculum" (p. 1020). As teachers find it harder to personalize their practice in response to students' individual preferences and feel pressured by legislation that demands performance on tests while ignoring important contextual considerations, these policies become seen as obstacles to doing their jobs effectively (Barrett, 2009; Sunderman, et al., 2004). Within this environment, school leadership and district personnel also increasingly monitor teachers and direct how instruction is supposed to be carried out. Systematic trends that skew an

educator's role towards focusing on meeting student growth targets may have created "a mostly mechanistic view of teacher's actions in relation to accountability-related curriculum policies", which cause teachers to be, "portrayed in undimensional ways", such as either being compliant or oppositional (Sloan, 2006, p. 121).

However, in spite of their problematic features, accountability measures can prompt teachers to focus their energy on traditionally underserved populations and adjust their practice to better align with baseline standards (Haycock, 1998; Sloan, 2006). For example, No Child Left Behind drew attention to vulnerable subgroups of children and forced schools to make staff adjustments to focus on their needs. As such, we cannot entirely dismiss the impact of teachers who, "actively read and appropriate facets of accountability-explicit curriculum policies to deliver instruction that not only is higher in quality and more equitable than their 'normal' classroom instruction, but also leads to a stronger sense of teacher agency" (Sloan, 2006, p. 124). If teachers can find ways to balance adherence to policy frameworks with utilizing their own knowledge about optimal pedagogical techniques, these guidelines may seem less restrictive and detrimental than originally feared (Lasky, 2005). Consequently, "the overall quality and equity effects of accountability policies depend on a variety of complex and interrelated factors" (Sloan, 2006, p. 146) that include multidimensional influences on a teacher's approach to navigating their role. Responses to policy directives that impact classroom practice then become a matter of interpretation, "mediated by such things as instructional materials, teachers' professional capacities, and methods of student assessment" (Cohen & Spillane, 1992, p. 12), as well as the availability of supports like professional development and collaborative planning that may ease accountability pressures.

Despite challenges to individual decision-making, “any teacher, in any system of schooling, interprets and enacts new instructional policies in light of his or her own experience, beliefs and knowledge” (Cohen & Ball, 1990, p. 335). This contributes to why the application of standards in one classroom may look completely different from the classroom down the hall – and allows us to consider why some lack of uniformity may actually result in positive outcomes for students. Teachers contribute to the effect of policy on practice as active participants in the implementation process, working to reconcile broad standards with specific realities, as well as other influences on their professional identity. Within the following discussion of teacher evaluation mandates, this study must then consider the interaction between teacher and policy, and investigate how it impacts teachers’ lives, as well as how teachers themselves can influence a policy’s impact on local settings.

Teacher Evaluation Policy’s Place in Teachers’ Lives

School environments are embedded with political decisions that lead to workplace rules and expectations like those contained in evaluation metrics. In the case of policies that delineate performance management goals, “teachers may perceive that the system used by their school district for teacher evaluation is based on an image of the teacher and beliefs about teaching that are inconsistent with their beliefs about teaching, and thus, even if given feedback from such an evaluation system, teachers might not be inclined to reflect on their practice” (Peterson & Comeaux, 1990, p. 5). Under these conditions, “teachers may view standardized evaluation procedures as simply something to ‘pass.’ Teachers may develop model lessons that are reserved only for formal evaluations, and their typical teaching may be unaffected by either the substance or the format of the teacher evaluation procedures that they experience as beginning and practicing teachers” (p. 4). Teachers particularly fear the

consequences of being evaluated when “they perceive insufficient control of both a particular dimension of teaching and insufficient control over how such a dimension is evaluated” (Conley & Glasman, 2008, p. 29). In those cases, data gleaned from performance rubrics only gains utility when it is well connected to prior pedagogical goals and context-specific decision-making.

Furthermore, evaluation itself might not be enough to positively impact teacher behavior in areas where improvements are necessary. “Control-oriented reforms tend to favor school and teacher accountability, taking a carrot-and-stick approach to school improvement...most advocates of standards-based reforms acknowledge, however, that the carrot-and-stick approach alone will not lead to dramatic gains in student achievement” (Smith & Rowley, 2005, p. 128).

Therefore, if school leaders view “failure in implementation as demonstrating lack of capacity or a deliberate attempt to ignore policy” (Spillane, Reiser & Reimer, 2002, p. 391), they may need to recognize that complex processes like adjusting pedagogy may not always yield immediate results. Initial data about a policy’s success often provides limited information about related instructional practices that were either more or less effective. In contrast, teachers who have opportunities to investigate the effectiveness of pedagogical decisions and analyze learning impacts on students are more likely to incorporate suggested reforms into their practice and interpret them in productive ways (Ermeling, 2010; Coburn & Stein, 2006). Differences in how evaluation data are being utilized therefore contribute to evaluation’s place with teachers’ lives.

Moving into the study: lessons from prior research. Because recent iterations of teacher evaluation policy are markedly more extensive and outcomes-oriented than prior

performance management efforts, their specific effects on teachers' lives are largely unknown. However, trends in policy-to-practice relationships signal a likely connection between the nature of policy implementation and the potential for evaluation to contribute to teachers' conceptions of good pedagogy (Berryhill, Linney & Fromewick, 2009; Day, Flores & Viana, 2007; Lasky, 2005; Loeb & Miller, 2006; Olsen & Sexton, 2009). Research indicates that because the quality of implementation can be inconsistent, much depends on key variables like how well evaluation policies and procedures are related to teachers' current approaches to instruction (Cohen & Spillane, 1992; Crossley & Vulliamy, 1995; Fullan, 2001; Honig, 2006; Spillane, Reiser, & Reimer, 2002).

In addition, as evaluation is carried out in local schools, its significance can be understood as relative to other influences that contribute to a teacher's professional identity development (Beijaard, Verloop, & Vermunt, 2000; Gee, 2000). If local implementation of a policy's original intentions is mediated by a combination of school and person-based influences, the interactions between these factors may determine the degree of evaluation's impact, and generate congruence, co-existence, or conflict (Cohen & Ball, 1990; Lasky, 2005; Sloan, 2006). Therefore, evaluation's success partially depends on how well performance management plans and procedures fit into the ecosystem of school context, as well as a teacher's broader picture of ideal classroom life (Feiman-Nemser, 1983).

Lastly, evaluation is an example of an accountability policy. Common concerns about accountability efforts and standardized expectations for sound teacher practice include their effect on teacher agency, or the ability to flexibly make decisions about practice, as opposed to being restricted by rigid guidelines. In addition, teachers typically have concerns about whether the performance ratings generated by accountability policies are designed to

be high or low stakes, and summative or formative in nature. However, while broader literature on teachers' historically strained view towards accountability measures, especially post-NCLB (Barrett, 2009; Berryhill, Linney, & Fromewick, 2001; Day, Flores & Viana, 2007; Labaree, 2011; Loeb, et al., 2005; Ma & MacMillan, 1999; Ravitch, 2012; Sachs, 2001), allows us to predict that teachers may have similar attitudes towards evaluation, there is limited data specifically addressing how teachers are reacting to the evaluation measures implemented in states like North Carolina over the past four years (from 2010 forward). Even if new information simply confirms existing views towards the effects of accountability policy on teachers' lives, documenting teachers' experiences will still lead to a richer understanding of evaluation's relative influence.

Moving into the study: qualitative approaches to examining complex relationships between policy, practice and identity. As this study examines how teachers might be influenced by evaluation policies, it will focus on potential connections between performance management and resulting beliefs about instructional practice. It will also explore the range of complex social structures that inform a teacher's view of her role. Darling-Hammond, Wise and Pease (1983) argued that this approach is needed to counteract research that only, "seeks to measure the worth of teachers by reference to the product or output of their work" (p. 292), and sees students as raw material that can be molded at will. In the past, "the general failure of policy analysts to look at the ways in which policy wended its way towards and ultimately came to rest in schools and other social agencies left them without a useful explanation for the limited effects found by the research designs and methods then in use" (Darling-Hammond, 1990, p. 340). Because the translation of professional identity and performance expectations into professional identity development

and conceptions of good classroom practice is a multifaceted and continual process, it is difficult to determine the full effect of policies like teacher evaluation by simply focusing on set outcomes. In addition, since research tends to examine influences on practice in isolation from each other, it often fails to adequately describe the intertwined nature of such variables and their combined role in shaping teaching personas.

Within this study, North Carolina serves as a case example of where evaluation fits within the range of factors that can contribute to teachers' growth trajectories within the field. Capturing more details about local teachers' experiences with evaluation policies helps reveal how, and to what extent, performance management measures impact teachers' impressions of preferred practice and either support or detract from their development. As part of this data collection, teachers discuss the degree to which their own professional development processes have been influenced by systemic policies. I also examined whether the feedback teachers received as part of the official NC Teacher Evaluation Process promotes meaningful conversations about teachers' instructional choices and professional growth (Conley & Glasman, 2008).

To this end, I asked teachers to describe 1) how evaluation is being carried out in their school; 2) which factors determine the nature of its implementation; 3) what they are "taking away" from evaluation; and 4) how "in sync" evaluation is with other influences on their conception of good practice and professional identity development, such as person- and school-based factors. Within this study, person-based factors are the characteristics and prior experiences that a teacher brings into the classroom. School-based factors are the organizational and structural characteristics that shape local school context and a teacher's workplace environment. One goal of the study was to determine how these variables interact

when it comes to policy implementation within local school contexts (*see Figure 2.1*). In brief, the research study traced the path from 1) evaluation policy intention to 2) school-level implementation and 3) resulting individual interpretations of evaluation's value (see *Appendix A*), examining “*How, and to what extent, does teacher evaluation policy influence teachers' conceptions of good practice and professional identity development?*”

CHAPTER THREE

A TWO-PART QUALITATIVE EXPLORATION OF THE EFFECTS OF TEACHER EVALUATION ON TEACHERS' CONCEPTIONS OF GOOD PRACTICE AND PROFESSIONAL IDENTITY DEVELOPMENT

The Need for Additional Qualitative Research About Evaluation Policy

When a new policy is introduced, those in charge of its design and implementation often consider how major stakeholders might react to their efforts and anticipate potential changes to those stakeholders' lives. In the case of teacher evaluation, policymakers are typically most concerned with determining how well performance appraisals measure teacher effectiveness and help monitor instructional quality. For example, related research might focus on how well evaluation mechanisms measure yearly growth and isolate a teacher's impact on student achievement (Baker, et al., 2010; Ekert & Dubrowski, 2010; Rothstein, 2010; Kane & Staiger, 2008), or assess the consequences of linking evaluation ratings to pay-for-performance bonuses (Clotfelter, Glennie, Ladd, & Vigdor, 2008; Figlio & Kenny, 2007; Podgursky & Springer, 2007; Dee & Keys, 2004). However, policymakers pay less attention to the impact of implementation on the daily experiences of teachers, who contribute to evaluation's success or failure as "street-level bureaucrats" (Lipsky, 1980; Marshall & Gerstl-Pepin, 2009). This oversight minimizes the impact of teachers' choices about how to execute performance management standards within their classrooms.

Qualitative research methods, such as case study narratives and discourse analysis of local policy contexts, have the power to illuminate contextual factors that shape how policies like teacher evaluation are carried out. They can also bring teachers' voices forward to

explain the nature of the relationship between evaluation standards and teachers' own perceptions of their role. This data is especially revealing in cases where the unique characteristics of each school setting cause variation in how evaluation policy guidelines translate into reality. By holding a magnifying glass over the intersection between state evaluation policy and school-based implementation, policymakers can compare their original intentions with individual interpretations of the evaluation process. Although policymakers do not always allocate resources to gathering information about educators' own views of performance guidelines, this type of qualitative data might help them assess evaluation's impact relative to other pedagogical drivers.

The primary questions that guided this study are, "*How, and to what extent, does teacher evaluation policy influence teachers' conceptions of good practice and professional identity development?*" In keeping with this focus, **Part One** of this study *examined the evaluation priorities within federal and local policy guidelines*, while analyzing related public discourse about those performance management processes. In **Part Two** of the study, teachers considered *how those evaluation expectations related to their personal views of their professional identity and the responsibilities associated with the teaching role*. The resulting data were used to create narrative examples of teachers' experiences with evaluation policies in North Carolina. Analysis of this data addressed key sub-questions:

- *How do North Carolina's expectations for teacher performance compare to a teacher's own ideal practices and instructional objectives?*
- *What is evaluation implementation like for local teachers?*
- *How does school context affect evaluation's ability to influence these teachers' lives?"*

Analysis of State and Federal Policy Documents

The **first part** of data gathering and analysis included discourse analysis of federal and state policy documents that marked the introduction of more rigorous evaluation measures in North Carolina (from 2010 forward). These data set the scene for the use of North Carolina as a case study in evaluation implementation and provided contextual information about how these policies define successful teaching. Detailing the messages within evaluation policy about expectations for teacher performance and examining how the teaching role is presented and positioned with the official North Carolina Teacher Evaluation Process also provided a basis for comparison with teachers' own beliefs about their role. Within this analysis, I examined data in two stages to establish a baseline understanding of the public messages about teaching responsibilities. First, I synthesized prior analysis of federal-level communication about evaluation priorities to determine how the environment for more rigorous local evaluation was established. Next, I outlined the nature of North Carolina's response to federal pressures by examining the current state evaluation process, along with related training materials.

In the first part, data included (1) transcripts from 11 federal-level speeches Secretary of Education Arne Duncan made in 2010 that reference teacher evaluation, during a key initiative to persuade states to consider the need for greater evaluation, (2) state-level policy documents that reflect the local changes in evaluation policies that followed, such as sections of the North Carolina's Race to the Top application addressing teacher evaluation and the North Carolina Teacher Evaluation Process manual, and (3) a PowerPoint presentation used to introduce the new North Carolina Teacher Evaluation process to local educators during state-mandated training sessions.

Teacher Case Studies

The **second part** of the study brought forward the lived experiences of local teachers who had been subject to evaluation measures by using narrative details to formulate composite examples of how they believe evaluation has impacted their practice. Data sources included information from surveys, general focus group discussion, a structured focus group activity, and individual follow-up interviews. Other factors that shaped teachers' perspectives towards ideal teaching behaviors were positioned relative to the effect of receiving observation feedback under the new evaluation policy guidelines. Together, this data led to a better understanding of *how evaluation affects teachers' professional identity development and subsequent conceptions of good practice*, which was the ultimate focus of analysis. It also provided the basis for understanding teachers' (1) interpretation of evaluation expectations, along with (2) the implementation factors that drive how those expectations are carried out at their school; and (3) the relationship between evaluation and other factors that they consider to have significantly shaped who they are as teachers (See *Appendix A*).

By connecting policy content with narrative accounts of teachers' experiences with evaluation, several objectives were accomplished. First, I juxtaposed policy intention with how teachers internalize and interpret evaluation requirements. Secondly, I dissected the role of school settings in implementing evaluation to reveal what drives varied experiences with the execution of such policies. *Does context create significant variation in how evaluation policies are carried out? How does evaluation's influence compare to person- and school-based factors that may also play an important role in shaping a teacher's practice?*

Research Methodology: Discourse Analysis

Discourse analysis was the dominant methodology used in this study, particularly in Part One. Discourse analysis examines how language is used to build meaning and reflects on how these meanings are put into use (Gee, 2005). In this study, discourse analysis was used as the methodology for examining the interrelationship between policy discourse and teachers' own discourse about evaluation. Discourse in this context means, "a distinctive way to use language integrated with 'other stuff' so as to enact a particular type of (however negotiable and contestable) socially situated identity (type of person)" (p. 46). In other words, the language that we use to express ourselves helps shape our identity by connecting those utterances, along with any associated behaviors and actions, to societal conceptions of our position within particular groups. In this case, the framing of evaluation by policymakers, government officials, and state agencies established an intended purpose for implementing more rigorous evaluation. Teachers then discussed what they thought of evaluation implementation from their perspective, while describing the broader context of influences on their conceptions of practice.

In response to the need for more specifics about evaluation circumstances, critical discourse analysis, in particular, provided tools for identifying how widely held beliefs about teacher responsibilities come to be presented and by whom. This methodology was especially essential to examining policy documents in Part One of this study. By applying this type of theoretical framework to issues like teacher evaluation to see *who* is allowed to define teaching roles, *how* they do so, and *what* the effect of such mechanisms might be, I could "explore how language works in policy texts, and in particular how it can be used to document hybrid genres and discourses, and to highlight competing discourses and

marginalized discourses” (Taylor, 2004, p. 444). The critical approach offers an important counterpoint to other types of analyses: “Proponents charge that traditional policy studies, particularly those based on economic models of behavior, take a narrow and technocratic approach to policy choices, and that they diminish the meaning of politics and obscure the role of values in defining policy alternatives” (White, 1994, p. 508). In contrast, a critical approach examines the underlying ideologies and beliefs that drive policy decisions and identifies the power dynamics within social structures and institutions (Fairclough, 2013).

Researchers may consider policy documents to be drivers of a “conversation” between states/districts and teachers, establishing expectations that make significant contributions to the nature of school working environments (Fairclough, 1995). These documents can be further understood as evidence of an “interpretive system” (Taylor, 2004) that informs aspects of a teacher’s identity: “This interpretive system may be people’s historically and culturally different views of nature; it may be the norms, traditions, and values of institutions; it may be the discourse and dialogue of others; or it may be the workings of affinity groups” (Gee, 2000, p. 108). Documents that describe teaching roles, responsibilities and success markers therefore provide us with specific evidence of how perceptions of ideal instructional objectives might be formed.

Because this study first focuses on data within written texts such as teacher evaluation policy rubrics and guidelines, it makes sense to identify the broader themes within these documents, making their content and underlying values – or *what* is said – as important as *how* it is described. Gee (2004) suggests that while the form and function of language should remain central to any examination of discourse, an understanding of situated meaning

becomes germane to our interpretation of communication patterns within historical and cultural frameworks.

Examining policy documents to understand how evaluation is being positioned.

By examining what Gee (2005) calls “building tasks” of conversation, this study recognized how language works to create significance, activities, identities, relationships, politics, connections, sign systems, and knowledge that serve to organize our discursive process (p. 11). This approach incorporated foundational descriptions of critical discourse analysis with Gee’s notions of how social practices are organized. Gee (2005) tells us that as part of this complex interaction between the form and function of language, “We do recognize or assemble situated meanings based on context, but we also construe the context to be a certain way and not another based on the situated meanings we assemble” (p. 65). Widely distributed policy documents contribute to the context for policy implementation by establishing common understandings of their purpose. Even though readers view such documents as being products of the institutions that created them, the documents themselves contribute to our characterization of that institution’s intentions.

Woodside-Jiron (2004), for example, applies those principles to the policy arena, while considering how values and power dynamics function within regulatory and rule-bound institutions. Her research “draws attention to particular texts, discourse practices, and social practice issues that are particularly relevant to thinking about the engineering of social change through language and practice” (p. 176). When a policy is introduced, word choice gives some indication of the policy’s underlying implications and intentions. Tracking both thematic and structural trends within policy discourse about teacher evaluation in North Carolina helped illuminate how proposed evaluation measures were sold to the public and

introduced to those who would be the most directly affected by any changes to the law. This type of analysis also identified priorities within policy texts that sought to move teachers in a particular direction.

The dialogical nature of policy language. By using critical discourse analysis tools to recognize how “language plays a primary role in the creation of meaning” (Apple, 1996, p. 130) within layered social and political contexts, these methods began to pinpoint how policymakers attempted to fulfill performance management implementation goals, as well as the effects of that implementation on stakeholders within school communities, like teachers. Because language is dialogic, it was important to not only to examine what was said and written, but also what was heard or received. The players within this educational policy arena were all subject to each other’s views of evaluation, as they took on the simultaneous roles of speaker and listener. However, the educational arena does not necessarily provide a level playing field, as the policymakers, educational agencies, and teachers producing this discourse still couched their language within socially bounded genres that indicated the nature of their role. A better understanding of these dynamics:

facilitates the exploration of how policies that are presented as reality serve primarily as political rhetoric; how knowledge, power and resources are distributed inequitably; how programs...reproduce stratified social relations; how schools institutionalize those with whom they come in contact, and how individuals react (i.e. resistance or acquiescence) to such social and institutional forces (Young, 1999, p. 685).

In the case of teacher evaluation, it was valuable to compare patterns and themes within teachers’ contributions to teacher evaluation policy discourse with public depictions of evaluation by government officials who may have relatively more authority. In doing so, this study showed how stakeholders making higher-level decisions about policy implementation positioned the role of the teacher and made claims about teachers’ views towards being

assessed. It also characterized the part that teachers play in their own “evaluation story”, which may itself vary depending on context and perception.

Relationship between policy and teachers: Case study narratives. A broader qualitative examination of teachers’ experiences with evaluation helped me form the case study narratives in Part Two. Case study narratives are commonly used as a way to provide rich details about individual experiences and “add to humanistic understanding” (Stake, 1978, p. 7). In composite form, they draw from across a study’s participants to provide representative accounts of trends from throughout the data. In this study, teachers’ stories provided detailed examples of the nature of teachers’ discourse about evaluation policy implementation within their schools, which were situated within the broader North Carolina educational policy environment. By building case examples of teachers’ experiences with evaluation that consider context, a “specific and focused form of linguistic analysis can be connected to questions about social construction in organizations” (Fairclough, 2005, p. 926). In other words, drawing from teachers’ descriptions of how evaluation is being enacted within local institutions helps explain how school, district, and state-level culture affect a policy’s impact. Since case studies relate the complexity of a particular experience to broader contextual circumstances (Stake, 1995; Yin, 2009), I was able to relate teachers’ individual experiences with evaluation policy to widespread evaluation discourse and the educational climate that fostered increased performance management.

Because initial analyses of the messages within relevant evaluation policy documents would simply provide a beginning point for understanding how expectations for teacher performance are framed and presented, it was crucial for me to gather more information about their effects on teachers’ daily lives. In Part Two of the study, I discussed the content

of policy-driven messages and evaluation procedures with participant teachers, who could relate policy discourse to their own ideas about what they should be held responsible for, and share their thoughts about how evaluation had impacted their practice. As a result, I could connect the implications of evaluation policy identified within Part One to the narratives of real-life experiences that emerged in Part Two.

Details About Part One: Documentation of Policy Environments That Define Teaching Roles

In North Carolina, the primary policy document that describes the components of a teacher's job and outlines performance expectations is currently known as the *North Carolina (NC) Teacher Evaluation Process* (Public Schools of North Carolina, 2012a). The state Department of Public Instruction created this guide to evaluation in 2009, and slightly revised it in subsequent years to include more details about additional components like value-added growth measures that track student achievement progress. It contains a description of its purpose and outlines responsibilities within the review process, while providing a rubric to guide classroom observations and assign end-of-year ratings. Therefore, analyzing its content revealed how teaching success is being defined within state policy objectives that ask administrators to determine performance scores.

My examination of related communication, such as a PowerPoint presentation used to introduce North Carolina educators to the new evaluation process (Public Schools of North Carolina, 2012b), offered an understanding of how this policy was situated. These documents were my entry point into capturing the thinking behind increased teacher accountability measures and outlined key areas of focus. Analyzing their content provided discursive insight into how the teaching role is being positioned within school communities. As I tracked trends within the presentation of evaluation standards, the nature of their

potential effects on teachers' perceptions of their job was made apparent. By describing the priorities and values revealed within this policy discourse, I set up further exploration of the relationship between evaluation expectations and views of the teaching role, as they related to the choices that teachers make about how to approach their classrooms.

Data collected during this **first part** of the study addressed sub-questions like, *“What does analyzing federal and local policy guidelines reveal about evaluation priorities, and how do these priorities connect to one another? How do they describe the teaching role and related responsibilities, as well as indicators of success? What messages might they contain for teachers working to construct their professional identity and determine appropriate practice? Whose voices contribute to this dialogue around how to measure teaching success? What is emphasized in these documents (e.g. ongoing growth/process or end results/product)?”*

Details about data sources and analysis for Part One. North Carolina's guide to teacher evaluation includes information about the accountability measures and performance expectations that evolved from federal pressures to conduct more rigorous assessments of teachers' impact on student results. These guidelines were at the center of my discourse analysis of performance management trends within North Carolina, as I traced the path from the U.S. Department of Education's reform agenda to state level adaptation of updated evaluation policies and procedures:

- First, I analyzed 11 speeches made by Secretary of Education Arne Duncan from March through August 2010 that specifically reference teacher evaluation (in archived transcript format). My analysis narrowed in on the time period between RttT Round 1 and Round 2 winners being announced, when states had a chance to

adjust their application to reflect preferred federal evaluation approaches, including a growing emphasis on teacher effectiveness over instructional process or training qualifications. For example, North Carolina had the opportunity to commit to linking teacher performance to student achievement as the state responded to the U.S. Department of Education's feedback about their initial RttT application. As a result, 2010 marked a tipping point that cemented teacher evaluation's place on the education reform agenda. This context provided a broader backdrop against which to situate the North Carolina case.

- When I analyzed these speeches, first I identified all mentions of teacher evaluation. This included related terms such as performance management and teacher effectiveness. Next, I broke that data into subcategories based on my research interests. Because I wanted to establish the intended purpose of evaluation reform, as well as key messages to states within the RttT agenda, I tracked justifications for making evaluation more rigorous. I then identified discourse that included specific suggestions for updates to state-level evaluation systems. At this point, additional areas of analysis emerged based on the patterns and themes that surfaced throughout Duncan's evaluation discourse. As a result, I decided to capture how 1) uniform evaluation implementation and 2) the role of teachers were being framed. To do so, I analyzed patterns of positioning language used to form agreement and consensus. This language was also used to construct images of ideal teaching behavior (Fairclough, 2005; Gee, 2005).

- I then analyzed North Carolina’s RttT application (U.S. Department of Education, 2010a), as well as federal reviewers’ comments about the state’s proposed reforms to managing “teacher effectiveness” (U.S. Department of Education, 2010b; 2010e). This provided me with another source for demonstrating how evaluation changes were made possible via RttT funding, which I used to establish North Carolina as a specific case example of recent implementation of rigorous evaluation policies.
 - o My analysis again included an examination of how teaching roles and the purpose of evaluation were being described and positioned with North Carolina’s RttT application, in relation to stated RttT priorities. As a result, I looked for parallels between federal and state language, to determine if the federal agenda had any influence on state evaluation actions. For instance, did the language in North Carolina’s proposed evaluation reforms echo Duncan’s rhetoric? Did it match application requirements and reviewer expectations? Did reviewer comments reinforce federal messages about evaluation priorities? How did the state adjust or clarify its evaluation plans in response to those comments?
- Finally, I analyzed local policy documents that related to evaluation implementation. The primary document that I examined was the aforementioned guide to the official *NC Teacher Evaluation Process*, which includes the rubric for evaluation observations within an outline of the performance management process. This document also contains directions for completing the evaluation cycle, as related to the observation rubric and its standards for teaching, as well as

templates for designing future professional development plans. As I looked across the official evaluation guide, I identified common ways that teaching priorities or the primary responsibilities that comprise a teacher's job are described. My analysis also focused on the ways that the purpose of evaluation was depicted, the nature of suggested teaching behaviors, and the use of formative versus summative language. Next, I triangulated my content analysis of the NC Teacher Evaluation Process guidelines with data from a PowerPoint presentation created by the North Carolina Department of Public Instruction to train teachers on the new evaluation process during this guide's initial distribution. In doing so, I could confirm how these rubric requirements and objectives were initially presented, positioned, and framed by state and district personnel who introduced these concepts to local implementers, school employees, and the general public. Overall, reviewing this data helped me identify key points - or *cruces* tensions (Fairclough, 1995) – within examples of how evaluation was supposed to be carried out.

- Within my analysis of the NC Teacher Evaluation Process guidebook, I first identified stated purposes for updated evaluation measures. I also looked at patterns across language that described a teacher's role and responsibilities. Then, I looked at the action verbs and phrases within evaluation standards to determine what types of behaviors teachers were expected to display.
- Within the evaluation training PowerPoint, I examined data that showed how the role of evaluation, as well as the purpose of being rated based on

standard measures of practice, was being positioned relative to teachers' professional growth trajectory.

Ultimately, I used each data point to build evidence of how North Carolina policymakers and the state Department of Public Instruction defined the purpose of this particular type of teacher evaluation. Analyzing these documents also contributed to my understanding of how North Carolina's current teacher evaluation standards define good practice. For local teachers who must adhere to these standards, the NC Teacher Evaluation Process that resulted from the state's response to federal priorities contains messages about how they might best contribute to their school community, and how those contributions will be measured.

Details about Part Two: Building Rich Case Examples to Illuminate Evaluation Experiences

The second part of the study describes the degree of influence that performance management directives might have on a teacher's perspective of her pedagogy. It also places evaluation within the context of school environments, which contain a variety of potential influences on beliefs about the teaching role. As performance management mechanisms are implemented, their effects are mediated by a combination of school and person-based factors. The result of the interplay between these "variables of interest" creates a path from implementation to teachers' interpretations of evaluation's significance (see *Appendix A* for a framework that traces this path). To create a basis for comparing policy intention, implementation realities and personal perceptions of ideal practice, I juxtaposed evaluation procedures with teachers' descriptions of their actual evaluation experiences. I also positioned official performance standards with teachers' views of their job responsibilities. Consequently, the data in this part of the study helped me show how interactions between evaluation and other influences on teachers' professional identity development might affect

teachers' conceptions of good practice, as well as further growth within the teacher role. Therefore, my data collection and analysis in this second part specifically related the details of teachers' experiences with evaluation to questions such as, *"How do evaluation rubric expectations relate to teachers' personal views of their professional identity as educational professionals? What factors affect evaluation implementation? What other factors, aside from evaluation, drive practice decisions and approaches to the classroom?"*

Selecting participants. For this study, I recruited twelve PreK-3rd grade school teachers from across three different public, North Carolina elementary schools. In recent years, new assessments have been constructed to review teachers at all levels, including the early elementary years. Most prominently, rubrics for rating teachers based on observations of their practice provide a generic template for principals to evaluate every teacher in the state. To better understand how evaluation using these rubrics has evolved, I ensured that participants had been in their current school since at least 2011. Furthermore, although I wanted to recruit multiple teachers within the same school setting to discuss how evaluation was being presented within their workplace, and their principals were aware of this study, participants independently agreed to share their experiences, rather than being directed to take part by administrators.

As such, I solicited volunteers through their prior participation in a multi-year, collaborative research project that had worked with teachers to identify typical classroom practices, discuss potential reforms to their curriculum and instruction, and engage in related professional development. To better understand how localized implementation variables can affect teachers' perceptions of their responsibilities, this new study drew participants from three of the prior project's school sites to represent a range of working environments and

compare teacher experiences within and across those venues for instruction. Although the prior project had focused on ways that these teachers might better serve African-American, Latino, and low-income students, I asked the participants in this separate research effort to engage in broader conversation about how they determine appropriate practice and develop instructional strategies. In addition, I asked them how teacher evaluation fits into school contexts that are crowded with other reform priorities, of which their prior research project is just one example. While this sample of North Carolina teachers was recruited primarily out of convenience, the prior work that participants had done to explore their teaching practices may have prepared them to feel comfortable analyzing the factors that contribute to their pedagogical choices and teaching goals. As a result, they were prime candidates for this type of research exploration.

The participants also had similarities beyond their involvement in the same prior project. Although they are located in disparate regions of the state, in different, non-adjacent districts, each of these teachers' schools serves between 450 and 650 students. The first school is located in the smallest, most rural town out of the group, with about 11,000 residents. About 75% of students are minorities, and about 80% receive free or reduced lunch; in the past, the school has performed below the state average in math and reading (Public Schools of North Carolina, 2013). The second school serves a portion of the small cities and towns within a county of about 67,000. Approximately 12% of students are minorities, and about 50% receive free or reduced lunch; in recent years, their school performed above the state average in both subject areas (Public Schools of North Carolina, 2013). Lastly, the third school is in a semi-rural area of approximately 15,000 residents, close to a larger metro area. It serves a population of about 65% minority students and 70%

of students receive free or reduced lunch; the school performed below the state average in both subject areas in 2012-13 (Public Schools of North Carolina, 2013). Overall, these schools were not completely struggling, but had areas for growth and improvement.

Along those lines, during the prior research project that these teachers participated in, data gathered using the Classroom Assessment Scoring System (CLASS) revealed that as a group, these teachers were providing an average learning environment for their children. For instance, on a seven-point scale, the average CLASS score from across the “instructional support” domain was close to the middle of the scale, at 3.21 (Pianta, La Paro, & Hamre, 2008; Oertwig & Gillanders, 2012, slide 2). The instructional support domain assesses how well a teacher provides concept development, quality feedback, and language modeling to their students (Pianta, La Paro, & Hamre, 2008). From across the “emotional support” domain, which includes assessments of positive climate, teacher sensitivity, and regard for student perspectives, the average score was a 4.2 (Pianta, La Paro, & Hamre, 2008; Oertwig & Gillanders, 2012, slide 5). While some of the individual teachers in this current study may have scored even higher than the group average from across their schools, this data indicates that participants were starting from at least a middle-of-the-road performance level.

Participant interactions. There were four opportunities to collect data from participants over the course of the study, although only 10 out of the 12 participants volunteered to participate in the final follow-up interview. First, all 12 teachers filled out a survey to expedite the uniform collection of basic background information about their school context and their experiences with evaluation (see *Appendix B*, Survey Questions). Demographic information collected from this survey revealed that these participants were white females with bachelor’s degrees or higher. Their experience levels ranged from two to

26 years in the classroom. One might divide the group into those teachers who were extremely experienced (four teachers with over 15 years of experience); those who were experienced (six teachers with five to 15 years of experience) and relatively new teachers (two teachers with one to four years of experience). Nine of the participants head up PreK-1st grade classrooms, while the other three teach in 2nd-3rd grade, causing the earliest elementary years to be particularly well represented.

During focus groups I conducted with teachers' colleagues at their schools, I used the survey data as a jumping off point for understanding how evaluation was being implemented and perceived within their setting. Because participants filled out the survey directly before focus group discussion began, there was a natural transition from thinking about evaluation individually to talking about it with others within their same context. Questions I posed to the group focused on general impressions of evaluation policy, as well as implementation factors that made evaluation either more or less effective within their school setting. In doing so, data from these focus groups helped capture whether state-level guidelines were being applied in the same way across different school settings, and if not, whether areas of difference were beneficial or detrimental. *Appendix C* (Focus Group Questions & Activities) contains more details about specific questions that were used to guide conversation.

During their focus group session, all 12 participants also engaged in an activity designed to compare evaluation with other drivers of teachers' practice. As part of this activity, participants considered a comprehensive list of potential influences on their teaching (see *Appendix C*, Focus Group Questions & Activities). This list was divided into person-, school-, and policy-based factors. Person-based influences included experiences that individual teachers commonly carry with them into the classroom, such as prior schooling,

teacher training, or family upbringing. School-based considerations were more contextual, such as leadership, student demographics, and professional development opportunities. Lastly, possible policy-based influences encompassed elements of the educational environment that went beyond specific school settings, such as statewide initiatives or the current educational climate within North Carolina. I specifically named teacher evaluation as a potential influence within this category, to generate data about its relative importance. Although this activity had a set structure to allow for consistency in data collection across groups, teachers were able to write in “other” influences not included in the generalized list if there was anything that they wanted to add. They were also encouraged to add notes under each potential influence that would provide a brief description of what it meant to them. After considering how these possibilities applied to their own circumstances, participants then ranked each factor’s relative importance.

During this activity, teachers often made comments to each other that were further captured within extended focus group discussion about their responses. They were asked to share what was the most important to them, and describe how evaluation fit into the bigger picture of the numerous factors that drove their pedagogical approaches. For the most part, teachers participated in this activity as planned. However, there were a few challenges in trying to capture the complexity of teachers’ lives. As participants worked through the provided list, rating potential positive influences on their practice was fairly straightforward. But then they were asked to indicate if some aspects of these factors could actually be considered less helpful, or negative, influences on their teaching. Several participants acknowledged these distinctions, but still found it easier to rate the importance of potential positive influences than to differentiate between less helpful elements of classroom life.

Although this did not majorly detract from the study, as some of these nuances were further explored during subsequent conversations, a few teachers were reluctant to name these pressures and commit them to paper. In addition, it was challenging for teachers to think about a scale of relative importance while simultaneously distilling the multiple facets of an influence's scope. Future work in this area would therefore require modification of the activity's structure to better capture those details beyond what was clarified within later focus group and interview discussion.

Follow-up interviews to triangulate data and gain further detail about their evaluation experiences and conceptions of good practice were conducted with 10 teachers out of the total group of 12 participants (two from the first school, four from the second school, and four from the third school). Ten of the 12 teachers participated in the final interview. During these interviews, teachers were asked about emerging themes from across previously collected data, either for clarification or to provide more specific examples of trends within their evaluation experiences (see *Appendix D*, Interview Questions). Teachers also elaborated on responses they had given during the focus group activity that asked them to rate influences that impacted their practice. I hoped to uncover areas of negotiated decision-making within semi-structured discussion of how each teacher approaches the classroom, where teachers described the interaction between other drivers of pedagogical decisions and meeting evaluation standards. I was also interested in the costs and benefits of providing teachers with structured guidelines for their work. At certain points, broader discussion about how evaluation might influence their practice was grounded in examining the exact guidelines within the NC Teacher Evaluation Process. However, because many teachers did not see details of these guidelines as particularly relevant to their daily practice, they tended

to talk about them as a whole, rather than speaking to individual aspects of evaluation mechanisms and procedures.

As participants further explained who they are as professionals, how they see their role, and how their visions of success compare with those provided by evaluation policy directives, these details began to form a rich narrative account of teachers' lives in the classroom and their workplace experiences. Once all four data points were synthesized, the data contributed to the construction of composite descriptions of how evaluation tended to unfold, while providing representative examples of evaluation experiences within particular school settings. Because of the continuous nature of this data collection process, which occurred over the course of five months, case study narratives utilizing teachers' stories were constructed in several stages. See *Appendix E* for more details about each element of data collection within Part Two and the role it played as analysis, organized according to the timeline for gathering this information from participant teachers.

Building case study examples. Although the experiences of a small number of teachers could never fully represent evaluation experiences for an entire state or district, their stories can be instructive to policy makers and school leadership personnel who are seeking to better understand what the evaluation process has been like for the teachers under their direction, and gather details that may be missing from summative policy analysis. Such narratives recognize the differences between teachers' views towards evaluation policy and implementation objectives, and make them explicit. As I asked participants to share how their sense of "teacherness" and conceptions of good practice might relate to performance standards, I obtained enough specifics to begin to more precisely pinpoint how they defined good teaching, along with significant influences on their approach to the classroom. This

helped me to engage in research that, “investigates a contemporary phenomenon in-depth and within its real life context, especially when the boundaries between phenomenon and context are not clearly evident” (Yin, 2009, p. 18). Moving from theory to practice, this study investigated what might be gained or lost in the journey from policy documents to lived reality.

Although the sample size was limited to create opportunities for involved storytelling, representational data extrapolated from the details of teachers’ experiences still related data to a bigger narrative that encompasses district, state, and national evaluation impacts. Within this study, the 12 participants shared how policy guidelines and workplace constructs interacted with their individual inclinations about how to lead their classrooms. Since participants were clustered in schools, they also described varied contexts for implementation, which could be documented and compared. This helped illuminate when experiences are typical of a particular school setting or district, rather than isolated to just one individual’s perspectives. Stake (1995) tells us that in order to build confidence around patterns of findings, “we must take more time, looking them over again and again, reflecting, triangulating, being skeptical about first impressions and simple meanings. For the evidence most critical to our assertions, we isolate those repetitions and correspondence tables most pertinent, challenging ourselves as to the adequacy of these data for that assertion” (p. 78). This does not necessarily mean that researchers need a great volume of data, but that the data that is collected should be a close enough observation of real life practices to provide evidence of common circumstances.

By combining an in-depth look at each participant’s experiences with an analysis of patterns and themes that resonated across the broader data set, my analysis attempted to

capture the nuanced effects of person- and school-based influences as they intersect with policy frameworks that are supposed to be applied in uniform fashion across multiple settings for teaching and learning. Because case study narratives can contain both individual retellings of personal experiences and composite impressions gathered from the entire body of contributing evidence, readers gain both the detail necessary to adequately represent the complexity of educational workplaces, and a certain level of generalization that can provide useful conclusions for leadership personnel to consider. Stake (1978) explains that, “As readers recognize essential similarities to cases of interest of them, they establish the basis for naturalistic generalizations” (p. 7). The level of detail within each case description then allows for distinctions between aspects of any one particular story that resonate with a broader base of constituents, as opposed to more trivial or unique experiences and attitudes that remain too specific for us to glean useful information from. As case studies give us more information to consider than we previously had access to, they can play a useful role in, “adding to existing experience and humanistic understanding” (Stake, 1978, p. 7), and grounding our views of policy’s impacts in rich descriptions of lived practices and implementation realities.

Analyzing data to answer essential questions. During data analysis, the central question of “*How, and to what extent, does teacher evaluation policy influence teachers’ conceptions of good practice and professional identity development?*” guided my initial organization of findings. Sub-questions about teacher attitudes towards evaluation, implementation factors, the role of context and school culture, and where evaluation fit within the range of influences on teachers’ professional identity development also helped shape the nature of my early analysis, which included *a priori*, preselected coding (Coffey &

Atkinson, 1996, p. 32). I determined these initial, broad codes by considering 1) my study's focus, 2) my analysis of themes and trends within evaluation policy from Part One of the study, and 3) my review of literature that pointed towards these categories of interest (see *Appendix F*).

The first preselected code was “attitudinal”. I chose to focus on teachers’ general attitudes towards evaluation and how they characterized evaluation’s purpose, because I wanted to compare this data to my discourse analysis of the official NC Teacher Evaluation Process completed in Part One. Previous research indicated that teachers’ attitudes towards a policy’s value and utility are central to implementation success (Gee, 2000; Lasky, 2005; Spillane, Reiser, & Reimer, 2002). The next code was “implementation”, which also stemmed from my central interest in how evaluation was being carried out across the three school settings in this study. Researchers have established that implementation factors significantly affect a policy’s impact on practice (Cohen et al., 2009; Fullan, 2001; Ozga, 2000). Finally, I coded for “influence”, given my interest in determining how evaluation’s impact on practice compared to other key drivers, as well as prior research indicating that a complex array of factors affect teachers’ professional identity development (Beijaard, Verloop, & Vermunt, 2000).

However, as I looked for trends and patterns from across teachers’ own descriptions of evaluation, their insights contributed to my emerging organization of data. For instance, although my research questions already focused on implementation, teachers themselves indicated that within that broader category, specific tension points might include issues with turnover, alignment, and time constraints. Teachers also suggested key factors that I had not initially considered, such as whether teachers were personally motivated to use evaluation

data to advance their instruction, as opposed to having a negative view towards receiving any kind of critical feedback. Peirce (1979) described this type of combined analysis structure as based on *abductive* reasoning. This mix of deductive and inductive approaches takes both previously established research questions and research directions suggested by subsequent analyses of fresh data into account. In doing so, it generates “a repeated interaction among existing ideas, former findings and observations, new observations, and new ideas” (Coffey & Atkinson, 1996, p. 156). Consequently, my initial coding schemes were refined and enhanced by a preliminary review of the data to determine the final themes from across teachers’ real-life experiences.

Finally, as I neared the end of my analysis, I specifically identified participants’ discourse about what good teaching looked like. In the context of this study, I defined their discourse as language used to build meaning and create a socially situated professional identity (Gee, 2004). I also saw their descriptions of ideal teaching behaviors and goals as indicative of ground-level dialogue about evaluation standards within the policy-to-practice dynamic. I directly asked teachers to provide more detailed examples of what good teaching looked like after my preliminary analysis of focus group data. Although some participants had already touched upon this topic during focus group sessions, my initial review indicated that more information was needed to better compare participants’ views to federal and state evaluation policy rhetoric. My decision to further analyze this discourse was also driven by returning to my primary research questions about conceptions of good practice, as well as sub-questions like “*How do evaluation rubric expectations relate to teachers’ personal views of their professional identity as educational professionals?*” By doing so, I could connect

the ways that teachers talked about ideal approaches to the classroom with official evaluation standards.

Examining teacher discourse about evaluation to determine how it related to dominant policy narratives allowed me to bring these teachers' perspectives to the forefront (Fairclough, 1995; Taylor, 2004; Young, 1999). It also illuminated how policy discourse is interpreted by teachers, who then decide how to incorporate directives like performance expectations into their existing conceptions of good practice (Gee, 2000; Taylor, 1994). Overall, using teachers' views towards evaluation to capture evaluation's relative influence within their lives helped me build narrative case examples based on their ground-level experiences (Stake, 1995; Yin, 2009). In addition, it brought forward attitudinal dimensions of real-life experiences that cannot be easily pre-determined or described within strictly empirical research. Blumer (1954) called these "sensitizing concepts", which he saw as important to understanding aspects of people's lives like "culture, institutions, social structure, mores, and personality" (p. 7). Within analysis of data, these sensitizing concepts "suggest directions along which to look" and contribute to "a general sense of what is relevant" (p. 7) within particular contexts.

Initial use of each data point. To accomplish my analysis goals, I drew from all four points of contact with teachers. First, I used survey data compiled by Qualtrics to describe general trends in attitudes towards the official NC Teacher Evaluation Process. Within the Qualtrics report, simple descriptive statistics were used to calculate the average number of responses within questions rating evaluation experiences and this policy's relative importance to teachers. Survey responses to sub-questions about general feelings towards evaluation, which parts of the evaluation process mattered (to both their school and to them

personally), the impact of evaluation on decisions about practice, and evaluation's relative contributions to their sense of what it meant to be a good teacher were the most relevant to determining if this group found evaluation to be both useful and personally meaningful. I also related these responses to my broad initial coding about teachers' attitudes towards evaluation. This established a basis for asking teachers about the reasoning behind their ratings during follow-up discussion.

Focus group data from discussions with three to five participants clustered in each school included further descriptions of implementation issues, which I later coded to identify areas where implementation was a factor in attitudes towards performance management. This data also helped me characterize each school context, allowing for comparisons between schools. For instance, I was able to identify areas where participants either spoke positively or negatively about specific components of the evaluation process and how they were being carried out within their school setting. Consequently, focus group data was my primary source of data during my initial, broad analysis of teachers' overall experiences and orientations, using the pre-selected codes described earlier.

Next, I used data from the focus group activity where individual teachers rated evaluation relative to other positive influences on their practice to pinpoint trends in what did matter to teachers. During my initial analysis, I compiled data from across individual participants to identify how many teachers total rated each potential factor along a scale of relative importance. In some cases, the details that individual teachers provided about how each category of influence related to their lives also provided material for follow-up. For instance, I focused my next round of coding on capturing more details about top-rated

influences from focus group discussion about the activity. I also decided to ask teachers about influences that were highly important to them during individual follow-up interviews.

Individual interview data was the final source for understanding evaluation experiences. Along with focus group discussion data, my examination of teachers' individual accounts of their evaluation experiences heavily contributed to narratives about trends in evaluation's relative influence. I used interview data to confirm my earlier analyses, flesh out descriptions of data that fell under *a priori* codes, and more concretely identify emerging trends within participants' experiences that might provide additional insight. As a result, coding interview data allowed me to further establish broader themes from across the group, and then determine how to break down this information into subtopics and organize it.

Coding across the data points. Coding across these data points first focused on broad evidence that I could use to indicate whether or not evaluation was having an impact. When I looked across all four data points to identify overarching statements about evaluation's effects, I looked for teacher responses that would address whether or not such policies were considered to be a substantive influence on teachers' conceptions of good practice and professional identity development. For instance, responses to survey questions that asked teachers to rate evaluation's relative impact on their practice were combined with data from the focus group activity that specifically asked teachers to rate evaluation's importance relative to other importance influences, along with more detailed descriptions of teachers' general attitudes towards evaluation that emerged from analysis of focus group and interview discussion. Within the attitudinal category, teachers' perspectives towards evaluation were broken into sub-categories like "pressure" or "performance". I also noted

demographic factors that appeared to play a consistent part in perceptions of evaluation's value. *Appendix F* contains additional information about the attitudinal sub-codes that emerged from looking at the entire scope of this study's data.

Then, I re-examined data that identified the specific reasons behind teachers' assessments of evaluation's relative influence on their lives. This area of focus again stemmed from the study's previous established sub-questions, which included, "*What factors affect evaluation implementation?*" As part of my initial a priori coding, I had already tracked broad mentions of implementation-related factors that affected teachers' perceptions of evaluation's value and utility within their school. However, after frequent mentions of implementation-related data were identified as a pattern across teachers' experiences, sub-codes emerged that were used to break down implementation into associated factors like leadership, turnover, uncertainty, and logistical constraints. I then tracked mentions of those specific factors across both focus group and interview data, to build a group narrative.

Throughout this process, I synthesized information from across multiple teachers and data points to describe common problematic trends in teachers' relationships with evaluation policy. However, I then wanted to compare data about more negative experiences to aspects of the evaluation process that did work well for teachers, along with identified person- and school-based drivers of positive implementation. Teachers self-identified these particular factors as being important when asked to speak about their views towards evaluation's impact. I used a combination of focus group, activity, and interview data to identify the most prominent other influences, aside from evaluation, on teachers' pedagogical choices.

Finally, my analysis of teachers' responses to the focus group activity that asked them to rate positive influences on their conceptions of practice led me to a further examination of

how teachers in this study would define good teaching. As a result, I identified data from follow-up interviews in response to my pre-determined questions about how participants would describe ideal teaching objectives and behaviors. Then, I organized their discourse about this topic according to the types of examples that teachers frequently emphasized throughout our conversations. Taken as a whole, this qualitative analysis of multiple data points led to the desired rich narratives that described evaluation implementation and its role in teachers' conceptions of good practice and professional identity development.

Limitations to scope of data. Throughout the process of constructing evaluation narratives, my own view towards local policies and knowledge of their potential impacts undoubtedly drove the scope of questioning and the direction that interviews took, as well as subsequent analysis of that data. Although participant responses played a major role in determining patterns and themes for further exploration, I also identified important moments and representative experiences, based on prior research and my own understanding of evaluation discourse, which led to this study's design. As a former kindergarten teacher, my experiences in the classroom may have further shaped my perceptions of whether evaluation matters to teachers, and what implementation factors might either promote or hinder potential benefits of performance appraisal practices. Since, "traditional research relationships are generally asymmetrical, with power disproportionately located on the side of the researcher" (Glesne, 2006, p. 138), these perceptions influenced data collection and analytical outcomes via their role in my study design and question selection. "We select our data, our research problems, what strikes us as interesting, and what to focus on and follow up with our informants" (Coffey & Atkinson, 1996, p. 141), and member checking can only partially counteract those inclinations.

In addition, while participants were selected because of their direct knowledge of recent historic shifts in how North Carolina rates teachers, the scope of their experiences was largely limited to their school's particular approach to carrying out evaluation. Because they had all taught within their school setting over the past few years of policy implementation, one may assume that they were well qualified to comment on the transition from minimal accountability guidelines to their state and district requiring specific expectations of teachers and mandating observation and rating procedures. However, their orientation towards this process is rooted within their own personal views towards the classroom and their unique school contexts, which may determine how evaluation is delivered and received within the day-to-day workings of their job setting.

My prior position working as a data collector for the research project that these teachers were previously involved in also played a role. Although I had minimal conversation with these teachers before their participation in my own study, I entered the study with a basic familiarity with their school environments that may have affected my perception of their workplace norms and values. Furthermore, their experience analyzing their instruction and utilizing detailed data about their classrooms via prior research involvement could have better prepared these teachers for conversations about their practice. It could also have given them an atypical perspective that differs from the standard North Carolina educator. Since no one identified themselves as a struggling teacher, and their schools were performing at least within reach of state achievement averages, that may also have skewed their perception towards the impact of receiving evaluation feedback.

In addition, because I chose to focus on a smaller sample size, favoring depth over breadth in terms of representative experiences, any extrapolation of patterns or themes within

interviews about evaluation discourse had to remain rooted in the contexts from which they were generated. I recognize that this methodology is designed to unearth details that tell a particular story about each teacher's relationship with policy directives and characterize North Carolina-based evaluation practices, rather than produce broad generalizations that apply to all teachers' experiences across the country. On the other hand, it is also important to recognize that even when narrowing the focus to a handful of teachers, it can be challenging to extract relevant details about one policy from a crowded landscape that includes a variety of factors that affect teachers' outlook towards evaluation. As Denzin & Lincoln (1998) describe, working with data from this type of qualitative research requires an artful analysis similar to that of a quilter who, "stitches, edits, and puts slices of reality together" (p. 5). Looking both within and across the case examples provided by this inquiry, related analysis needed to reconcile providing relevant information to the field at large with remaining true to context and preserving the authenticity of individual experiences.

Member checking. To boost the validity of findings, cases were compared from teacher to teacher and fleshed out through continued conversation with participants. Each participant was part of an initial survey, focus group, and focus group activity, and most provided additional insights and an opportunity for member checking during an in-person individual follow-up interview. Because these teachers were being asked to share their evaluation experiences throughout several touch points, I was able to introduce the potential impact of performance management requirements within a format that lent itself to ongoing reflection and reassessment of assumptions. This form of triangulation connected my initial coding for reoccurring ideas, descriptive patterns, and commonly used language to an opportunity for participants to verify my initial perceptions. Multiple points of contact

helped boost the probability of accuracy and understanding, by connecting my interpretations with participant insights about the extent to which evaluation influences their conceptions of good practice and professional identity development. Although self-reports of the effects of evaluation on practice were just a entryway into illuminating teachers' perceptions of their role, their characterization of influences on their practice, coupled with rich descriptions of the part that evaluation plays in determining their professional identity, still provided significant insights for policy makers, administrators, and other practitioners to consider.

CHAPTER FOUR

PART ONE: TRACING THE PATH OF TEACHER EVALUATION DISCOURSE AND EXAMINING POLICY RHETORIC ABOUT EVALUATION'S INTENDED PURPOSE

Federal and State-Level Policy Makers Assign High Importance to Rigorous Evaluation of Teacher Performance

Part One of this study traces the path from federal incentives for states to reform teacher evaluation to North Carolina's response to those federal policy recommendations. It sets the stage for understanding the policy context that local teachers were operating within when the current evaluation process was introduced to their schools. It also describes how good teaching performance is defined within the North Carolina Teacher Evaluation Process guidebook, which is currently being utilized across the state. The guidebook's directions include details about evaluation's stated purpose, structure, and intentions, as well as recommendations for implementation. Policy discourse analysis of such documents therefore establishes how coordinated federal and state reform efforts produced the evaluation standards and procedures that were presented to teachers.

Setting the Stage on the Federal Level

In 2010, the Race to the Top (RttT) competitive grant program was making immediate waves. Participating state education departments were jumping at the chance to gain additional federal funding if they emerged victorious. However, the U.S. Department of Education required states to prove that they were both willing to engage in reform and capable of making significant shifts in policy to align with federal priorities. The Obama

administration had already established several hallmarks of its educational agenda, such as increasing the number of charter schools, implementing the Common Core standards, bolstering data systems, and creating greater teacher accountability for student learning. As a result, the application structure that the administration created for the RttT competition reflected a desire to bring states on board with these reform approaches (U.S. Department of Education, 2009). Evidence of evaluation systems that explicitly tied measuring teacher performance to student achievement scores was given a relatively high point value within the “Great Teachers and Leaders” section, within a sub-section that required applicants to focus on “improving teacher and principal effectiveness based on performance” (U.S. Department of Education, 2010a, p. 87). In addition, key state stakeholders, such as union officials, had to sign off verifying that they supported these proposals in order for an application to even be considered. In order to be considered, federal authorities required that, “there must not be any legal, statutory, or regulatory barriers at the State level to linking data on student achievement...or student growth...to teachers and principals for the purposes of teacher and principal evaluation” (U.S. Department of Education, 2009, p. 4). A robust teacher evaluation system was described as having “multiple rating categories that take into account student growth (as defined in this notice) as a significant factor” (U.S. Department of Education, Race to the Top North Carolina Proposal, 2010a, p. 87). In addition, these components were supposed to be used to inform “compensating, promoting, and retaining teachers”, as well as decisions about granting “tenure and/or full certification” or “removing ineffective tenured and untenured teachers” (p. 87). From the outset, these policy structures were established as necessary aspects of a successful evaluation plan.

Pushing for alignment. Once initial applications were submitted, the Department of Education decided to try and push a broader range of states to become even more closely aligned with its evaluation vision. In a strategic move, the Department selected a limited number of Phase One winners, who would serve as models for other states jockeying for the remaining dollars promised in Phase Two. Delaware and Tennessee were chosen to receive enviable award amounts that would inspire others to strive for the same backing and recognition that would ease budget woes in a struggling economy. For instance, Tennessee, a state already known for utilizing value-added measures to calculate how well a teacher had been able to advance student progress, received \$500 million (U.S. Department of Education, 2010c). In explaining his decision to initially select only a few states to win significant dollars, Secretary of Education Arne Duncan was able to emphasize the types of programs that the administration wanted other states to emulate and provide further detail about the federal expectations that states needed to address in order for their reforms to be similarly funded, including incorporating value-added measures into teacher evaluation. This “carrot and stick” approach (Marshall & Gerstl-Pepin, 2005) therefore came with an obvious push for implementing policy updates modeled by these winning states, such as more rigorous teacher evaluation that would emphasize responsibility for student results, along with more severe consequences for poor teacher performance, such as reduced salary opportunities.

Pushing North Carolina to adapt more aggressive reforms. In comments from reviewers of the teacher evaluation section of its RttT Phase One application, North Carolina received feedback from multiple DOE reviewers that noted that while the state planned to do so, “student growth measures are not yet incorporated directly into the evaluations” and “there is very little that speaks to the inclusion of student growth data in the feedback” (U.S.

Department of Education, Race to the Top Technical Review 1, 2010b, p. 5). Another reviewer reiterated that while the state's current accountability system was designed to incorporate student growth, "it is not clear, however, how student achievement growth will be 'woven' throughout the relatively new Teacher Evaluation Process", which "remains vague about how big a factor student achievement growth will be in the rubrics" (U.S. Department of Education, Race to the Top Technical Review 2, 2010e, p. 4). While there was an overall positive tone to the comments, due to the headway North Carolina had already made in this area, additional concerns included that the evaluation plan had too many qualitative components, and not a clear enough path towards using student achievement data to inform personnel decisions. In the second round (or Phase Two) of the RttT competition, North Carolina was given the opportunity to amend their application to provide further evidence that the state was committed to "developing, rewarding, and retaining effective teachers" (U.S. Department of Education, 2010e, p. 1), and capable of carrying out state-wide accountability measures directly linking effectiveness to proof of student learning.

Federal Messages About Role of Teaching During RttT Competition

Because federal priorities surrounding more rigorous evaluation were broadly and publicly communicated throughout the RttT competition, the suggested adjustments to state policy outlined brought to the forefront by RttT were also reiterated by Secretary Arne Duncan and other federal education authorities who wanted to prompt states to act accordingly. This contributed to a national view of 1) how teacher performance should be measured (i.e., by implementing mechanisms such as value-added models), 2) the impacts the results of such evaluation measures should have on teachers' careers, and 3) the key student results that teachers should be held responsible for producing. Because evidence of

student learning was equated with producing higher test scores, this view asserts that an increase in these scores is what matters. In particular, the administration's efforts created a direct link between teacher inputs and student learning outcomes, and encouraged the use of value-added models to try and isolate the effects of this relationship so they could be measured more precisely and consistently.

To examine how the Secretary of Education showcased the importance of a particular type of teacher evaluation, while expressing his views towards the teaching role and its primary responsibilities, the following analysis of 11 speeches that Secretary Duncan made between Phase One and Phase Two of the RttT competition focuses on instances when Duncan most overtly contributed to evaluation-related discourse. Between March and August of 2010, states had the opportunity to respond to federal feedback and adjust their applications to better align them with Duncan's viewpoints and reform efforts. Shortly afterwards, Phase Two winners were announced, and North Carolina was awarded approximately \$400 million (U.S. Department of Education, 2010f, p. 1). The speeches that Duncan made during this time highlight his strategic approach to getting eventual grant winners like North Carolina on board with specific changes to reforms like sweeping evaluation policy mandates. Patterns, repetition, themes, key messages, and trends across all 11 speeches reveal repeated instances where evaluation policies or references to related reform are being framed as positive change. Duncan also set up a vision of good teaching versus bad, pushed for a federal role in consensus-building instead of local control of such measures, and pitted competition against non-differentiation, creating tension points within his discourse during this period. As a whole, these speeches can be synthesized to create an evaluation policy "story" designed to take the audience from bemoaning the broken state of

the current education system to celebrating the promise of RttT initiatives and championing Duncan's hopes for future improvements to performance management systems.

Establishing essential role of evaluation in identifying teacher quality. Across the 11 speeches analyzed, Duncan maintained a singular message that because good teachers needed to be recognized, evaluating their performance would help highlight their contributions relative to less successful performers. For instance, Duncan argued that, "We know that literally tens of thousands of teachers are doing a great job with students that are years behind - and helping them catch up - but the current system doesn't recognize or reward or learn from that teacher" (Duncan, 2010e, p. 1). On the flip side, Duncan suggested, evaluation would help school systems identify, "which [teachers] are simply not getting the job done" (Duncan, 2010j, p. 1). According to his administration, "too many teachers are unprepared when they enter the classroom..." (p. 1), and end up significantly impeding their students' progress. In some cases, Duncan submitted, we should have mechanisms in place to, "counsel out of the field those teachers just not suited to this challenging profession" (p. 1). Notably, he paid less attention to the middle ground of teachers, who might improve with support. Instead, Duncan chose to set up a dichotomy of good teachers versus poor ones, giving the impression that if you fall into the former camp, evaluation can only benefit your career.

Establishing outdated systems as part of the problem. Another consistent message across the speeches included attention to what Duncan referred to as an outdated feedback system, as he continuously argued that, "the system of evaluating, recognizing and rewarding teachers is broken" (Duncan, 2010g, p. 1). As Duncan spoke about why schools have historically done an inadequate job of measuring teacher performance, he described obsolete,

unhelpful evaluation systems that were doing little to progress teachers' development. He believed that more informal, cursory performance feedback is too lax, as principals and teachers are perpetually "making excuses for poor performance" (Duncan, 2010e, p. 1), and failing to hold themselves accountable for meeting basic learning objectives.

Duncan also linked these issues to the plight of effective educators who suffer from a lack of differentiation within the workplace and remain frustrated by a lack of official recognition that would set them apart from less-worthy peers and hold them up as role models. He positioned himself on the side of "teachers who feel that their good work goes unrecognized", while lamenting that weak evaluation also "ignores other teachers who would benefit from additional support" (Duncan, 2010j, p. 1). In his view, because loosely documented "gut feelings aren't good enough" (Duncan, 2010f, p. 1) ways to assess and track performance, a better solution must be quickly established to address such insufficiencies and help teachers from across the spectrum of instructional talent. Duncan wanted to rally educators around boosting the career status of excellent performers and holding their colleagues responsible for their lack of progress. By asking "educators to be more responsible for what happens inside the classroom", he appealed to their sense of personal responsibility and fairness while making it clear that accountability should be better tied to student learning results (Duncan, 2010e, p. 1). Notably, however, Duncan also presented evaluation feedback as a source of support for the perceived majority of teachers who would fall into the "good" category, rather than such feedback being punitive in nature. In doing so, he used this tactic to try to build consensus from educators who may have otherwise perceived evaluation updates as a threat to their job stability.

Providing exemplars of updated evaluation systems driven by RttT. A third message consistent across Duncan’s speeches is that states with strong teacher evaluation systems will be rewarded. For example, he said that, "The State of Tennessee has been collecting value-added data since 1992, but it wasn't until this year that Tennessee changed its law to allow its use in teacher evaluation and to identify the state's lowest-performing schools. That change in the law helped Tennessee win its Race To The Top grant” (Duncan, 2010l, p. 1). The implication was clear – make similar changes to your own state’s rules, and you, too, may find yourself the beneficiary of a federally funded windfall. According to Duncan, evaluation is again explained as a mechanism for recognition, since winning states realized that, "...we must recognize and reward the schools that show the largest growth and the teachers making the largest gains” (Duncan, 2010e, p. 1) by differentiating between them. Duncan also encouraged applicants who have already made steps in this direction to join the crowd by continuing along that trajectory. In the first round, “17 states reformed teacher evaluation systems by including – among other things – student achievement” (U.S. Department of Education, 2010i, p. 1). With that kind of buy-in surrounding “strong plans to create more meaningful teacher evaluation systems” (Duncan, 2010a, p. 1), more rigorous evaluation was positioned as foregone conclusion, rather than as a subject for debate.

Framing evaluation as a solution that will help close the achievement gap for struggling students. Another message across these speeches is that evaluation will “save” students from the persistence of low-quality instruction. Within that framework, using achievement data to measure teachers’ abilities to move their students forward was described as one way to ensure better outcomes. Ultimately, students would no longer be cheated by an ineffective educational system, as states would be "using this data to help...accelerate student

achievement” (Duncan, 2010a, p. 1). When explaining how children were being negatively affected by a lack of regulation of teaching effectiveness, Duncan made it difficult to argue the converse without appearing to be in favor of damaging students’ learning opportunities. For instance, when he asserted, “We can do a much better job...are we closing the achievement gap?” (Duncan, 2010g, p. 1), he created a rhetorical link between a common desire to improve the educational system to better serve all students and the specific aims of teacher evaluation policy. Underperforming children were particularly highlighted as the victims of a system that had failed to meet their needs, let down by adults who protected themselves from having to acknowledge their part in perpetuating ineffectiveness. As a result, Duncan argued, "When we develop fair ways to identify our best teachers, we can use that information to ensure our neediest students are being taught by the teachers they deserve" (Duncan, 2010g, p. 1). He simplified the potentially complex process of capturing a teacher’s impact by reducing it to a simple input/output relationship, where 1) good teachers are identified and 2) this data then becomes directly linked to higher achievement.

Firming up a defense of RttT strategies against criticism. A fifth message across Duncan’s speeches was that implementing stricter evaluation standards should not be a cause for great concern. To enforce this message, Duncan’s strategy was to acknowledge potential critiques of the federal evaluation agenda, but then minimize them. For instance, although the U.S. Department of Education wanted to align states around evaluation efforts, it was also sensitive to the historical precedent of states having local control over educational decisions. When considering states that had not yet committed to the more aggressive components of the U.S. Department of Education’s ideal evaluation plan, Duncan reassured applicants that states, districts, and schools would continue to have authority over how

evaluation data is utilized. While he outlined key components of a robust performance management system, he also said, “Local school districts must...decide for themselves how they want to share this information” (Duncan, 2010l, p. 1). Repeating the notion that the details of how to execute evaluation and utilize performance data will still be in states’ hands, he repeated, “That’s a local decision...”, “The local leadership can also choose...”, and again, “It’s a local decision...” (Duncan, 2010e, p. 1; 2010h, p. 1). Furthermore, because “community input is essential” (Duncan, 2010g, p. 1) to hashing out these details, “The administrators and unions need to lead the conversation; they also need to be thoughtful about how they engage the broader community” (Duncan, 2010l, p. 1). His framing of how performance management reforms would be carried out positioned evaluation implementation as an inclusive process steered by familiar faces, rather than being dictated from unknown authorities from above.

In addition, the most important part of these reassurances involved bringing the issue back to serving teachers’ needs – those key stakeholders that would be most directly affected by evaluation mechanisms. Duncan circled back to the idea that teachers will welcome critical assessments of their instruction, claiming that, “Teachers want - and need – this information. They want the feedback. And they want to get better” (Duncan, 2010l, p. 1). He also used humanizing anecdotes from cities where evaluation has already been publicly contested, asking us to:

Consider the words of two other teachers who ranked among L.A.’s lowest performers -- according to the analysis. Instead of being defensive, one of them was quoted saying: ‘Obviously what I need to do is to look at what I’m doing and take some steps to make sure something changes.’ He also advocated sharing the data with parents to keep him and his colleagues ‘on their toes a little bit more.’” He goes on to say that, “When another teacher saw her low score, she asked, ‘What do I need to do to bring my average up?’ Such responses, I believe, are real courage in action and I see that from teachers everywhere” (Duncan, 2010l, p. 1).

In telling this story, Duncan presented the worst-case scenario of being scored among the lowest within a district as being an opportunity for personal reflection and growth. He described teachers who take on this attitude using positive terms, and positioned himself as being on their side by expressing his admiration for their forward-thinking outlook. Once again, this shifted the focus from concerns about the evaluation process being used to monitor and punish weak teachers to viewing evaluation tools as a way to assist teachers in their pedagogical development.

Duncan's speeches also served to present more rigorous teacher evaluation as a non-problematic endeavor with few obstacles to effective implementation. Because evaluation reforms were presented as being both logical and imperative, for both teachers and children, Duncan led us towards viewing any potential downsides to evaluation as minor tradeoffs. Although he avoided saying that low-performing teachers could be fired, he acknowledged that, "It may mean making difficult decisions around staffing" (Duncan, 2010e, p. 1). He also asserted that teachers do not have to worry about attaching higher stakes to performance management ratings, as these ratings will be accurately calculated. Duncan explained, "We're also funding the creation of new and better tests that more accurately reflect how students and teachers are performing - and once we do that, we can do a much better job tracking student growth" (Duncan, 2010g, p. 1). The positive language here of "new" and "better" possibilities allowing educators to "accelerate" and improve student "growth" counteracts mentions of less-appealing personnel consequences (Duncan, 2010a-2010l). This pattern of positioning evaluation as a net positive can be traced across multiple speeches, making a powerful statement about the administration's objectives.

Presenting major evaluation reform as a widely agreed-upon necessity. Above all, these 11 speeches send the overarching message that major stakeholders have agreed that more rigorous teacher evaluation must move forward without hesitation. Within the analyzed speeches from this crucial time period, Duncan used the word “we” more than two dozen times just within the quotes that are explicitly focused on teacher evaluation. Although he may have been directly referring to the audience attending each speech, his sweeping use of “we” was often extended to include anyone with an interest in improving the public education system. This served to create agreement about the need to reform evaluation, while driving a shared focus around the administration’s intended outcomes. To build consensus around his priorities, Duncan’s mentions of evaluation begin with phrases like “Working together to improve evaluation systems”; “We need to...”; “Many forward thinking union leaders...agree with us...”; “Everyone agrees...”; “We must...”; and “Everyone knows....” (Duncan, 2010f; 2010k; 2010j; 2010l; 2010e).

Along those same lines, Duncan reiterated that his position is based on logical, indisputable facts. A more rigorous approach to evaluation, according to his administration, is “just common sense” (Duncan, 2010i, p. 1). He often begins talk of evaluation with framing devices like “We all know...” or “The truth is...” (Duncan, 2010k, p. 1; 2010j, p. 1). By repeatedly establishing that “Everyone agrees that teacher evaluation is broken” and “Everyone agrees that our current evaluation system is fundamentally broken” (Duncan, 2010i, p. 1; 2010l, p. 1), Duncan positioned evaluation as an obvious fix to repair an ineffective system. He also speaks in the present tense, once again establishing such reforms as a current reality that many states already seem to be participating in. For instance, Duncan asserted that, “Today, school district leaders and union leaders across the country are

working together to improve teacher evaluation systems” (Duncan, 2010f, p. 1) and that “This is going to start happening all over the country” (Duncan, 2010l, p. 1). Glossing over complication, Duncan laid out an action-oriented plan to guide states “as we move ahead...” (Duncan, 2010l, p. 1) with rating teachers according to evidence of student learning. Present-tense verbiage like “is” and “are” contributed to this perception that evaluation was already moving forward.

Utilizing action-oriented language. In order to maintain momentum around these reforms, Duncan consistently used language across his speeches indicating that he wanted to disrupt the status quo through an influx of accelerated change. He painted participation in the competition for RttT dollars as an energizing endeavor, which would ultimately result in positive gains forward for students. This was reflected in phrasing like “*cleared* that bar”, “*raise* their standards”, “*accelerate* student achievement”, “*accelerate* growth”, and “*drive* a cycle of continuous instructional improvement (Duncan, 2010e; 2010h; 2010c)” [emphases added]. Duncan also talked about the “value of competition” (Duncan, 2010c, p. 1) when illustrating how RttT has already spurred significant shifts in long-standing regulations that would have prevented some of the proposed changes from being considered in the past. He saw the process of participating in RttT as a victory in and of itself, and counteracted the less savory side of competition by claiming that instead of creating discord, it pushed a broad array of educators to, “ensure that our neediest students are being taught by the teachers they deserve” (Duncan, 2010g, p. 1) and provide a “world-class education to every child” (Duncan, 2010b, p. 1). Duncan characterized RttT as a “game-changer” (Duncan, 2010j; 2010l), even in its earliest phases, because of its unprecedented motivating power. Although states wanted to beat out other applicants, internally, they were “working together to

improve” (Duncan, 2010a, p. 1; U.S. Department of Education, 2010f, p. 1). Even looking across applicants, Duncan believed that because the competition application was structured to elicit responses aligned with federal guidelines, it created a unity of purpose and contributed to a national reform direction.

By framing the administration’s essential evaluation plan characteristics as part of a cohesive movement, Duncan established his case for immediate reform action and policy change. As a result, states like North Carolina emphasized that they had the attitudinal and structural capacity to begin implementing teacher effectiveness measures within updates to their original RttT applications. Because North Carolina policy makers and educational agencies were motivated to address the federal agenda in an effort to receive funding, they began describing the state’s evaluation process according to the parameters established by the U.S. Department of Education. State-level discourse about evaluation was therefore strongly tied to more widespread RttT efforts.

North Carolina’s Response to Feedback About Their Evaluation Plans

As North Carolina continued to seek RttT dollars after the state was not selected as an early winner in Phase One of the competition, the state quickly responded to the federal government’s call for further action in the evaluation policy arena. Because Duncan’s recommendations to applicants included strategies for improving teacher evaluation, the next step within this study was to connect his rhetoric with evidence of how state-level policymakers internalized those messages. In this section, evaluation discourse from North Carolina’s RttT Phase Two application is compared with comments from federal reviewers about the state’s increased willingness to carry out upgraded “teacher effectiveness” measures. This data highlights 1) North Carolina’s continued efforts to align with the

Department of Education's evaluation agenda, and 2) comments from federal reviewers confirming that North Carolina had acknowledged specific reform priorities like explicitly holding teachers accountable for student achievement results. Tracking feedback from the Department of Education within sections of North Carolina's RttT application that were focused on teacher evaluation shows how the state worked to prove that it was ready and willing to make desired adjustments to upgrade its performance management system.

Within the speeches analyzed for this study, Duncan's frequent mentions of the importance of updating state evaluation plans to include more rigorous components like linking teacher ratings to student achievement data were critically timed to influence Phase Two of the RttT competition. Because applicants who did not receive funding in Phase One had the opportunity to respond to feedback from the Department of Education at the next stage in the process, states like North Carolina were able to address reviewer concerns about their proposed evaluation measures. For instance, North Carolina's measures were initially perceived as being overly reliant on qualitative data, such as notes from classroom observations. Once they responded to this feedback, Phase Two reviewers examining North Carolina's proposed measures of teacher effectiveness noted that now, "The State clarified that each of the components included in the teacher and principal evaluation systems must be satisfied for the teacher or principal to be judged successful. So, in a sense the student growth measure is weighted 100% as are the other individual components" (U.S. Department of Education, 2010e, p. 7).

Although some federal reviewers were still fuzzy on the details of how these changes would be carried out and enforced by state officials, one reviewer who gave North Carolina additional points in this area reiterated that along with clarifying that student growth would

be included, “the performance management system...is not silent on the issue of multiple rating categories” (U.S. Department of Education, 2010e, p. 7), taking evaluation beyond a simple satisfactory or unsatisfactory designation. Another reviewer who also slightly raised North Carolina’s Phase One score for “developing evaluation systems” in Phase Two noted that “This rater’s concern about the possibility of excessively low weighting of student growth has been allayed” (U.S. Department of Education, 2010e, p. 6). When North Carolina then became a Phase Two winner, commitments to an enhanced teacher evaluation process led to the rapid refinement of existing performance standards in preparation for systemic implementation. While the state education agency had already begun putting some of these evaluation measures into place before securing federal support, North Carolina’s eventual RttT victory fast-tracked their adaptation.

The Further Development of North Carolina’s Current Teacher Evaluation Process

The current North Carolina Teacher Evaluation Process was formalized in 2009, originally drawing from existing professional standards to create rubrics for rating teacher performance to be applied consistently across the state. Its primary components measure how well teachers demonstrate leadership, establish classroom environments that support diverse learners, use content knowledge to develop curriculum, facilitate the instructional process, and reflect on their practice (Public Schools of North Carolina, 2012a). Teachers are supposed to connect feedback from observations with their own professional development goals. They are also asked to produce artifacts, such as student work, to serve as evidence of their effectiveness within each performance standard. In 2012, the process was updated to include value-added measures that utilize standardized assessments to track student achievement progress, “as calculated by the statewide growth model for educator

effectiveness” (p. 41), which was a direct response to RttT commitments. As a result, while observation rubrics that rate teachers on a continuum and support their ongoing professional development remain central to the evaluation process, this latest model of performance management also tries to make explicit links between teacher effort and measurable learning outputs.

Introducing the updated teacher evaluation process to teachers. Part of the goal of RttT was to create uniformity in teacher evaluation across states, so that data could be tracked and managed within a central system. This effectively holds states and districts accountable for consistently implementing required evaluation procedures throughout the schools in their area, prompting North Carolina to create support materials to train staff on how updated evaluation measures would work. To fulfill North Carolina’s RttT proposal to implement a comprehensive, statewide performance management plan that would measure teacher effectiveness over time, the state began holding a series of required teacher and administrator trainings that would familiarize personnel within each district with how evaluation’s many components should be utilized. The procedural elements of teacher evaluation were explained within presentations that the state Department of Public Instruction provided to North Carolina schools to share with their teachers. In some cases, teachers attended district-wide sessions to receive this information. Online modules explaining each standard have also been recently created. The goal was to create a uniform understanding of implementation expectations. However, it is notable that the two present-day online modules are estimated to take approximately twelve hours total to complete, which seems indicative of the complexity involved in trying to measure teaching performance (Public Schools of North Carolina, 2014). Although the North Carolina

Teacher Evaluation Process was built based on existing professional standards that were expanded and tailored to meet the requirements of an updated performance management system, utilizing them to formally rate teachers on this level was new to many local educators. As they were introduced to these performance standards, they were told that bringing themselves up to speed on how evaluation would operate was a required responsibility within this “new vision of teaching” (Public Schools of North Carolina, 2012a, p. 7). For instance, teachers were responsible for the timely completion of evaluation components like personal ratings, reflections, and professional development plans. When attending training about the new state evaluation measures, teachers were also provided with a copy of the completed guidelines, procedures, and performance rubrics that would be used to evaluate their progress.

Stated purpose of North Carolina evaluation. North Carolina’s detailed guide to new teacher evaluation standards and rating procedures includes descriptions of its purpose and utility. The introduction to the official NC Teacher Evaluation Process lists evaluation’s stated goals and objectives for “21st Century” teachers and outlines plans to enforce more rigorous performance standards, as established by RttT priorities. Particularly within the part of the manual entitled “Purposes of the Evaluation”, this additional data provides clear messages about evaluation’s intended impact. It also presents evaluation as a beneficial tool for teachers, administrators, and other members of the educational community like coaches, mentors, and teacher educators. The following analysis of this document’s contribution to policy discourse reveals that North Carolina has numerous lofty goals for utilizing performance management data, which are supposed to support each stage of a teacher’s career in a delicate balance between formative and summative components. The section of

the state-level guidebook that explicitly outlines the purpose of teacher evaluation lists both individual-level and administrative uses for formalized performance feedback. For instance, this information might guide teachers as they “reflect upon and improve their effectiveness” or “enhance the implementation of the approved curriculum” (Public Schools of North Carolina, 2012a, p. 5). It could become an important tool for coaches and mentors seeking to structure and ground conversation. However, descriptions of this process also reveal a strong desire to track classroom-level contributions towards school-wide achievement goals, as this “measure of individual performance” can help “focus the goals and objectives of schools and districts as they support, monitor, and evaluate teachers” and become “the basis for instructional improvement”, perhaps even informing “professional development” or “teacher training programs” that aim to prepare teachers to meet North Carolina’s expectations (p. 5). According to the state, “all of the instruments and processes are designed to encourage professional growth, to be flexible and fair to the persons being evaluated, and to serve as the foundation for the establishment of professional goals and identification of professional development needs” (p. 4). Yet, a concurrent focus on assessing “teacher performance” by producing evidence of “student learning” may create tension between valuing continual, formative feedback and acknowledging pressure to showcase more summative results.

Tensions between multi-faceted evaluation objectives. The tension between evaluation components intended to monitor teacher’s progress and those intended to provide professional development support was apparent throughout analysis of the NC Teacher Evaluation Process. Since North Carolina created this guide in response to RttT commitments to not only track teachers’ effects on student learning, but tie those results to

personnel decisions, its components reflect a tension between the need to emphasize accountability priorities and the desire to capture additional details about the entirety of a teacher's efforts. For instance, it is worth noting that while evaluation ratings remain relatively low-stakes in some local districts, others have plans to use the ratings as major factors in personnel decisions, tie them to bonuses, or publish teachers' scores publicly (Charlotte Observer, 2011; Ladd & Fiske, 2014; Wagner, 2013), just as the Department of Education suggested with RttT guidelines (U.S. Department of Education, 2009). This may be the natural next step for an initiative designed to create accountability by publicly classifying teachers as either "highly effective", "effective", or "in need of improvement" (Public Schools of North Carolina, 2012a, p. 6), and tying those categorizations to measures like student learning outcomes. Recent legislation has in fact asked districts to use evaluation ratings to select a top 25% of teachers to receive extended contracts, along with a small financial reward, effectively using a merit-based view towards teaching performance to eliminate established career status (Ladd & Fiske, 2014).

In addition, the vision of success promoted within the North Carolina Teacher Evaluation Process further connects teaching goals to other student-based results, such as graduating high school, becoming "globally competitive", and being prepared for "life in the 21st century" (Public Schools of North Carolina, 2012a, p. 4). While some of these goals can be more concretely defined and measured than others, they all involve holding teachers responsible for accelerating students forward. In fact, even mentions of staff-focused indicators of success direct teachers to collaborate around shared objectives to increase student achievement. The idea that collaboration should remain focused on end results further reveals local evaluation's dual purpose of wanting to reward teachers for behaviors

that go beyond boosting test scores, but then needing to link those behaviors back to those primary targets. These tensions echo those within Secretary of Education Duncan's speeches on the federal level, which tried to reassure teachers that evaluation would be more holistic and less punitive than they feared, while simultaneously asserting that producing student achievement results was paramount, or admitting that difficult staffing decisions would have to be made if teachers did not show adequate progress within their performance reviews.

Assigning high importance to evaluation's objectives. North Carolina's Teacher Evaluation Process also attempts to explain why teachers should care about evaluation and invest in the performance management process. This is reminiscent of Duncan's efforts to get teachers, along with traditionally opposed groups like unions, on board with significant updates to the evaluation system. In the section describing the origin of evaluation standards, state guidelines answer, "Why are these Standards important to you?" by explaining that they, "are the basis for teacher preparation, teacher evaluation, and professional development" as the teaching profession moves "into the 21st Century" (Public Schools of North Carolina, 2012a, p. 7). In other words, if teachers want to stay on top of their profession, they need to get on board with these expectations. Because many elements of the official evaluation process are described as "required" elements that teachers "shall" fulfill, they are not presented as debatable or optional. In fact, the word "required" is used 36 times throughout the document (p. 1-50). Furthermore, teachers' summative ratings will be used to categorize them according to how well they met "expected growth" (p. 41), confirming that value-added models will set the baseline for performance ratings. The manual detailing the NC Evaluation Process also explains that not meeting these standards will come with consequences. Language used to describe those consequences includes words like

“probationary” (p. 17) to explain the repercussions of being in need of improvement. This establishes the major purpose of evaluation as an accountability measure, even though, as alluded to earlier, it is also being promoted as a tool for professional growth. As a result, the Professional Development Plan (PDP) component of teacher evaluation requires that teachers list “evidence of progress” (p. 42) in areas where improvement is needed and then obtain signatures from administrators and mentors verifying the accuracy of that documentation. In addition, the PDP itself can be “monitored” or “directed” (p. 42). As Duncan had envisioned when trying to get states like North Carolina on board with more rigorous evaluation, teachers should be held responsible for meeting these standards. Consequently, North Carolina has reinforced the idea that these new policies are a serious undertaking, with weighty consequences for teachers who do not measure up to performance expectations. The state has also painted evaluation as more strictly tied to specific markers of ideal teaching behavior than prior expectations that were more open-ended.

Early messages about the teaching role and performance expectations. North Carolina’s evaluation efforts move beyond ideology into pedagogy by including explicit messages about what good teaching looks like and how it should be measured. Looking more closely at the presentation materials used within initial teacher training about evaluation shows how the increased role of such policies, as well as related teaching standards and rating scales used to assess performance, were being framed. These directives again echo RttT priorities driven by a federal-level desire to be able to more precisely track teacher performance and pinpoint areas of weakness or effectiveness, and provide details about how the state will be carrying out that vision. Consequently, the PowerPoint presentations created by the state of North Carolina to review evaluation procedures included explanations of how

teachers would be judged and rated. From the first introductory PowerPoint, teachers are told that they will have to demonstrate that they have met each standard in concrete, measureable ways. Teachers can take on this burden of proof by maintaining documentation of their practice and collecting artifacts of their teaching over the course of the year as “supportive evidence” (Public Schools of North Carolina, 2012b, slide 6). However, it is also up to their administrators to look for examples of ideal practices within the classrooms that they observe. When seeing a classroom in action, administrators will be noting “observable items”, such as “evidence of collaboration” (slide 9). A self-assessment component is also important, and links these ratings to professional growth. Yet, the utility of these reflections will in part depend on a combination of teacher motivation and clear communication from principals about what teachers can do to improve their instruction, ensuring that subsequent professional development plans are “specific, measurable, attainable, relevant, and time-bound” (Public Schools of North Carolina, 2012a, p. 25).

Similarly, the evaluation process does allow for conversation about performance during pre-, post- and summary-observation conferences, but career status teachers will not be observed as often, and therefore will not have as many required touch points. In addition, this presentation assumes that some of these conversations may be about “discrepancies” (Public Schools of North Carolina, 2012b, slide 19) and positions them as procedurally-driven opportunities to ensure that the provided evidence matches each rating, rather than rich discussions of a teacher’s progress. These end results could lead to further consequences, such as being placed on a “monitored” or “directed growth plan” (slide 24). As a result, a formative process that includes developmentally based ratings of teacher performance still has a clearly evaluative end-goal of being able to assign numerical or

categorical values to teachers' feedback. This means that the identified components of good teaching within the evaluation standards must connect to some sort of documentation.

Because federal reviewers of North Carolina's RttT application were wary of the use of nebulous, subjective qualitative data, the state's resulting performance management guidelines tried to at least tie more subjective assessments of a teacher's practice to physical proof of their efforts. This caused the state to emphasize measurable outputs within its rollout of evaluation changes.

Defining the teacher's role within a "new vision" for practice. Examining North Carolina's new evaluation standards revealed that the state wanted to make a distinction between outdated and fresh ways of approaching pedagogical responsibilities. Within the NC Teacher Evaluation Process, teachers are charged with proving that they are meeting recently upgraded performance standards. The unveiling of an new, standardized teacher evaluation process came with an updated conception of what teaching should look like in a new era where "simply covering the material" (Public Schools of North Carolina, 2012a, p. 4) would no longer be enough. Instead, teachers are being directed to build their capacity to effectively handle instructional tasks like developing skills in areas such as critical thinking, communication, collaboration and technology, which are more reflective of the needs of today's job market. Primary teaching tasks therefore include being a leader, making contributions to the school culture, making content "engaging, relevant, and meaningful", "uncovering solutions", infusing skills like problem solving and critical thinking, including 21st century content like "global awareness" within curriculum, integrating content with relevant ties to students' home communities, reflecting on practice, "authentically assessing" students, demonstrating the value of life long learning, and encouraging students to learn and

grow (Public Schools of North Carolina, 2012a, p. 4). This laundry list of responsibilities captures the complexity of teaching, which requires that one individual simultaneously serve as a leader, role model, cheerleader, content expert, instructor, and guide, while building relationships with students, families, and co-workers. It also reflects the idea that instructional approaches need to be updated in response to the job prospects within today's economic landscape, which values more globally aware, tech savvy, collaborative, and critically thinking citizens. These ideas can be directly compared with Duncan's stated desire to make students globally competitive and receive the same level of educational as their international peers, which were apparent throughout his reasoning for needing to fix broken elements of the current educational system, including lackluster evaluation. As a result, North Carolina wanted to send the message that outdated teaching approaches would no longer be acceptable in a forward-thinking era of higher expectations.

Creating a picture of ideal teaching behavior throughout observation rubrics. The descriptions of distinguished teaching behaviors within the NC Evaluation Process provide teachers with detailed examples of what ideal practice should look like, and what actions a model teacher might take to achieve that level of recognition. Looking across the rubrics used to measure each of the five observation-based standards, the major areas being measured are leadership, creating a classroom environment that supports diverse learners, content knowledge, instructional delivery, and reflection (Public Schools of North Carolina, 2012a, p. 8-12). When listing the demonstrable qualities of good teachers, the types of action verbs used within NC Teacher Evaluation Process suggest that the teaching role consists of a combination of 1) action-oriented behaviors and attitudes, 2) organizational or utilitarian tasks, and 3) social connections and accommodations made in service to the children in their

classroom. For instance, within descriptions of the most distinguished teaching behaviors, evaluation rubrics paint a picture of forward thinking teachers who use their strong planning abilities to design units of instruction that will connect their vision to the specific needs of their students. However, these standards also expand the teaching role to include modeling ethical behavior, interacting with colleagues, paying attention to professional development, and contributing to the field at large.

Good teachers harness a combination of skills to act on a clear vision for their specific students. Looking across the action words used within indicators of each standard, teachers are supposed to “lead”, “demonstrate leadership”, “advocate”, “demonstrate”, “communicate” and “link” within their instruction (Public Schools of North Carolina, 2012a, p. 8-12). These are all strong verbs that imply that the power to make a difference within education is at least partially within their control. For example, teachers are supposed to tackle the difficult challenge of addressing barriers to families’ involvement in their children’s education by being the type of professional who “conscientiously seeks solutions to overcome them” (Public Schools of North Carolina, 2012a, p. 24). Similarly, a good teacher can readily respond to changes in their classroom dynamics, as a problem-solver who “actively investigates” (p. 30) alternatives to standard instruction. This focus on action is particularly in line with federal evaluation discourse that asks teachers to drive accelerated progress towards big goals like closing the achievement gap and changing existing educational paradigms. But, teachers are also supposed to remain attuned to their particular students and respond to their unique needs as they “provide”, “treat”, “adapt”, “know”, “recognize”, “integrate” and “help” them (p. 8-12). Since teachers have been traditionally depicted as public servants, these responsibilities imply a level of deference that places the

good of the classroom at the forefront. Framing and organizing these tasks then requires teachers to “work”, “align”, “make”, “plan”, “use”, “utilize”, “analyze”, and “function” across a variety of utilitarian tasks that require advance planning and analysis of information (p. 8-12). If teachers think “systematically and critically” (p. 29), they will be able to make better-informed choices about how to move their practice forward. These tasks link to the use of information about student achievement to drive instruction that Duncan promoted as part of his general advocacy for accountability measures that mirror business-like, bottom-line objectives.

Good teachers make an impact that goes beyond the classroom. The inclusion of leadership skills and reflection about practice is notable in that these components expand a teacher’s role beyond lesson delivery and consider teachers’ development within their professional contexts for teaching and learning. For instance, under leadership, a distinguished teacher demonstrates classroom-based skills if he “encourages students to take responsibility for learning” or “empowers and encourages students to create and maintain a safe and supportive school and community environment” (Public Schools of North Carolina, 2012a, p. 21). However, to receive top ratings in this category, a teacher would have to make an impact outside of student-focused work to show that he “collaborates with colleagues to improve the quality of learning”, “assumes a leadership role in implementing school improvement plan”, “seeks opportunities to lead professional development and decision-making”, and “actively participates, promotes, and provides strong supporting evidence for implementation of initiatives to improve education” (p. 21-22). An ethical dimension is also introduced by requiring teachers to model “the tenets of the Code of Ethics for North Carolina Educators and the Standards for Professional Conduct” (p. 22). Although the NC

Teacher Evaluation Process guidebook provides some recommendations of how teachers can prove that they fulfilled this standard, suggested artifacts that would demonstrate “service on committees”, “Professional Learning Communities”, “formal and informal mentoring”, “membership in professional organizations”, or “National Board Certification” seem to apply more directly to some elements of this standard than others (p. 23). These types of indicators are more holistic than the U.S. Department of Education’s primary objectives, but link to the idea of good teachers becoming models for others.

Good teachers use data and research findings to make instructional decisions. There is also a repeated emphasis on good teachers being able to justify instructional choices, using data, research findings, or other evidence of promising practices to guide pedagogy. Under the standard describing how teachers should “facilitate learning for their students”, teachers are advised to “keep abreast of evolving research about student learning” (Public Schools of North Carolina, 2012a, p. 27). They should also stay informed about “emerging research areas and new and innovative materials” (p. 27) that can be incorporated into lessons. Collecting data about student progress is integral to using “multiple indicators to...monitor and evaluate student progress and to inform instruction” (p. 28). The concept of teacher as monitor of student achievement is reinforced throughout the developmental continuum for that measure. Regular assessment is a major component of effective teaching, as teachers should “evaluate student progress and growth as they strive to eliminate achievement gaps” (p. 29). In fact, teachers can collect artifacts that will serve as evidence that they have utilized evidence, like lesson plans or professional development materials. Use of data is also mentioned in rubrics for other standards; for example, under the leadership standard, the first sub-standard says that, “Using a variety of data sources, they [teachers] organize, plan,

and set goals that meet the needs of the individual student and the class. Teachers use various types of assessment data during the school year to evaluate student progress and to make adjustments to the teaching and learning process” (p. 21). This pattern of emphasis establishes these behaviors as cornerstones of the profession that teachers should be rated on in numerous ways over the course of their evaluations.

Equating good teaching with producing student achievement. The only standard that has not been fully implemented is Standard Six, which was added to the original five to incorporate data from value-added models that track student growth and aim to isolate a teachers’ effect on a student’s academic progress. According to the official guide, “A teacher’s rating on the sixth standard is determined by a student growth value as calculated by the statewide growth model for educator effectiveness” (Public Schools of North Carolina, 2012a, p. 41). Plans were in place to use the “growth value for the entire school” (p. 41) for non-tested grades. However, it is not entirely clear how this is being implemented in the present day. It is implied that the ratings for this particular standard will be providing by state-level authorities, rather than calculated within each school by administrators. Here, the emphasis is on tracking “measurable progress” (p. 41) that can be objectively compiled. Because North Carolina had agreed to integrate quantitative measures of student learning into their evaluation process, this standard addresses that RttT promise. In doing so, it equates good teaching with being able boost student growth and produce higher achievement on standardized measures for each grade level. Although when looking across all of the required standards, teachers clearly must bring a multi-faceted skill set into the classroom, the underlying message is that instructional inputs should result in concrete, quantifiable learning outputs.

Overall, North Carolina's response to federal pressures to make evaluation more concretely tied to evidence of student learning was well aligned with the RttT agenda. Evaluation was locally framed as an important, high-stakes endeavor that would be placed at the forefront of new state policy changes. The state Department of Public Instruction facilitated the design of the evaluation process and set implementation standards to be immediately applied across districts. Districts then worked with the state to quickly train administrators and teachers on using the new process. District-level personnel have some oversight in terms of ensuring that administrators complete evaluations properly and submit ratings data. However, it is mainly up to each school site to carry out evaluation in the day-to-day and to keep teachers informed about how feedback mechanisms will function. Fulfilling these responsibilities is strictly required within the guidelines of the official evaluation process (Public Schools of North Carolina, 2012a).

Therefore, the official NC Teacher Evaluation Process was first introduced to teachers as a set of uniformly required standards and procedures that would focus on the results of teacher efforts and enforce increased accountability for performance. Along those lines, the purposes of evaluation highlighted its use as a tool to monitor teachers and ensure that they were fulfilling their job responsibilities, which could be further proven through documentation to be used as evidence of adequate or better efforts. Within descriptions of expected teaching behaviors, the upgraded, modern version of the teaching role is tied to levels of specific behavior that can be measured and rated. Good teachers are defined as actively making informed choices about their pedagogy, as part of data-driven instruction. In addition, teaching responsibilities are expanded to include a teacher's role within the broader professional community. Teachers whose students perform well on standardized exams are

seen as particularly strong exemplars for others. In the next part of this study, teachers' interpretations of these standards, as well as their views of what good teaching practice consists of, will be compared to these baseline requirements.

CHAPTER FIVE

PART TWO: ANALYZING TRENDS WITHIN TEACHERS' EVALUATION EXPERIENCES

“Who Has a Relationship with a Rubric?”: Current Teacher Evaluation Policy Does Not “Really” Matter to Teachers

Part One of this study established the priorities that led states like North Carolina to adopt more rigorous teacher evaluation measures. Analysis of recently introduced North Carolina evaluation policies and procedures provided context for what was identified as significant within the state’s descriptions of why and how teaching performance should be rated. In Part Two, I juxtapose this policy context with ground-level interpretations of evaluation’s purpose and impact. Part Two includes first-hand accounts of evaluation implementation and describes evaluation’s role in teachers’ lives, especially in regards to their evolving perceptions of what ideal practice looks like. In addition, this section explores how contextual factors like school setting might affect evaluation experiences. Analysis of the data about teachers’ experiences with policy implementation helps answer, *“How, and to what extent, does teacher evaluation policy influence teachers’ conceptions of good practice and professional identity development?”*

Teachers’ Impressions of the Official North Carolina Teacher Evaluation Process

While the North Carolina Professional Teaching Standards that formed the original basis for the state’s updated evaluation measures had already been adapted in 2007 (Public Schools of North Carolina, 2012a), using them to create observation rubrics that would drive teacher performance ratings took these performance expectations to a new level. The advent

of RttT funding earmarked for this purpose spurred North Carolina to require use of these rubrics within the official NC Teacher Evaluation Process and train administrators and teachers on how this process would unfold, primarily between 2010-2011. For most local teachers, 2011 was when the state's timeline for full implementation required their district to carry out the official process throughout all grade levels. District personnel played a role in the introduction of updated evaluation procedures and provided administrative supervision. However, the daily work of carrying out those procedures was largely handed off to administrators within individual schools.

Currently, local implementation leaves room for principals to determine exactly when teachers will be observed, and by whom, as well as details like how much documentation of practice through the suggested collection of artifacts aligned with each standard they will ask teachers to produce. However, fulfilling their evaluation responsibilities requires a mixture of context-specific decision-making and compliance with set guidelines. All principals must upload teacher performance ratings to an online system so districts, as well as the North Carolina Department of Public Instruction, can collect and review them. Before doing so, they must also ensure that teachers receive training about evaluation updates, conduct a self-assessment, attend pre- and post-observation conferences, and get observed anywhere from one to three times a year. While the number of observations depends on their career status, teachers receiving multiple observations are also supposed to have a summative meeting with their principal at the end of the ratings cycle. Feedback from evaluation should then be reflected in an official professional development plan. In 2012, the state integrated value-added measures into the existing standards, at least in part due to RttT promises to explicitly link teacher performance to evidence of student learning. For teachers who do not already

have standardized testing on their grade level, school-wide growth values are supposed to be utilized until additional assessments can be developed (Public Schools of North Carolina, 2012a). Teachers in this study were asked to comment on the role of each of these specific components of the evaluation process in their teaching lives, although observation-driven feedback was at the center of their current experiences.

Overall failure of evaluation to inspire significant change. For the majority of teachers in this study, the current NC Teacher Evaluation Process did *not* have a significant influence on their perceptions of good practice or their ongoing professional identity development. Although participants accepted that evaluation was a necessary part of the current education landscape, it typically fell short of holding great value. This lack of value sharply contrasted with the intended value touted by Duncan, for example. As one participant described when explaining her attitude towards participating in pre- and post-observation conferences, “Yes, I do it ‘cause that’s what expected of me, and I want feedback, but sometimes I feel like, it’s just...they’re doing what they gotta do, and I’m doing what I gotta do, but it’s not really going to change things that much”¹. In her mind, evaluation was “fine”, but lacked consistently visible drivers. Even teachers who reported enjoying reflective elements of the evaluation process or felt relatively positive about the feedback they had received as part of the evaluation process still failed to report that it had impacted their practice in the same way as more central influences. One teacher explained, “It has the potential for being more helpful, but it doesn’t always pan out that way”. As a result, while their attitudes towards performance management ranged from tolerance to optimism, evaluation did not appear to “really” matter to them, or affect many of their daily

¹ All teacher quotes within this chapter come from participant data gathered during the focus group and interview sessions described in the methods of this study.

decisions about curriculum and instruction. Instead, performance ratings were seen as “a number on a piece of paper” that was unlikely to become something that teachers would “focus on” or see as transformative. As one teacher admitted when considering whether evaluation had influenced her instruction, “Do you want the honest answer? Then, no.” Others within her school agreed that while they appreciated their administration’s efforts to become familiar with their teaching, the official process itself was not going to make a true difference.

While these teachers recognized that contextual variables might drive varied effectiveness of evaluation implementation, their overall outlook towards evaluation tended to fall towards the middle of the spectrum, with only a few taking an overtly positive view towards the potential for evaluation to drive shifts in instruction. For instance, 11 of 12 respondents said that historically, the overall quality of feedback received after being formally observed was average to excellent, with the group split between those two options. Similarly, eight teachers reported that evaluation standards “somewhat” contributed to their sense of what it means to be a good teacher, with the remaining respondents divided between either more or less positive answers. Yet, participants’ beliefs about evaluation’s utility as a direct influence on classroom practice was particularly mixed. When teachers were then asked if evaluation contributed to decisions they made about how to approach their practice, seven of the 12 teachers said evaluation somewhat contributed and two teachers admitted that it hardly contributed at all. There were also indications that evaluation was taking on an *imposed* importance, since 100% of respondents thought that overall, evaluation was “highly important” to their school, and identified end-ratings and test scores as having major value to

their administrators; yet, only half said that their evaluation results were highly important to them personally.

In addition, the handful of teachers who rated evaluation more positively within specific questions about its impact were all from the same school, suggesting that implementation with certain school settings could result in better attitudes towards evaluation. However, when that sub-group elaborated about their experiences, tensions within the evaluation process emerged that continued to limit its reach. This contributed to a relatively low assessment of evaluation's importance (as compared to other person and school-based influences on practice) from across participants, since teacher evaluation again fell towards the middle of the scale during the focus group activity that specifically asked them to rate evaluation relative to other influences on practice. Data from this activity showed that seven teachers in this study rated teacher evaluation as only being "somewhat important", while four thought it was "not very important", echoing trends in earlier survey responses to similar questions.

Evaluation is not a driver of teaching practice. On one hand, teachers characterized the standards they were being held to as fairly unobjectionable – for example, "all things that good teachers would do anyway". They also appreciated that the current evaluation process had some "holistic" elements, since it included standards on leadership, reflection, and classroom environment. However, evaluation feedback addressing these standards was rarely inspirational enough to be "life changing" or even stimulating. One participant explained that although she cared about her ratings, and found the standards to be "well-intentioned", they weren't necessarily as much of a "motivator" as other key influences on her teaching. Multiple participants expressed that evaluation "isn't going to change what

I do”, or “isn’t going to stop me from doing what I do in my class”. They explained that evaluation feedback only occasionally offered “a different perspective”, or specific thoughts about instructional approaches and teachers’ role within their schools for them to consider. For instance, a teacher who had made a shift to a new grade level identified feedback from her administration as relevant to that adjustment. Similarly, out of three observations she had so far this year, a colleague remembered one observer that had helped her integrate phonics with guided reading. But, in general, a handful of small ideas or prompts from evaluators were not enough to cause a shift in pedagogical trajectory, particularly because this group of teachers was in many ways already established in their practice, with substantial early childhood and elementary experience. As a well-credentialed teacher distinguishing between key instructional drivers and the influence of feedback guided by evaluation rubrics explained, “It just doesn’t *motivate* me”. Because she was a teacher who tended to look towards the big picture, it did not provide a strong enough “vision” for moving her professional development forward. Others similarly expressed that high-stakes evaluation was “just part of life”, but would never gain high personal value, despite the possibility of tying their evaluation ratings to bonus eligibility or career status. One teacher said, “Even the money, more of it is even better. But I have to look in her face [gestures to colleague in focus group] and my students’ face every day. And that’s what really matters.” Another colleague in her focus group supported the idea that while evaluation might contribute to their practice on the surface level, “Change me - it’s not going to, is the thing”. In a different school, teachers reiterated this theme, saying that when they reviewed their ratings, “It’s good to see, but it’s not going to change anything.”

However, while evaluation was clearly not transformative for this group of teachers, it was also not considered to be detrimental. Instead, the official performance management process fell into a middle space that simply fell short of becoming truly inspiring or meaningful. Although no one shared extensive complaints about evaluation over the course of data collection, teachers instead tended to see the formal process as merely ineffective or minimally impactful. As one teacher put it, the utility of evaluation was random, since, “the feedback, I don’t necessarily use it. I think it’s just given because they have to. They have to give you something, you know. Sometimes it’s helpful.” Looking across teachers’ experiences, those who gave examples of components of the evaluation process that did work relatively well for them, such as being able to connect reflection to a professional development plan or getting encouragement about their teaching, were describing exceptions to the norm, or bright spots within tepid policy implementation. In most cases, the “good intentions” of the NC Teacher Evaluation Process were not themselves enough to result in sustained, recognizable links between performance feedback and changes in teachers’ reported views towards their classroom responsibilities.

Specific reasons why evaluation failed to live up to its full potential. One primary goal of the NC Teacher Evaluation Process was to enhance teachers’ ongoing development while building their instructional capacity. Unfortunately, according to the participants in this study, it often failed to live up to that goal. In addition to the fact that teachers placed greater value on other drivers of their practice, the majority of reasons why the current evaluation process did not “really” influence teachers’ conceptions of good practice and professional identity development were related to implementation. However, teachers also identified flaws related to the uncertainty surrounding evaluation’s future use, its incomplete

applicability to early childhood and early elementary classrooms given this particular teacher sample, and the administrative nature of evaluation paperwork. Repeated patterns across their experiences indicated that trying to establish evaluation as an important influence on professional growth and instructional improvement was a complex and difficult endeavor.

Implementation factors: the difficult task of effectively carrying out the evaluation process. Across all three schools in this study, teachers gave examples of how implementation factors impeded evaluation's potential influence. Because local implementation hinged on a school's principal, in addition to the delegated contributions of other administrators like assistant principals or instructional coaches, a great deal was riding on how school leadership personnel decided to carry out evaluation within their school context. Teachers described the path from policy guidelines to implementation realities as similar to "a game of telephone", where the original message might not match the end result. One explained, "You know how one person says one thing over here, and then another person says something else a little different, and then by the time it gets to you...?" A teacher from another school had similar views towards policies getting jumbled in translation. As she described, despite extensive training from the state, "The conversations that you have here don't always trickle down to the intent that it's... it's like this person went to the meeting that was supposed to relay it to this person and the intentions aren't...and it is frustrating." As they traced the winding nature of evaluation's trajectory, participants articulated how those who held administrative power at the state, district, or school level ultimately determined how evaluation was presented. Furthermore, they raised structural issues within their schools, as well as flaws within the evaluation measures themselves, that led to implementation challenges.

Varied types of relationships with school leaders as a barrier to consistently effective implementation. First and foremost, teachers reported mixed relationships with leadership over the course of their career, and emphasized that the person providing them with instructional feedback heavily shaped the quality of their evaluation experiences. Sixty-seven percent of participants rated leadership as either “important” or “very important” to their overall practice, suggesting that even one key individual could impact, for better or for worse, a teacher’s growth as an instructor. As one might expect, less trust and connectivity with a principal meant that their evaluation feedback would have less influence. One teacher explained that right off that bat, when asked her opinion of the current evaluation process, the central role of the principal came to mind. She said:

I think for me it always ends up being the relationship you have with that administrator and, umm, the type of leader that they are. Administrators that valued you as a teacher and ones that didn’t value you as a teacher, so the feedback that you got was going to be different, just like with your students, you have that relationship with them, you give them that solid feedback. So it’s kind of hit or miss depending on who the facilitator is, who the leader is.

This was indicative of a universal desire for principals to provide the same supportive environment for teachers that they would expect teachers to provide for their students. As part of this desire, teachers wanted their leadership team to take the time to get to know them fully as a teacher, so they could feel as if their administration both understood their pedagogical approach and cared about their progress. Otherwise, “You have some administrators that get in there and know who you are, and then you have some that could care less. You know they’ll see for your evaluation and see you for your scores and that’s it.” If leadership did not make an effort to personalize their observations and engage teachers in individualized conversation about their instruction, teachers quickly became less receptive to evaluation as a tool for professional development. Therefore, qualifying language like

“depends” was repeatedly used by multiple teachers to qualify the direction that evaluation could take in the hands of either a well-liked, relationship-oriented leader, or a less-desirable leader who was solely focused on the bottom line. Within one focus group, a teacher suggested that evaluation, “completely depends on the quality of that leader”, and another agreed, “completely, ‘cause you got your leaders and ones that are into growth and ones who build relationships”, along with other types of principals whose attitudes were unengaged, and would “lay back, you just don’t care”. In those cases, style and “personality” played a role in performance management implementation.

In addition, some teachers had markedly negative past experiences with dictatorial principals that had made them aware of the difference administrators could make in their professional lives. In previous situations where they felt that they could not trust “crappy” principals to have their best interests at heart, “You know that your evaluation and your reflection about what you do is more important than that person who came into your room for fifteen minutes who doesn’t truly understand what’s going on all the time. Even though they’re your leader, sometimes they don’t have the picture.” In those instances, the impact of evaluation feedback had been limited by the perception that the person providing them with ratings did not have a full view of their practice. As one teacher currently in a more amenable situation speculated, “They’re the deciding factor, the ones that come and observe you. But it depends on their personality type and how they see you. If you have a real strict dictator and you’re not doing exactly right, that could be a tough evaluation”. As a result, teachers’ own assessment of their progress was deemed to be more worthwhile than commentary from an administrator who made a quick visit to their classroom for a formal observation a few times a year. Teachers were skeptical that such an individual could

completely understand the complexity of all their interactions with children and accurately determine appropriate goals for their classrooms. Ideally, they instead wanted administrators who were conducting evaluations to provide a combination of targeted, relevant advice and warm encouragement.

As a participant explained, if principals have to point out areas of improvement, “They should be our instructional leader and know that next step.” She felt that in order to maintain positive connections with teachers that would maximize productivity, administrators should extend the foundational relationship that had been established between the two parties over time, and use it to build trust in their specific pedagogical guidance. Teachers thought that, “...it would make a difference if you have a relationship with your administration, because if you have a good one, that would make a big difference,” and similarly, “a strong and caring leader makes a huge difference”. This foundation was essential to their ability to take evaluation feedback seriously and see it as a means to an end, rather than a chore or an empty exercise. Otherwise, teachers would say, “Ok, this is my leader and they’re telling me that it’s not the best, and depending on the leader that you get, sometimes your ratings are not all that wonderful, but they can’t tell you what else to do. They don’t give you that feedback that helps you grow.” Because the only teachers that rated leadership in the highest category of importance when rating positive school-based influences on their practice were from the same school, this suggested that teachers within other contexts for teaching and learning still found leadership central, but not quite as uniformly effective. Unsurprisingly, the school that assigned strong importance to leadership had generally positive feelings about the supportive nature of their entire school culture, even though they were still adjusting to a new principal. While all participants were sensitive to

leaders' role in their lives, especially if they had also recently experienced principal turnover, they could not always count on receiving deep emotional backing or specific pedagogical guidance within their workplace.

Frequent principal turnover as a barrier to consistently effective implementation.

Because a long-standing foundation of trust was seen as essential to productive conversations about areas of improvement, frequent administrative turnover was another issue that clouded teachers' feelings about evaluation. The teachers in this study reported that, historically, in their schools, some form of evaluative observation had been conducted for various purposes by principals, other administrators/assistant principals, instructional coaches, district personnel, researchers, and their peers. This meant that even when their primary principal had remained the same since evaluation was first implemented, which only four out of the 12 participants could claim, chances were that at least one of the people tasked with observing classrooms would contain an employee who was new to that role. For instance, at one school where other teachers played a hand in providing their colleagues with feedback, observation pairings would change from year to year. Participants from across the study were often excited about the possibility of new administrators making positive changes, and valued receiving a variety of perspectives towards their instruction. However, they also felt that it was difficult to maximize evaluation's impact when the nature of its contribution to their professional development trajectory was subject to constant shifts in personnel. As one teacher at a school with recent turnover explained, "I think I've been in the business long enough to know that evaluations change dependent upon who is doing them. And I think hearing other people's feedback is good, but I think that self-evaluation goes a lot further, because we can have a principal today who tells me I'm fantastic and wonderful in every way

possible, and someone else can come in tomorrow and see something completely different." This lack of continuity prevented evaluation from feeling connected to sustained relationships that teachers could rely on with confidence.

In contrast, although teachers at the sole school where the principal had remained the same in recent years were ambivalent about the evaluation process as a whole, they did partially credit that long-term relationship as driving relatively more positive aspects of receiving performance feedback. Although they were not at the site where teachers reported the strongest overall school culture and implementation practices, at least conversations with their principal followed predictable interaction patterns. Within that context, they reiterated that they did not have to worry as much about their observations, because "she knows me", and "because of the type of relationship that I have with our principal here, she knows very well what I do...it's a part of that trust piece". In different instances, teachers reported a shift in their views towards the evaluation process after a new administrator handled post-observations conferences differently than her predecessor. One teacher who had previously felt in sync with her principal said that in more recent meetings she "only had about five minutes" to review her ratings before a rushed conversation. This experience made her feel as if her evaluation was less personalized, and therefore less useful, than before. She felt personally devalued, but also felt as if the process itself had become less likely to take on real meaning. Others feared that their current comfort could be short-lived, since, "You never know when you're going to get a crazy." A teacher who currently had a decent working relationship with her principal still described the frustration that could result when "you've had a good relationship with one administrator, and have scored very high, and they see these things in your practice really without you having to explain yourself" and then all of a

sudden, “someone else comes in who has a different filter, a different background, a different personality and they don't necessarily see the same things”, causing teachers to have to justify practices that had previously been lauded. In the best case scenario, well-received administrative changes could result in a teacher immediately feeling “comfortable with her even though I didn’t know her”, while at other times, being observed by an administrator who they had not formed a “connection” with could create feelings of uncertainty.

Uncertainty about the use of the evaluation process within their schools as a barrier to consistently effective implementation. Teachers who had significant experience in the school system, like most of the teachers in this study, were both wary and weary of a host of constant changes, which went beyond issues with turnover. In their view, the nature of evaluation, like many other mandates, would probably shift over time to accommodate evolving implementation nuances and changes in legislation. On an internal level, changes in their school building led to an additional lack of clarity around how the evaluation process would continue to unfold and how it would affect their professional status. As teachers filled out a survey asking them to identify the components of the evaluation process that were in place in their schools, as well as important to their administrators, they often paused to ask each other for clarification about how evaluation was actually going to be carried out. For instance, when trying to determine if value-added models of measuring student growth were currently in place within lower elementary grade levels, one teacher asked, “But does it impact us right now?” Some of her colleagues replied, “Not yet”, or “Not yet for us” but others chimed in that “It’s impacting us”. At another school, the conversation had a similar direction, as one teacher noted, “This year they are currently using the student achievement, right? I know it wasn't last year but now it is...” Her colleagues replied, “Maybe, maybe

not”, and “Not sure”. While 100% of participants reported that test scores were being used as an evaluation tool, they were not certain whether this meant that formative assessments, periodic benchmarks, or some kind of new value-added measure would be taken into account moving forward. One teacher asked, for instance, “Do you get test scores?”, while the others within her focus group replied, “I guess they kind of look...”, and “They look at them, they see ‘em, they’re starting to...” Along those lines, while 100% of participants were confident that their schools were following the NC Teacher Evaluation Process closely, 50% answered “not yet, but will in the future” when asked if value-added scores currently factored into their evaluation ratings, while the other 50% said that they were “not sure”. There was a sense of some planned components of the state evaluation being on the horizon, although they had not become an established reality across their school building. This led to uncertainty about what the future implementation of evaluation would encompass, as several teachers asked what the term “value-added scores” meant and how it would be defined on the PreK-3rd grade level. In general, even though teachers at all three schools in this study recognized ways that their principals were trying to improve the process from prior years, and saw the value of data that helped them “see that they’ve made growth” with their students, they had not spent much time outside of required training internalizing the guidance within the NC Teacher Evaluation Process. For instance, when being interviewed, teachers would ask to look at a copy of the guidebook that described the process more carefully, since they could not always speak to the numerous details within its standards off the top of their head.

Because teachers were not certain exactly who going to be observing them each time, that also exacerbated the feeling that their ratings could become hasty assessments made by a subjective, yet unfamiliar party. The participants in this study were in agreement that

variation in how scores were determined detracted from their value, “Cause it may depend, like she said, on who it is. You know what I mean.” Since, “The interpretation is going to be different with each individual,” the feedback that they received might be misaligned with their own perceptions, which they felt were based on a deeper understanding of their students’ needs than a series of casual observers would be able to achieve. As a teacher explained, “I’m just saying two different people are evaluating you, one person sees all of this great stuff and then this other person’s like, ‘Mmmmmm’. I’m just saying...” This echoed the common sentiment that, “It depends on who your observer is, too”, since at times, there “just might be different personalities” at play, and “the way that it’s framed, it’s being told like 16 different ways, and so nothing’s really as consistent”. Teachers were concerned that if this led to unfair depictions of their practice, and the rumors of evaluation ratings affecting their eligibility for performance incentives like bonuses came to fruition, “that’s just very...cutthroat”. They were slightly alarmed that the stakes of evaluation performance might rise, now that a few districts had already begun “putting in there that, oh, you get a bonus if you get a score. I mean you’re, like I said before, I think what’s eventually gonna happen...” They were not sure how much weight should be given to evaluation ratings that seemed to be in the hands of varied individuals who could “finagle stuff”, since, “...to me the objectivity of it sometimes can be a little...[pause]”, implying that it was questionable.

In addition, teachers were the most critical of prior situations where they perceived administrators as turning in ratings that failed to reflect a deep, holistic view of their work, instead limiting their comments to a handful of formally observed lessons. In describing a less ideal circumstance, a teacher said, “In the past, you know, that one shining moment of that person may not have even been the person that you see, but you’re sending in a

summative evaluation.” Because formal evaluations were brief and infrequent, “They only really come in two, maybe four times, max a year so that’s four lessons out of - you know, it’s not really indicative, I don’t think”, which caused teachers to feel unconvinced that their observers were familiar enough with their pedagogy to make an accurate assessment. One teacher gave the example of how her observations happened to be during lessons where she did not integrate technological tools, “but I wouldn’t say I don’t use technology, it was just that one specific time she was in there. My Smartboard, iPad, I use them all the time. It’s, like, it’s hard to know that they really can see that, that they know that you’re doing that at all times.” As another teacher articulated, although the evaluation process could theoretically encompass many “wonderful” purposes, “It’s like the total parts don’t add up to the whole. The whole is so much bigger than all of those little pieces.” Across the group, participants were torn between taking those constraints in stride as part of the nature of the evaluation system and dismissing the resulting feedback as limited.

Uncertainty about future state-level decisions about the evaluation process as a barrier to consistently effective implementation. On top of school-based uncertainty about how evaluation might end up being carried out, participants were not sure how to feel about potential state-level proceedings like the possibility of their evaluation ratings being made public. One said, “Does that mean we’re online?” Another dismissed the question as irrelevant, saying “I want to say that’s not important.” But the uncertainty about what that would mean for their professional reputation gave others pause, as the original questioner said, “Well...”, and another teacher dubiously chimed in, “Yeah, that’s kind of scary.” Their trepidation over evaluation ratings contributing to their reputation within the community tied into to a general concern that teachers get “beat up” and “blamed for stuff

that's not always in their realm of control.” While the participants in this study spoke about trying not to pay too much attention to chatter about public attitudes towards teachers, they still found “stuff coming from above us, rules made by people who have never taught, legislation, testing,” to be “frustrating” and “ridiculousness”. Because “A lot of things that have been reported have been so skewed...just statistical data has been so spun, it’s just been awful,” teachers worried that “the governors don't seem to have the respect for us, especially the lower grades”. Their perception of evaluation measures as unreliable tools for high-stakes salary and personnel decisions made them skeptical of proposals to use ratings to reward only the top percentage of teachers or to replace career status as the major determinant of job security. When describing proposed ways of linking evaluation ratings to job stability, one teacher noted although one system might be in place now, “you could change that the next week”. They were secure in the knowledge that their “parents for the most part are very complimentary, very appreciate of what we do, and they’ll quickly speak up to that”, but noted that within the public at large, “It depends on the ... it depends on who's talking”. Given the current direction of local education policy, one teacher explained that, “You hear about them doing away with tenure and all those not good, negative things, and it just makes you wonder if this process is going to be taken into account when those things arise. This is statewide, anybody can pull it up and see it, and I don't know. It could be held against you.” This exemplified fears about the future that had not yet been realized, but weighed on their minds in the background.

Teachers also worried about aspects of the state’s design and subsequent application of evaluation mechanisms that they saw as beyond their ability to influence. For instance, they wondered how their performance would be rated if they decided to move to a new grade

level, or take on a challenging group of students. In that case, “Heaven forbid, you move me to first grade, but I’m not gonna be distinguished anymore because it’s a learning curve. And that needs to be taken into consideration, too, and not go, ‘Oh, well, she was distinguished last year, oh my goodness.’” Teachers in one school asked how the incorporation of test scores might detract from teachers’ willingness to take on special education students, and doubted that the evaluation process would consider that “Your children’s growth and your ability to teach are...not always directly correlated”. Adding to the confusion, they were not sure if the state wanted them to demonstrate student proficiency or progress, since “I don’t know how that works”. One teacher gave the example that it is, “not okay if this year’s class was at 95 percent proficient and next’s year’s class was 65 percent, so I went from a good teacher to a crappy teacher, in one year...I look at the whole child and how much they grew. There’s not a test or a measurement that can measure all of that...every child is different”. Even if value-added models of measuring student achievement were more growth-oriented, they did not believe that such models were perfect, either.

On average, the teachers in this study rated value-added scores as “somewhat important” to them, indicating that they thought this was something their school cared about more, and put more stock in, than they personally did. Teachers explained that if they were forced to integrate growth models into their daily reality, they would have to give them more attention, but as it was, they still had questions about how well they could isolate a teacher’s impact on a student while accounting for the complexity of variables within the educational system. For example, one teacher who had worked at several different schools discussed how during her first year of teaching in a more affluent area, her students got high test scores that she did not feel she could take all the credit for. On the flip side, now that she was

working at a Title I school, she felt that making gains was more difficult, even though her instruction was much stronger than it had been as a novice. Uncertainty about how the state would address these issues made it difficult for teachers to believe in the continued efficacy of the evaluation system, as they questioned how it could be fairly implemented across the state amidst frequent changes to classroom composition and teacher placement.

Constraints on administrators as a barrier to consistently effective implementation.

Although the teachers in this study could identify several reasons why evaluation implementation ran into difficulties, they were sympathetic to the logistical challenges that naturally arose from trying to deliver high-quality feedback to a large number of teachers. While they wished feedback delivery could be improved, they saw administrators working hard and “scrambling” to meet district deadlines to turn in evaluation paperwork. Teachers could see that when those pressures arose, “It’s hard for administration”, and “It’s a crunch for them to get it done”. In some cases, even though bringing the ratings system online had cut down on physical paperwork, teachers saw their principals struggling with the system continually “crashing” and causing further delays in processing. Several identified “time in general” as a scarce resource, explaining that the limitations of the school day made it hard for administrators to make the formal observation process, which also included pre- and post-conferencing, meaningful and substantive, rather than a rush to turn in all the required components within the allotted period. It was a puzzle that could not be solved within those boundaries, as teachers saw principals who made an initial effort to get to know teachers at the beginning of the year as having difficulty maintaining that regular presence within their classrooms. One said that there had been some years in the past when, “the kids don’t know who their principal is”, which she saw as an example of the potential for division between

administrative offices and classrooms. As such, she appreciated visible efforts from current administrators to try and rectify that distance, saying that “I can’t imagine having to do that in all of these rooms and feeling connected but it did feel like, okay, they want to be in the rooms”. However, during hectic points, other high-pressure priorities like extensive state “testing” in upper elementary classrooms still led to a decrease in a principal’s presence in the younger grades. One teacher noted that, “You should see them having to scramble with the new ‘Read to Achieve’, testing, and all that hitting everyone when all the observations were due”. This resulted in many teachers having long stretches without administrator interaction within their classroom, as a teacher from another school echoed that, “We haven’t had any [observations] lately. I don’t know if it’s with all the 3rd grade testing”. Quicker walkthroughs and pop-in visits could help rectify that disparity, but they were not a full substitute for extended conversation.

Although principals were trying to squeeze in as many briefer, casual interactions as possible, formal observations that were explicitly tied to professional standards sometimes did not happen until almost halfway through the school year. Often, teachers reported that “I’ve not had a formal observation this year yet”, even though they had been working with their students for several months, or “I just had my first one” in January. While they acknowledged that their administrators were probably busy giving more attention to newer, needy teachers or focusing on testing grades, this created a further disconnect from the official NC Teacher Evaluation Process. And, although teachers did not blame principals for these circumstances, and “understood” why a lack of time was a constant problem, they also believed that it did not “mean it’s right” for them to be getting the short end of the stick when it came time for evaluation follow-up and relationship-building. Because “It’s complicated

to look at what goes on in a classroom in a day...you can't then say, you come in for twenty minutes and, checklist, well, we're done". While conceding that, "They have a lot of paperwork, they're in charge of every single child in the classroom, every single bus", teachers drew parallels to their own work life, with one saying that, "I mean, I get it, but at the same time, I know that my children who I have good tight relationships with perform better for me in the classroom". When asked what improvements they would make to the current evaluation system, teachers typically made suggestions like wanting, "More time, sometimes, because it's really not very personal after they leave. They either send you an email or a note", or said, "With the timing too, a lot of times it's just like we got to get it done. Let's be done, let's meet, let's go over it". Although they saw why bringing the evaluation process online was more efficient, a brief email was not going to compensate for reduced meeting time or form the basis for facilitating extended professional development. The limited time allotted to the official evaluation process consequently led teachers to perceive that making performance feedback more meaningful was low on the priority list.

Lack of alignment: evaluation does not fully align with broader teaching goals and expectations. Teachers reported that areas of focus within the formal teacher evaluation process were rarely connected to other professional development opportunities, and were not always aligned with briefer but more frequent "walkthrough" feedback resulting from informal administrator visits. They also felt that evaluation did not adequately address all components of early childhood and early elementary education, including the extent of their relationships with students, lessons on character building, and the importance of developing socio-emotional skills that went beyond academics. As one experienced early elementary teacher put it when talking about her administrator's attitude towards her practice, "She *likes*

us, but she doesn't know what we *do*." In general, they did not consider data from a few formal observations a year to be an adequate representation of their instructional efforts, not just because of time constraints, but also because of administrators' limited capacity to offer instructional advice that was appropriate for their grade levels.

Fragmented professional development links leading to a lack of alignment. When asked if they ever discussed the evaluation rubric, observations, or feedback with their colleagues, six out of the 12 teachers in this study reported that they sometimes did, while two said that they either rarely or never had those types of conversations. This pointed towards a missed opportunity that arose from evaluation feeling like a private, individual judgment that would not be shared with others. While sharing their exact performance ratings may not have been productive, teachers had little opportunity to solicit targeted ideas from others and tie evaluation into what their learning community or grade level was working on. They also felt that while their evaluation feedback and reflections were sometimes tied to other aspects of their teaching, such as when mentors for new teachers were included in reviewing a teacher's plans for improvement, they were not as obviously related to the professional development support provided by their school or district. A teacher reflected that, "I don't really know if it ties in with my professional development so much. That could be something completely different from what I'm observed on", and explained that if there was overlap, it was coincidental. This group of teachers tended to dismiss the utility of those professional development sessions in general, since quality widely varied, and "the good ones are important, bad ones, not". They also described a multitude of areas of focus, from the Common Core to specific grade-level strategies, which required their attention but were never integrated with the NC Teacher Evaluation Process or the official

professional development plan that was included within evaluation's components. In fact, one teacher laughingly pointed out that the latest evaluation guidelines had not been updated to reflect that the Common Core State Standards had been adapted. As a result, teachers tended to feel as if the evaluation process was disconnected from other aspects of their professional growth, remaining in this space where it did not conflict with other teaching objectives, but did not explicitly support them. One teacher said, "I feel like all of our feedback is kind of quick. I feel like sometimes it would be better to have a little more constructive, like, suggestions of what you can do to make this lesson better or to make your teaching better." Even though she found that the impressions administrators formed during informal visits to her classroom were sometimes reflected in her formal evaluation ratings, links between brief comments and next steps were missing. In addition, a colleague in the same school felt even less connection between different pieces of feedback, saying that, "It's not as smooth", and "When you look at your summative piece, it may not directly relate to one area or another", indicating that there was a lack of consistency across teachers' experiences.

Incomplete inclusion of early elementary objectives leading to a lack of alignment.

Because the same rubric is used to measure teacher performance across all grade levels, it is up to administrators to make baseline standards come alive for individual teachers by relating them to their particular grade level and circumstances. One common issue for early elementary teachers was that their administrators might not have training in working with younger children. They were worried that evaluation standards emphasizing academic content delivery were symptomatic of a broader trend towards ignoring other important facets of a child's personal growth, such as socio-emotional development. As one teacher

described the types of questions that drive her growth goals for students, she said, “Are they noticing that they are part of a bigger picture? Children. Are they working on community service, are they working on self-regulation? All these things that are so important for them when they get older...we have been focusing on the wrong thing.” One of her colleagues added that “I don't think that enough focus is being put on...on the self regulation and I know that there's a lot of articles that are out there recently, research-based, good data to support the importance of the self regulation and how that eventually, when they get to 3rd grade, if they can attend to a task if they can self regulate, then they will be more successful in the upper grades”. They saw these elements of their teaching as essential building blocks for the future, but they were not always confident in their evaluation observer’s ability to recognize their value. Although one teacher thought that her observers should be able to consider the age level she was working with, she hoped they would understand that incorporating movement during “wobble breaks” was appropriate. A prekindergarten teacher was frequently put in the position of “giving to our principal...a breakdown, especially for early childhood”. She had to explain what she was doing to administrators, rather than administrators being able to offer her instructional support. She also found formal evaluation measures provided by the state to lack alignment with other key measures of early childhood classrooms. For instance, “I have teachers that are my same grade level that are coming to me and going, ‘I'm doing what ECERS [Early Childhood Environment Rating Scale] says, I'm doing, I'm getting five stars on my rating and my principal is marking me down because they can't find anything in my classroom that's multicultural’”. Another teacher gave the example of expectations like students managing their own learning teams being “developmentally inappropriate” for younger children according to the trade definition of

that concept, which originally related to upper elementary objectives. It was confusing to figure out whether they had covered all of their bases for each set of standards that was being used to assess their classroom environment. One participant explained that she had to go beyond the universal NC Teacher Evaluation Process guidebook to tie in a more concrete early childhood supplement provided by her district, which she appreciated but required an extra time investment. In addition, teachers could not trust administrators to “see that connection”, because they would “have to have experience” working with young children in order to completely understand the intentionality of their instruction.

Over-evaluative: Too much emphasis on end ratings, rather than growth model.

Even though the NC Evaluation Process contained many growth-oriented structures, such as rating teachers on a continuum from developing to distinguished, prompting reflection, and tying evaluation to a professional development plan, teachers still perceived end-ratings as having greater importance within their school setting. They wished that evaluation did not feel so “high-stakes”, because:

I don't care that you've observed me, that doesn't bother me. I don't care about knowing it, I don't care how formal it is, but if I can truly do it for a reflective practice and looking at what is it that I'm doing - what do you see that I could better? And if it's just purely that conversation for making me a better teacher or helping my students in a different way I haven't thought about, it wouldn't bother me in the slightest, but when it's tied to my success as a teacher for my job and grading me on it, that changes it, and then it's a whole different ballgame.

Interestingly, even though the North Carolina Evaluation Process has not yet become explicitly tied to career status or salary on a statewide level, the suggestion that it might in the future, along with the simple idea of being rated and wanting to perform well, was enough for most teachers to feel like evaluation was a high-stakes enterprise.

Pressure and defensiveness resulting from evaluative emphasis. The pressure that even experienced, confident teachers associated with evaluation was one of the clearest trends across this study's data. Teachers frequently used words like "panicking", "anxiety" and "stress" to describe their anticipation of an observation or a post-conference. One veteran teacher said, "It shouldn't stress me out as bad, but it does, just as bad as anybody else". Another highly experienced teacher recounted how "I would rehearse lessons, and in my sleep, dream about lessons, and be thinking about lessons when I'm driving down the road. Something that's really not that important." While most of the participants in this study felt fine about what had actually happened during evaluation observations, and tended to think the anticipation was worse than reality, a few had been deeply affected by demoralizing experiences where "It was just not fun for me, and it frustrated me" and often thought, "Oh, my God, it's too much", since "To know that everything weighs on that formal observation, a lot of your...that's a hard thing to swallow sometimes." While on one level, most of the teachers in this study were well established within their practice and their schools, very few could take a relaxed attitude towards being watched in the classroom by an administrator. At best, they would say, "I really do pretty good, considering", or concede that although they did not like being formally observed, they were used to it. It appeared that for many participants, the process of being judged and committing ratings to paper would "freak people out" and become more tedious than energizing, since, "It's a lot when you're reading through it, you almost get like a headache thinking about it".

Teachers also took a somewhat defensive approach that related to their broader concerns about limited observation time resulting in less than accurate ratings. They would push back against criticism, or worry that the observer had seen something beyond their

control that they should not be held accountable for. For example, a teacher protested that, “You were in my classroom for 45 minutes...so I just think that it's also frustrating, also cause I really took that hard and I didn't think that was fair.” Another fretted that, “You have control of your students as much as you can, but there's always just that unknown of what they're going to do. That's what makes me the most nervous”. Similarly, a different teacher in her school explained that sometimes, “It's like a gotcha instead of, like, a real decent look at what you're doing, 'cause it depends on when you come in. You could have had, you know, Johnny flips out and it throws everything off with your day, but this is your 45 minute observation.” Others tried to argue, “Is it done on a snow day or a day when you have early dismissal or early release?” Their emotions ran high when describing how, “I take it too much to heart...I take it as a push down...and then sometimes when it becomes comparative, I really take offense to it”. The most experienced teachers in this study reported that their long-term investment in teaching caused them to internalize perceived criticism, even though they would not put themselves in the category of “poor teachers” and in fact, said that they often received the highest ratings possible. Regardless of experience level, it was “hard to only take it reflectively” when “Somebody's tearing you apart for what you do and that's very personal”.

One established teacher felt that because of her relationship with her principal, her attitude towards receiving critique had evolved to the point where she could move past her personal emotions. However, she could see her colleagues struggling to perceive evaluation as beneficial, instead of punitive. Because teachers “were those kids who wanted to be pleasers”, and there is “not a whole lot of comfort that goes with testing or evaluation”, that resulted in “a lot of defensiveness”. For instance, although it appeared that this had not

actually happened to them very often, teachers still asked, “What if they come in that one time at the end of the year and things aren't right on track that day?” or shared partially unfounded fears that “I guess I'm always worried, too, that they're just gonna come in and...you finish something early and you're doing, like, a filler and it's not exactly what's on your plans, I'm like, ‘Ohh, what if they come in right now?’ - you know?” This indicated that teachers saw observations as a means of monitoring them, rather than as a support mechanism. Typically, while concrete examples of times when administrators had showed a flagrant lack of understanding of their classrooms were minimal, a few less-than-ideal anecdotes were enough to taint the evaluation process.

Some teachers even thought that the pressure to perform could result in “pretend, pretend”, where “I just feel like sometimes I'm teaching to the evaluation, you know what I mean?” They explained that it was hard for them to take evaluation feedback seriously when they knew it was based on such a performance, since “I feel like when you're trying to prove yourself to someone, it's not necessarily the most valid classroom. It's not the most accurate representation of you.” Within a focus group, one teacher suggested that, “Sometimes it's better when it's all natural, just flowing”, and another added that it was best when “teacher-student interactions are going on, and it's not staged”. This idea that a prepared lesson for a formal observation was going to be inauthentic also contributed to the idea that observations of such lessons would yield ill-informed ratings. As one teacher talked about how the evaluation process did not seem like something worth investing in, she pointed out that in her school, “I think you have some that are just going to put on a show when somebody is in there. So then that doesn't really show who that teacher is.” Even if she felt that her work was accurately represented, the process as a whole left room for error.

Perceptions that own professional judgment is in doubt resulting from evaluative emphasis. In addition to feeling general pressure around evaluation, participants in this study were sensitive to the possibility of evaluation becoming an affront to their own ability to make decisions about their classroom. They valued certain parts of the evaluation process like planning out their professional development, but did not always believe that administrators truly valued their input. A teacher shared that, “I feel like the self-reflection is one of the most important pieces, but I don't feel like it's seen as important to the people who should be looking at, ‘How do you feel about yourself?’” Rather, since evaluation was primarily seen as a monitoring tool, they viewed administrators as reducing the process to an attempt to, “make sure we’re covering it” and fulfilling baseline responsibilities. As one teacher articulated, “That always ends up feeling like you’re not trusted to people and you’re not being treated as a professional. When you get into all those, you know, do you have your documentation? And then you lose the artistry. To me, teaching is an art because all of that interplay with people.” Other teachers agreed that a back and forth, in-depth conversation was necessary, since “It’s a dialogue...and if it's used that way, it makes a little more sense, but when it's not, that's obliterating,” to your sense of professional self-worth.

Since these teachers conceded that they were nowhere near the “danger zone” of poor ratings, they also felt that the constant drive towards continuous improvement could sometimes wear thin. Although one teacher appreciated a growth-mindset, she said, “It'd be nice to know that sometimes just to hear you're doing a good job. Sometimes I just need to hear, ‘Good job, we appreciate the effort you're doing, and we're seeing it in your kids’. There's a time for it.” In her mind, “There's a time and a place for reflection, there's a time and a place for those questions of saying, ‘Ok, what could you do differently? How could

you do this?’ Or, you know, think about something. But you don't need to hear that every single time, because then it makes you start second guessing everything you do.” Her colleagues agreed that the structure of evaluation feedback and its end goal of producing summative ratings could make it hard for their administrators to focus on recognizing their teachers’ professional expertise and affirming the strengths within their practice. For instance, one remarked that because she did not think there were any teachers in her school who needed to be “weeded out”, the implication that an evaluation system was still needed to monitor teachers was demoralizing. She explained, “I just feel like a lot of times accountability and test scores is pushed, I feel like it's a lot of mistrust”. While teachers acknowledged that the official evaluation process did attempt to capture the full extent of their teaching efforts, there was a sense that administrators were always supposed to be looking for areas to change, across all levels of teaching ability.

Empty: This sort of evaluation is not personally meaningful. While evaluation could have productive elements, the NC Teacher Evaluation Process was not typically seen as meaningful or well connected to personal drivers of their teaching. Teachers reported that the standards themselves were too generic and broad, and conversations about observation feedback that did result in useful thinking were rarely sustained over time. They compared the flat feeling of reviewing of a standard rubric unfavorably to more relational, individualized influences on their practice, and often failed to find inspiration or motivation from evaluation metrics.

Standards too generic, resulting in less personal meaning. The teachers in this study all accepted the evaluation standards as a good standard for baseline practice; when asked, they in fact thought that at first glance, evaluation’s intended purposes sounded pretty nice.

In a focus group, a teacher asserted that the elements of evaluation were “just the elements of a good teacher. As long as you're teaching them you're teaching within those elements, good teaching is happening...and then I don't think you should have to worry about that piece of paper.” However, while other teachers agreed with her, they also found that the standards were too open and “vague”, and too generically constructed to make an impact on the complex facets of their instructional choices. As another teacher put it, “They might as well go back to the satisfactory and unsatisfactory.” Because these participants did not see education as cut and dry, they doubted that it could be boiled down to the constraints of a rubric, especially one that used qualitative inputs to produce quantitative ratings. An early elementary teacher explained, “I think education is a different animal than business is...there are things that are cut in dry in business that aren't in education. These are children's lives. Somebody doesn't make so many things on this line over here, it's not the same...the wholeness of what made that child get to that point doesn't get factored in.” In her mind, business-like structures were being used to measure completely different type of tasks, within a much more complex, person-focused setting.

When conducting peer observations, one teacher described trying to fit what she saw to the evaluation structure, since, “a lot of the things I noted, it was almost kind of hard to figure out where to put them.” For instance, she observed that she could tell what children had been learning in the classroom from looking at their work on the walls. The children's learning was visible, and felt like a more authentic representation of their efforts than pre-packaged classroom decorations. After searching through the evaluation guidelines, she decided that notation might best relate to a standard related to providing a safe and orderly environment, but that that still did not quite fit what she wanted to capture. She also had

questions about the precise difference between various levels of proficiency, asking, “Does it go under ‘demonstrates’? You know, which one of these words does it go under? Maybe that’s a big piece of it too, even doing your own self-evaluations, I don’t know which box this is...I’m not sure, it doesn’t really break it down, which one you go in.” The prevailing sentiment about the standards and the ratings continuum was that “I don’t know that it’s specific enough...there’s so much room for interpretation.” Other teachers within her school gave additional examples of when they had to ask for clarification about components of the evaluation system, such as exactly what 21st Century learning or building global awareness might look like on the early elementary level.

Teachers reported that as a result, this also watered down the quality of the feedback they might receive. One teacher dryly recounted, “I’ve actually been told as long as they get home safely and that no one dies, I’m good.” Her feeling was that the standards themselves did not go much beyond basic objectives that seemed obvious to veterans of the classrooms. Her colleague provided another example of when she noticed that she had gotten “distinguished” ratings in almost every category except for one having to do with meeting the ethical standards expected of North Carolina teachers. When she asked her administrator why she had only received average marks in that category, worried that she had done something untoward, she was told that it was simply that, unable to make much distinction within that broad standard, “she just gives everyone proficient on that one”. This made for an amusing anecdote, but deepened her concern that a worst-case scenario had unfolded, where ratings according to these standards did not mean much, but could have lasting effects on her reputation.

Not part of long-term, sustained growth trajectory, resulting in less personal meaning. Although teachers had been asked at various points to fill out professional development plans (PDP's) and collect artifacts of their teaching as part of the evaluation process, one problem was that there was rarely enough administrative bandwidth available to help teachers make ongoing connections to their long-term professional growth trajectory. As a teacher who had diligently amassed a binder of artifacts pointed out, "It's been sitting up there [points to top of high shelf] from two years ago. I never changed anything and nobody has ever looked at it." A teacher at another school confirmed that sentiment, suggesting, "It wouldn't be as tricky if we did keep those portfolios". Her colleague agreed that, "If we had them from year to year, we could look at our evaluations, we could say, 'Okay, what did I do last year that I got marks for this and this year that I didn't?'" Other teachers commented, "I feel like they vanish", and "They go into space", before ultimately concluding that they must be in "personnel files somewhere". One recounted the involved process of trying to track down the history of her performance reviews over the past five years from the state agency that had stored those results in an amorphous database. Without an explicit, sustained focus on their long-term growth, teachers said, "It's going to get lost; you're losing the importance of what the evaluation really means". In addition, while integrating more frequent walkthroughs created more touch points with teachers over the course of the year, and they appreciated that "they're really getting a picture of you at all different times of the day", teachers saw that feedback as random and "jumping all over the place". Because for some teachers, "normally they don't give you suggestions anyway", there was little concrete dialogue to build from. They also were unsure what next steps looked like if they were to get a lower rating. One asked, "If you do poorly, is there a

consequence? I don't know if there is a consequence.” Other teachers reported that as a solid performer, the utility of evaluation feedback seemed to hit a plateau. They explained that because “distinguished” ratings were only supposed to be given out to teachers who were making exceptional contributions to the district, “Where do you max out and just become stagnant? And if I do max out and just become stagnant, what's going to happen to me?” At a certain point, repeatedly being rated as “proficient” or “accomplished” became meaningless, and detracted from the idea that evaluation could be used to track their progress each year.

Not based on authentic relationships, resulting in less personal meaning. When asked the difference between evaluation and more central influences on practice, one teacher matter-of-factly broke down the crucial difference – “Who has a relationship with a rubric?” Comparing feedback resulting from close relationships with colleagues and mentors to the type of cursory feedback they would get from formal observations made it clear that the evaluation process had much less of a chance of having an impact. This teacher further explained that, in some cases, “Because you have a relationship with that person, you're much more likely to listen and to take to heart what they say. Whereas, if you have someone in a suit who shows up twice a year to do an evaluation and you never see them any other time, then no matter what rubric you give it's not gonna matter.” To make the evaluation experience seem authentically connected to everyday interactions within the school community and feel “genuine” was a challenge. Yet, participants from across the study emphasized that if administrators made efforts to take on that challenge, it would pay off in dividends. A teacher at a different school confirmed that, “When people feel connected, that's what makes a difference. And I think all the fixes in education and in the world in

general, is when people feel connected.” Consequently, if administrators were unable to move teachers’ perceptions of evaluation beyond, “I know it’s their job and they gotta do it”, they would find it hard to get teachers to care about it and feel invested in the process.

Not motivating or inspirational, resulting in less personal meaning. Given the combined lack of buy-in and sporadic efficacy surrounding implementation of evaluation standards, participants in this study did not usually feel that evaluation was able to tap into their core motivational drivers. One teacher explained that, “It’s really just the way you feel. It’s not like after you get observed, I’m like, ‘Oh yeah! This means so much to me.’ I just feel like they’re doing what they need to do and I’m doing what I need to do. You know it only happens three or four times a year, where you’re seeing your kids everyday and it’s more in your face.” In other words, aspects of her professional life, such as interactions with her children, were part of her daily reality and therefore more of a focus. She would get immediate feedback from them, and could experience regular emotional exchanges within her classroom. As this teacher put it, when it came time for evaluation, “I understand and I do it but it doesn’t drive me.” Another expressed a similar sentiment, saying, “More often than not, it’s probably not as driving as they intend for it to be”, and explaining that, “I’m numb to it, so I’m like, whatever. It is what it is.” While this did not mean that teachers completely dismissed their evaluation feedback, their stance was more along the lines of, “I do think it’s necessary for people to grow. It’s just depending how it is approached. It’s the same way with how we treat our children. Do we give them a test to show them what they did wrong or do we show them what our next steps are?” If teachers felt that evaluation was primarily designed to point out flaws, without real support to move them forward, they were less likely to rely on the evaluation process to propel their future career trajectory.

Administrative: Evaluation process becomes merely an administrative task. With all of the challenges surrounding evaluation's perceived lack of meaning, it was very easy for it to seem like just another hoop to jump through, or "just another thing that you have to do". The guidebook for the NC Teacher Evaluation Process is about fifty pages long, and there are numerous components for teachers and administrators to complete. Absent of personal connections and buy-in, teachers often saw evaluation as the latest iteration of a host of accountability measures to "check off". They particularly disliked the need to document each aspect of their practice, since they felt that their administrators should know their staff well enough to be able to form an impression of their teaching and recall their contributions to the school community.

Gets lumped in with host of accountability initiatives teachers already disillusioned with. Because more rigorous teacher evaluation followed a series of measures designed to create accountability within schools, teachers felt that it "did make it harder" at times not to dismiss evaluation as yet another mandate designed to judge their practice and create more paperwork. After receiving "eight hours of it" during training sessions designed to uniformly introduce evaluation across the state, their first thoughts were along the lines of "Teachers are losing their jobs, and we're printing these [evaluation guide] books!" Even though those books were now online, they were not certain if the considerable resource investment in evaluation implementation was worthwhile. In focus groups, teachers talked about how "you have to sit through this again" and needing to go to "a gazillion" meetings. They were not inclined to value evaluation as a resource that could enhance their practice, instead seeing the process as placing an additional burden of proof on them to demonstrate that they were providing adequate instruction. One teacher in this study explained, "It is very time

consuming with all that's already heaped on top of us, but obvious things...I don't know, if there's some things, it should be understood.” Because teachers were not inclined to invest themselves in the evaluation process, they were also more likely to take on an attitude of, “this will pass”, lumping evaluation in with other measures that they had to endure until the next change came along. One teacher who had seen policies come and go said that evaluation “loses its ability to be powerful and useful because it's just another thing that's coming down from the state. I think it gets lost in the shuffle and it's just one more thing I've got to get done”. Teachers reported mixed feelings about accountability policies as a whole, with some claiming that they tried not to assign them too much personal value within their realm of influence, but others asserting that negatively framed pressures were a visible part of their working environment.

Artifact component creates excessive documentation. By suggesting that teachers collect artifacts to demonstrate their performance within each standard, the NC Evaluation Process makes an attempt to capture aspects of teaching that observations might not reveal. It also gives teachers the opportunity to influence their ratings, and use evidence of good practice to argue against marks they feel are unjustified. Unfortunately, the feeling that “sometimes it’s all about paperwork” cheapened the process for teachers. It also added a time-consuming burden to their already busy days, since “It's like you have to sign this, and this, and then sign off - it seems like there's so much, instead of just a simple submit”. When describing their annoyance, one teacher explained that, “To me, a lot of it then becomes falling on the teachers to then prove themselves, so instead of, let me go and teach and you evaluate my work, I'm now having to come back again and gather all my information to

present to you a second time.” Having to “backtrack and pull this and show pictures of this” seemed like unproductive “busy work” to them, and a poor use of their abilities.

Teachers said this also made them feel mistrusted, since in this past, some principals could not take their word for it, but instead would say, “Well, if you want to be pushed up, then you have to have your documentation of all of this and all of that”. Simply bringing up an example of how they fulfilled a standard was not enough since, “I can say, ‘Oh no, but I have it’, you know, but then you have to go and find that piece of documentation.” When this part of the evaluation process was framed as having “to prove it”, “it just feels completely like they’re there to get you”. The task of documenting each facet of their classroom could also lead to situations where, “It’s hours or weeks later that I think about it, and I’m, like, oh my gosh...and we did this, and we did this - I mean, I would panic, like staging the kids for pictures!” A teacher at another school who admitted that maintaining paperwork was not her strength said that sometimes, she would feel guilty about not writing down, for example, details of an intervention, but then, knowing how busy she was, would have to tell herself, “I did it, and that’s the important thing”, hoping that her administrators would agree with her. She also questioned whether excessive documentation detracted from more genuine motivation to fulfill her responsibilities. For instance:

I’m like, ‘Really, you want me to document that I went to their basketball games, really?’ Last time, we were told, you better - if it wasn’t on a piece of paper it didn’t happen. And that really depends on the quality and how your administrator treats that. Some treat it as a get ya, gotcha. You would hope that people who have been selected to evaluate you on that level would not be trying to catch somebody not doing what they’re supposed to. And I don’t do that for you to rate me on my time anyway. I don’t go to that because I get approval that I did community involvement.

Across all three schools in this study, teachers recounted instances where they felt that their personal relationship with their administrators and their evaluator’s broader knowledge of

their efforts should have superseded the need to justify evaluation ratings. This was typically brought up in relation to aspects of the evaluation standards that had to do with their own leadership and professional development, such as attending conferences, participating in professional associations, or presenting at grade-level meetings. As one teacher explained, “I think the frustration comes in that if I’m doing all these things and I’m having conversations with my administration, that they know I’m off work because I’m going to teach at a conference or I’m taking off early because I’m gonna go meet with a consultant, that then we’re also having to go back again and keep documentation...Seriously? You signed the paperwork for me to leave to go to that conference...” Her colleague explained that this sent a message that this process was not going to be based on real connections and long-term relationships, but instead was merely about checking off a to-do list and covering everyone’s bases. Reducing evaluation to “a game” made it difficult for it to matter.

Evaluation may not have the greatest impact on this population of teachers. As these teachers described why evaluation had a relatively limited impact on their lives, their particular characteristics also played a part in the formation of diminished attitudes towards the evaluation process. For instance, because they were mostly teaching in early elementary grades with less of an emphasis on standardized state testing, they described components of the process based on test scores, such as value-added measures, as “coming” on the horizon but not yet a full “reality”. Although they could anticipate these components being important to their schools and eventually carrying greater weight, they could not articulate exactly how they might affect their perspective towards their job responsibilities in the future. In addition, these participants were fairly established within their teaching career and readily earned average to good performance ratings. This made them unlikely to be concerned about

significantly improving their ratings, since, as one teacher put it, “Right now I’m not in the danger zone.” While feeling comfortable with their current instructional abilities could benefit them in other ways, these teachers’ lack of urgency around boosting performance ratings sometimes weakened their investment in earning better feedback. As a result, although evaluation did still have some importance to them, “It’s not gonna be something that you focus on.”

Evaluation can still add some value to teachers’ practice, depending on the circumstances. For most teachers, a combination of implementation issues and a lack of personal connection with the official NC Teacher Evaluation Process meant that their performance ratings had a moderate or low influence on their practice. However, several participants in this study were able to share reasons why some aspects of the evaluation process worked relatively well for them. First, teachers were unlikely to ignore evaluation’s presence in their schools and were motivated to earn good marks. They cared about their professional growth and were willing to consider evaluation feedback as a potential contributor to their development. In addition, teachers on either end of the experience spectrum were uniquely situated to benefit from evaluation’s structure. As a result, both very experienced and novice teachers had certain attributes that enhanced their view of evaluation policy implementation. Lastly, schools with an especially collegial culture had laid the foundation for better evaluation buy-in, making their teachers more receptive to incorporating feedback from observations into their instruction. Since these teachers had higher overall job satisfaction, it was easier for them to digest performance management dialogue. Taken as a whole, these particular conditions made it more likely for evaluation to influence teachers’ conceptions of good practice and professional identity development.

Teachers who care about their performance will invest time and effort into considering suggested improvements and desired areas of focus. Because the teachers in this study were people-pleasers and rule-followers, they wanted to do well on their evaluations. They were also inclined to seek approval and validation for their efforts, and described enjoying opportunities for their dedication to their craft to be recognized, which “always feels good”. For example, one teacher said, “I’ve always honored that time when you have that end of the year talk. Always like kinda like a big thank you for all the hard work that you’ve done.” Because they were not struggling teachers, and on the contrary, tended to report a history of relatively high marks, there was no reason to dread extensive critique, as negative feedback would likely be minimal. In and of itself, doing what they were asked with good humor did not necessarily make evaluation meaningful. However, the fact that teachers accepted evaluation as a highly visible policy that mattered to their schools meant that it was able to take hold within a school’s regular processes fairly quickly. It also meant that teachers wanted to know what they needed to do to move along the ratings continuum. So, while the root of their motivation was not always ideal, as it was largely performance-oriented, teachers consistently believed that the standards for practice outlined within the official NC Teacher Evaluation Process guidebook did have baseline value. Across the group, they agreed that, “I actually think, when you read the teacher’s evaluation tool and look at it, it’s really good information. I mean, yes, a high quality teacher should be doing those things.” In the best-case scenario, this baseline consensus at least opened the door for more in-depth conversations with administrators that had the potential to drive shifts in instruction.

In keeping with survey results where 100% of respondents rated evaluation as “very important” to their school, teachers continued to express that this process “must be very important to an administrator”. In turn, because they took their job performance seriously, some teachers took the attitude that “everything is important to me” and should not be summarily dismissed. Furthermore, although teachers did not always have optimal focus on the evaluation standards as a whole, they did commonly value the aspects of the evaluation process that forced them to document self-reflection. Teachers said that for them, this was a bright spot, since “I feel like for me, it’s a good self-reflection”, and they saw that piece of the evaluation process as essential to future professional development. One teacher said, that although she still felt that “Oh, gosh, I’ve got to perform” during formal observations, on the whole, “I like it in the sense that it gives me self-reflection, and I see areas I can improve and what I need to do, and I like seeing it from that aspect”. The mixture of compliments and friendly advice that she received added to her generally accepting view of evaluation.

Experience levels can also affect perceptions of evaluation feedback & openness to utilizing those suggestions to drive change. Most of the teachers in this study had a minimum of five years of experience in the classroom, if not many more. As a result, they were able to speak to the difference that their experience level made in their perceptions of evaluation as a possible influence on conceptions of good practice and a driver of professional identity development. Teachers compared their novice years to their current situation as they related key characteristics of those time periods to the utility of evaluation in their professional lives. Primarily through reflection, participants recounted how administrators provided more attention to the development of early career teachers, visited their rooms more often, and looped in mentors and coaches as part of professional

development planning. Teachers across multiple schools noticed that, “When you’re in the beginning teaching, you get more observations”. Their self-reported attitudes towards evaluation were also different in the earlier years, as they were used to be observed frequently during their student teaching and also felt highly driven to seek advice from others, making them more receptive to feedback about curriculum and instruction. One of the only newer teachers in this study explained, “You’re observed all throughout school anyway, so you’re used to feedback and commenting on your practices”. Being primed to receive “good pointers” often meant that their practice was open to change and adjustment. Less daunting expectations for performance meant that, “A first year teacher should not be, feel at all, like they need to be ‘accomplished’. You feel like you’re barely treading water.” To teachers who accepted these limitations, initially lower ratings on evaluation metrics were not as traumatizing, as they expected to be “developing” within the growth continuum.

On the other hand, while more experienced teachers sometimes missed the greater levels of support they had previously enjoyed as a beginning educator, they also brought a history of adaptation to new policies with them. Because these teachers had built their resilience to change, they could take new performance management requirements and resulting feedback in stride. A multi-decade force within the classroom related that, “I think you become more confident in yourself, and you do listen to these things, but you know how to pick and choose what affects you most and what you need to really work on and implement. I just think when you’re younger, I just think you immediately go to, ‘I’m doing a bad job if things are not in that perfect space’, instead of allowing yourself time to [say]... it’s okay not to know right now.” Furthermore, their confidence in their expertise often allowed them to move past defensiveness into productive conversations with their administrators.

Because of this, experienced teachers reported being able to have greater dialogue during post-observation conferences. They could articulate the intentionality of their work and justify instructional decisions that they believed were best for their children. However, they also had the ability to see criticism as constructive, since they knew one critical comment or lower rating was not going to make or break their careers. When given feedback, they could “Get past it to say, this is, yeah, I had a moment. That doesn’t mean the next 50 years have to be like that, there are things we can do.” One teacher explained, “There’s some things I don’t take as seriously, because I’ve been in it a lot longer and I know it takes a whole lot more for certain things to happen and stuff like that. It’s like, it’s really not the end of the world if something doesn’t go right”. At the same time, “I also am much more quick to stand up for myself for something...I’m not afraid to speak back at that summative conference that we have at the end of the year”. Experienced teachers like her were less likely to feel intimidated by the process of being judged, and more likely to take conversation about their ratings to a more detailed level. Since some of them had been through experiences like National Board Certification, they were used to having to reflect on their practice and explain their instructional thinking.

In addition, their official evaluations did not feel as high stakes or “overwhelming”, because “We’re well established in our community. And so for me, having families that...know you as a teacher and support what you’ve done and want their children with you. That to me speaks so much more...” Having worked in the school for a long time, at times even teaching the children of former students, meant that their network of support was widespread. As one veteran expressed, “staying power is a big strength”, which allows for greater comfort within the teaching role and less fear of one new change, such as a different

evaluation mechanism, turning their world upside down. When administrators took advantage of this perspective and leveraged experienced teachers as role models for others, at times asking them to share video clips of lessons or lead professional development sessions, they achieved an ideal balance between using ratings mechanisms to drive improvement and identifying instructional highlights. This balance allowed these teachers to feel valued enough to then accept comments that were not as favorable when applicable. For instance, they reported that their principal “respects what we’re doing” and was perhaps even “a little bit in awe” of the progress they were able to make with their students. Being established as experts within their school also made it more likely for experienced teachers to discuss instructional challenges within their team. Teachers who were part of a strong grade-level collaboration explained that as a result, “That level of experience within the team and knowledge of each other, and when they've done this together already”, led to easier dialogue about “How can we make it better the next year?” This forward-thinking attitude also prevented the benefits of their experience from turning into apathy or resistance, creating an ideal environment for continuous improvement.

Most importantly, supportive, collegial contexts can create an ideal foundation for considering evaluation feedback. Because the teachers who tended to rate evaluation more positively were primarily clustered in one school setting with markedly strong collegiality, their descriptions of school context provided evidence of how that work environment led to better evaluation implementation. In that particular school, there were fewer obstacles to effective implementation, both structurally and attitudinally. There was also a heightened sense of trust and respect between teachers and the administrative offices, so as teachers explained, “I don't feel embarrassed to say, ‘I don't know’”. Relationships between

administrators and colleagues were sustained over time, leading to greater evaluation buy-in. In addition, these teachers believed in the ability of their administrators, coaches, and grade-level leaders to serve as instructional guides. They were able to skip past typical roadblocks to effective evaluation-based conversation and jump straight to targeted discussions about how to best serve their students, because “I know that they have my best interest and the students’ best interests when they say, ‘Have you thought about this, could you try this?’” These conversations were perceived as having greater utility, since the “good advice” they received was directed and specific enough to have an impact on practice.

For instance, teachers at this school described their unique school culture as resulting from the deep investment that many teachers had made as founding staff members of the school. Because some of them had been within that setting from the very beginning of its existence, they felt strongly committed to making it work. A more recent addition to their staff explained, “I think that it's just not, 'Oh, I work here' but I think that there's something in the school, because it is so unique that there's an ownership in the school.” Teachers reported high levels of teacher leadership and involvement, and numerous opportunities for collaboration, on their grade level and beyond. One teacher, who had previously worked with her actual family members, explained that coming to work at this school was like working with an entire building full of family members who cared about each other, personally and professionally. Her colleague confirmed that she “felt like family immediately”, saying:

I think that where we are right now is fabulous. I feel like we have a great culture here together. We really support each other and we connect well, and know we can go to each other for anything. Even personal stuff, you know if I'm just having a terrible day...like I said we do support each other in everything you know – ‘I've never taught this before, what do I even do?’... ‘Well, it's this, and try this’.

In this setting, their principal ate lunch with teachers regularly, and teachers often socialized outside of required meetings. This meant that even if they did not feel particularly close to their evaluation observers, they could at least fall back on colleagues to move them forward. Across the data, these teachers valued a sense of “team”, and rarely saw teaching as an isolated endeavor. Interestingly, this feeling even extended to the district level, as one teacher reported, “I think honestly that we work in a great school that's a part of a really great county. And I don't always agree with the administration across the board, but I think that there are a lot of people who are trying to do what's best for children.” Several of her colleagues independently mentioned similar sentiments about feeling fortunate to work in a relatively supportive county, as compared to others within the state. This meant that mandates like more rigorous evaluation were less likely to be seen as yet another imposition coming down from the district.

Overall, although more recent changes in administration meant that evaluation implementation was not currently a smooth experience for every teacher at this school, the foundation for success had been established. This made the relative effectiveness of their evaluation feedback more likely to survive disruptions to their established routines and relationships that occasionally created implementation issues. Some of these issues, such as changes in administrator personality that led teachers to perceive evaluation as becoming less authentic and more rushed, have been discussed in prior sections. However, looking beyond those stumbling blocks, it was still clear that the majority of the positive elements within their school culture that supported teacher evaluation and made it likely to influence their teaching remained. For instance, teachers felt that it was okay for administrators to see “organized chaos”, since at the very least, the principal understood what the reality of

classroom life looked like and was generally on board with their style of instruction. And, even though they still wanted to impress her, their conversations about whatever had been witnessed tended to feel like a comfortable back-and-forth, rather than punitive.

Consequently, when evaluation tasks started to feel more like an administrative burden, teachers in this school were not thrilled, but remained willing to keep working together.

Teachers at another school with a historically less stable culture and administrative office were also able to report optimism about newly established enhancements to the evaluation process, which were starting to shift their attitudes about evaluation in a more positive direction. These teachers were able to leverage the supportive culture within their grade level to build better connections with administrators and serve as a model for their school. They were “excited” that a new principal had introduced “powerful” mechanisms to build closer relationships with teachers and use data more effectively to examine specific nuances of their instruction, like ability grouping. This plan involved regular one-on-one meetings with an assigned coach within their administrative team, who they would have the opportunity to develop a closer relationship with over the course of the year. By dedicating time to relationship building, ““We’re not, we’re not a number anymore. We felt like that in the past, when there was this paperwork that had to be done, and you’re not good enough.” It also allowed teachers to choose areas that they wanted to focus on after examining their students’ progress and discussing individualized aspects of their instruction, as well as the needs of their particular class. They would then be video-taped during a session focused on capturing the area of improvement that they had self-identified; for instance, one extremely experienced teacher recognized that guided math groups might benefit her children as much as they did during reading. These videos would drive conversation “that’s just going to be

personal between you and your coach” and from that foundation of trust, pinpoint changes that could be made in real time. As one teacher within this school put it, “That’s gonna move people” along a professional growth trajectory.

Within this more positive structure, helpful, “immediate” exchanges were happening, since when meeting with their principal, “There’s just always that ease of conversation. And she seems...she’s sharing information. I feel like she has a good knowledge from where she’s coming from she’s giving you suggestions or options.” For their part, as well-regarded teachers within their teaching and learning environment, they had agreed to embrace these extra efforts and make contributions to its success by sharing clips of exemplary teaching and leading related professional development. Ideally, “They really get to know what you’re interested in learning about and then can kind of direct their feedback on what you want to talk about. And there are, of course, some across the board things that we all need to work on. And that’s what our whole group PD can be”. In addition, although it is important to clarify that this new plan was *not* being explicitly linked to the NC Teacher Evaluation Process and its standards for performance, these teachers hoped that building a better relationship with administrators focused on targeted instructional growth would carry over to improve implementation of the state-level evaluation mechanisms. These feelings of deeper “connection” that allowed them to consider suggestions more seriously as part of a “growth model” and “reflect” together. If all went according to plan, “I do think that will also help them when they come in, when they do our formal observations, and they know our class more. And they know us more, so that can be very helpful.” The sense that their principal and the rest of the administrative team was becoming familiar with them and their students, to the point where the principal would know that someone’s mother was about “to have a

baby” or that another child was struggling with a certain issue at home, made a huge difference in their outlook towards evaluation implementation. These teachers emphasized that it was not just about administrators relating to the adults in the building, but also knowing “the children better”, so that observations of their classroom would be enhanced by that context for understanding what they were seeing.

Other key influences drive teachers forward and matter more to them than evaluation does. When asked to rate the relative importance of policy-related factors like evaluation within their practice, teachers continued to indicate that other person- and school-based factors were more essential to them. I was able to triangulate my data to show that focus group activity responses reflected earlier survey responses from teachers. For instance, ten out of the twelve teachers in this study rated accountability policies as either “not very” or “somewhat” important, which aligned with the seven out of twelve who specifically put teacher evaluation in the “somewhat” category. These participants explained that while accountability was important to their schools, they did not necessarily have the same personal connection to those programs and mandates. They also tended to describe the influence of policies like evaluation as pressures, rather than positive forces. The key distinction was again made that while evaluation fell into a middle ground, relational influences that were either brought into the classroom or developed over the course of their career meant “a whole lot more me than what she’s [referring to administrator] going to put on that piece of paper”. Within descriptions of what *did* matter to teachers, their personal beliefs, past role models and relationships with others within their school community stood out as the most important influences on these teachers’ conceptions of good practice and professional identity development.

Majority of lasting influences on teachers' conceptions of good practice and professional identity development are person-based. The combined impact of individually based prior experiences appeared to play a deeper role in teachers' lives than any one policy, requiring administrators to consider that teachers already walk in the door with a host of meaningful sources of pedagogical inspiration. Personal beliefs and values were strong motivators for the teachers in this study, along with past role models who made a lasting impression on teachers' career trajectory. Participants in this study described core beliefs like "the importance of school" and the value of education as central to their current dedication to their profession. These kinds of drivers also included attitudes towards their students, such as the belief that "all students need love and want to learn – just need the right space", or that "we all deserve to be seen and encouraged". Teachers saw these ideas as enveloping individual curricular components, and tying together their classroom community. When considering the broader goals of her classroom, one teacher said, "We all do bring those differences in, but we can still be working toward a common goal. And that's why I tell them, too, we're a family in here, we have to love each other and be kind to each other." Values related to both self-worth and seeing the worth in others drove them to incorporate messages like, "I'm very big on, 'Don't let your circumstances hold you back'" into their teaching.

Some teachers attributed their personal beliefs to their upbringing and family background. For instance, one teacher said that, "I am a Christian, and so I'm expected to do good in whatever I do, because that's reflective of everything, you know. Do everything in excellence, you know, you don't be a slacker." For her, this translated to defining reliability, preparedness, and effort as key components of how she carried out her job responsibilities.

Seeing her parents work hard at their jobs without complaint further contributed to the idea that, “You know, work is very serious...and, so, I mean, I’ve always seen them do that, so that was a huge piece.” Her colleague said that, “Being raised to think positively, and always have a positive outlook, and be proud of what you do, and accomplish what you want to accomplish and things like that” were a big part of her attitude towards her job, which allowed her to remain optimistic in the face of challenges. Another teacher explained that explicitly being taught that “School was very important” motivated her to “bring that in here – school is important, you’re here to learn, you know, you’ve got to do this”. This was indicative of a broader trend where teachers wanted to provide students with the same kind of structure and direction they had received growing up, which they did not believe many of their students were receiving in their own homes. Others simply appreciated “having that support system” that had bolstered their pursuit of their career, especially those who had teachers in their immediate family who had provided them with specific guidance about the education profession.

Within these participants’ lasting childhood memories, former teachers from their elementary school experiences were another example of influences that drove their current practice and conceptions of good teaching. Participants commonly credited “teachers I had – they were great, and that’s what made me want to be a teacher”. Across all three schools in this study, many participants cited role models from their own schooling experiences, indicating that “having awesome teachers growing up” drove their own career choices and created an ideal to strive for in the classroom. They modeled themselves after educators who had made a difference in their lives, because they wanted to make a similar impact on the children on their classrooms. As one teacher put it, “You know, you’re in school so

much, so I wanted to make a positive impact on kids, so that they liked school as much as I did.” The fact that their students spent the majority of their day with them was an additional motivator to create a “family” atmosphere within the classroom. One teacher recounted how different school was for her during her early elementary years, when she was continually in trouble for talking and misbehavior, as opposed to her experience with a seminal teacher who helped her shift her education trajectory by making her “accountable” for her teammates during group work, instead of isolating her from others so she would not distract them. She explained, “It sort of made me rethink about how what I do affects other people. I try to use that with the kids, that you are part of your team. Them understanding that we've all got one goal together and we've got to get there somehow...I still very vividly remember the bad things that happened with the old teacher, and then with this new one, the difference.” Her colleague explained that her former teachers had also been life changing, since although her parents were supportive of her:

I never felt like education was the biggest focus in my household, so if it hadn't been for my teachers, you know, I wouldn't be here. ... So, I feel like I can relate to them and I want to, you know, hopefully be one of those people who helps to push them in a direction to say you can do better...I can definitely say that if it wasn't for great teachers, I wouldn't be here. I don't know where I'd be. Stuck in Walmart or something.

This idea of wanting to connect with students in order to encourage them to maximize educational opportunities was a strong component of many teachers’ experiences that related to the role of prior schooling in their own success.

Lastly, teachers who attended strong teacher education programs, even if they took place many years ago, credited them as being a “very personal, one-on-one” influence that “set me on the right path”. One teacher who said, “My college experience was wonderful,” emphasized that “I cannot say enough about just, the in-classroom experiences I had, for

sure, “ since they exposed her to a variety of demographics and grade levels, and helped her determine her teaching preferences. Many of them remembered their cooperating teachers fondly, saying, ““When I did my student teaching, it just reinforced all that, ‘cause she was such a great teacher and I just loved watching her.” Teachers particularly appreciated mentors from their training who struck a balance between providing them with answers and letting them experience teaching for themselves, as they, “gave me all this wisdom and all this knowledge, but then gave me the freedom to make a mistake”. A principal at a school where one teacher had initially cut her teeth decades ago stood out as a similar type of foundational influence, since, ““She prompted you to reflect and to find and then sent you off to get that training. She made you a better teacher but in your own path. She'd let you reflect on what you needed.” That relationship exemplified the kind of professional support that this teacher continued to prefer many years later.

Relationships with key members of learning community like colleagues & mentors are most important influences on practice within school settings. The most clearly influential school-based factor, as perceived by this group of teachers, was “relationships with colleagues and mentors”. Every single participant rated it as very important, making it the only potential influence on their practice unanimously rated this highly, closely followed by school culture, which they saw as being related. About 83% of teachers found school culture to be either “important” or “very important”. Throughout the study, teachers repeatedly emphasized that because the relationships they had built within their school community were of such high value to them, and of such critical importance to their own professional development, they mattered on a completely different level than formal performance measures. One group of teachers concurred that, “What we do as a team with our PLC or

grade level means more to me...and even though that evaluation will end up being something linked to me and pay scale, I'm okay if we're [gestures to other teachers in focus group] okay, and my students are okay. Then I can live with whatever pay scale ends up needing to be". As teachers shared what they rated as having the highest value to them, one after another said something indicating that, "Teachers found that having "this huge support group" of "amazing people" was essential to their professional growth and well being, saying that, "I don't think I would be where I am without that kind of support". They credited their colleagues with helping them to survive the daily grind, as well as keeping their long-term goals at the forefront.

The value of collaboration and working together carried throughout participants' responses. One teacher explained, "I feel like we're all just very good about being collaborative and bouncing ideas off each other and it's not really a competitive thing. It's more what can we do for our kids." A teacher at a different school similarly noted, "Collaboration is so important to me, to be able to sit down and talk and bounce ideas off each other." With that shared understanding, teachers could work smarter, not harder, since, "It's hard for one person to tackle the world. Doing things together just makes it so much easier." They were not afraid to ask for help outside of structured opportunities to discuss their students' work, giving examples like, "People send emails saying, 'Look at this, check out this new app, this is really great, the kids will love, this try this program'". Given the challenges of their profession, "I don't know how people could get through it otherwise". As one teacher put it, "I have to have a team of people who are flexible and relaxed and if it messes up – okay, we'll just keep on going. I feel like if you cultivate that with your people as best you can, it just feels better." They also pointed at that feedback from colleagues was

readily accessible and immediate. When trying to problem-solve on the fly, “Those usually are the most effective, it’s the, ‘Let me tell you what this kid did today, please help me’”.

Their colleagues regularly added to their skill set by sharing their ideas.

Relying on colleagues was made easier because their relationships were not strictly professional. Additional relationship building happened after-hours or during breaks in the school day. As a teacher described her colleagues, she said, “We work well together, we hang out together outside of school, we go to the gym together, go get dinner, so you know, I just feel like having that kind of camaraderie, it helps to make things go a lot smoother.”

One of her fellow teachers on another grade level agreed that, “I get along with my team; we all balance each other out”, and, “We are all good friends, so if you spend a lot time, you want to enjoy your time when you’re here”. Several different teachers named their colleagues as the reason why they remained in their current position, claiming that, “I feel like it’s what kept me here”, and “That’s one of the main things that keeps me here.” These ties to their school community paid off in terms of instructional output, since if other aspects of their life were dragging them down, “You share the load. If someone has a hard time at home, you pick up the slack. And, you know, if someone's not doing their part, you call them out.” This mixture of support and honesty described by teachers across all three schools in this study revealed that such interactions were based on a store of trust that had been accumulated over time.

In addition, some teachers thought about the value of others within their school community more broadly, saying that the sentiment that, “a lot depends on who you’re working with” could be extended to include other staff members, administrators, students, and their families. Being “supported in whatever role by other staff members” was “very

big”. One teacher explained, “Families are part of it too. The concept of the relationship with the family, just the other adults in this building, has been important for me.” Others agreed that, “building those relationships are crucial, especially at the very beginning”. Any type of “long-term staff development” based on “good relationships and people who are willing to share research with you that really matters” was going to support their professional identity development within their school setting. For the 67% of participants who listed leadership as either important or very important to their practice, “I’ve most definitely had administrators I’ve had that I completely admire, and their suggestions really drove me”. For others, “It’s the students... ‘cause it’s obviously not the pay or the hours you put in. It’s that you’re making a difference in their lives and they seem to really love you and care for you. That’s your reward.” Multiple teachers expressed that connecting with children, and the enjoyment they got out of being with them in the classroom, as well as hearing about their success later in life. Experienced teachers especially described long-term relationships with former students as rewards that kept them motivated to put effort into their current practice.

Teachers’ own conceptions of good teaching focus more on emotionally based and relational aspects of good practice than official evaluation standards do. Teachers’ relational focus was one of the key differentiators between personal conceptions of good teaching and the teaching responsibilities laid out with the NC Teacher Evaluation Process. While certain strands of the official standards did incorporate relationship-driven elements, such as leadership standards that mentioned collaboration with colleagues, or requirements related to working with diverse learners, teachers perceived relationships as taking on heightened importance within their own views of their profession. As one participant explained as she looked through the standards, “It’s there, but I don’t know that, really, the

depth of it, and the importance of the way that I perceive it to be important” ends up being reflected within the evaluation process guidelines. In addition, teachers saw evaluation as emphasizing short-term evidence of outcomes, as opposed to long-term goals and lasting impacts on children’s lives that might be difficult to measure over the course of a year. Because “a good teacher has a vision” for their students, helping students to gain basic academic skills was seen as a mere starting point for nurturing a lifetime of human development.

For instance, one teacher explained that in the 180 days she had with her children, “You’ve got to put something in there that will continue and will last. I want to give them the spelling, the writing, and the reading and all that stuff, but to me, what’s important is that they believe that they can do it.” Instead of scrambling to cram in content before benchmark assessments, they wanted to give children “a lot of opportunities to explore”, along with the time to figure things out for themselves, and to be able to, “give them an idea and just say go with it”. When planning, they thought good teachers were adept at “seeing the long-term” instead of just thinking “day-to-day what you want to do”. That way, they could pay attention to informal indicators of growth like children’s conversations and interactions with others while, “knowing where that child is the curriculum and how to build to the next spot”. These teachers were growth-oriented, but wanted to take a more patient approach to learning, citing research that showed it takes “21 days to change a behavior” as part of the need to show children had they had progressed over the course of months, not days. A veteran teacher explained that the impact she had on students sometimes did not surface until decades later, as she recounted the story of a student who had recently reached out to her to let her know about successful admittance to college. In her view, “Whether I had anything to do

with him going to Duke, the fact that he remembered me and was someone important”, was more gratifying than reflecting on the concrete academic skills he may have walked away with during his time in her early elementary classroom.

The inclination to focus on long-term relationships also led the teachers in this study to describe good teaching as being less about content delivery, and more about socio-emotional development and growth. As early childhood and early elementary school teachers, skills like “teaching them some self-motivation of their own and being proud of themselves and what they're doing” were central to their practice. They worked to offer both “choice and autonomy” and directed instruction, so students could develop “independence”, as well as a “belief in themselves”. When teaching, they saw their role as giving children “that confidence” in their ability to make an “impact”, or a “big difference in this world”. Without that healthy self-concept, they believed children would not be able to see, “that they can do this, things are gonna get hard, that doesn’t mean that we quit.” Teachers worried that mechanisms like teacher evaluation would be unable to see that, “The academics is very important, but when you have a child that comes to you and they have been told that you are no good, you're trouble and they hear that all the time, and they're having to work on their, you know, self confidence and all these things and it's just, that is important, too.”

Within their classrooms, teachers described orchestrating harmony between students, as they balanced their time between providing them with individualized attention and thinking about their class as a whole unit. They thought it was crucial “to teach them how to be able to care back with each other and interact. I tell them all the time, ‘We're family in here, you have to treat each other with respect, you need to be kind’, you know just different things like that.” A focus on building “classroom community” was part of their view towards

students as future “members of society”, who they wanted to “contribute back in positive ways”. Participants believed they should dedicate time to modeling the behaviors necessary to become productive citizens, because, “if they know how to treat each other, that’s what’s most important. That they walk out of here, and they stand up to a bully for someone else, or they don’t let someone sit by themselves at lunch”. Ideally, their practice would provide opportunities for children to internalize a sense of whole group responsibility for the health of their school community. In that vein, advocating for their students’ needs meant fighting to retain these crucial elements of their instruction.

The teachers in this study used emotional language to describe the connections they had with students that meant more to them than completing functional tasks. One said:

I would define good teaching as just being there for the students and the families. Sometimes, I feel like there are days where these kids just need to be loved. We may not get through everything in the curriculum we were supposed to, but it's just one of those days where I can tell when they walk in the door they don't need me to be a teacher, they need me to just love them and care about them and give them hugs and tell them, ‘It's okay, you're great,’ and encourage them.

A teacher in another school shared that her approach to the classroom was based on realizing that at times, children simply needed to hear that “We’re gonna love you through this”. She was so invested in her children’s well being, as well as figuring out “what makes each one of these kids tick” that she admitted, “I dream these children at night”. It was important to her for children who left her classroom to know that she valued them for more than progress in a particular subject. Others agreed that it was “not just they’ve scored this. I know what they like, I know what they like to do, and I feel like they know the same about me. And I feel like that why our rooms run well most times, because they're not just students in our rooms, they are a part of your world.” In other settings, teachers explained that developing “that relationship” meant that students knew that they “could come to me if they have an issue,

you know, even if isn't school-related. Within this group, good teachers were described as someone who provided "support to their students" so their "emotional needs" could be taken account. Participants from across all three schools believed that in order to fulfill the extent of their professional obligations, teachers should clearly demonstrate that they cared about their students, liked being with them, and even loved them. They also saw information gathered about students' lives as a tool to make their lessons more relevant, since "knowing where they come from, knowing who they are as a person so I can incorporate that to help support them", was an essential "cornerstone" of their curriculum development. Since these teachers taught at schools with students that they perceived as struggling, some thought that, without a teacher they could trust, "Where else are they going to go?" As a result, their definitions of ideal teaching practices incorporated making these emotional connections.

In contrast, the NC Teacher Evaluation Process does not place nearly as much emphasis on either relationship building or children's socio-emotional development. It also takes a short-term approach to providing concrete proof of student learning, largely hinging evidence of such learning on value-added measures and yearly student achievement results. For instance, while state standards encourage teachers to use a variety of formative and summative assessments, the ultimate goal is to align their work with "21st Century assessment systems" (Public Schools of North Carolina, 2012a, p. 12) designed to "demonstrate evidence" (p. 12) of meeting expected growth targets within tested subjects. Although several participants again emphasized that they did believe good teaching included using data to drive instruction, their views of student success were inclusive of the "whole child". From teachers' perspectives, the evaluation process did not fully capture what inspired or motivated their practice, which generally meant that it did not bring enough

attention to the personalized interactions they were able to have with students, which they saw as essential to capturing student interest and engaging them in instruction. They also believed that the evaluation system defined good teaching as completing a “checklist”, as opposed to thinking about teaching in more responsive, fluid ways. Because the descriptions of model practices within the NC Teacher Evaluation Process are much more scientific and systematic in nature, they are not as focused on the “art” of teaching, which teachers described as orchestrating a series of intentional actions that would move individuals forward while simultaneously considering their class’s group dynamic.

Overall, because teacher evaluation policy was a relatively insignificant part of teachers’ professional identity development and conceptions of good practice, there was a disconnect between intended policy impact and ground-level reality. The aspects of their teaching that participants truly cared about, which tended to be more relational, were not fully encapsulated within performance standards. Since evaluation did not always provide teachers with meaningful, specific feedback or extended instructional support, it was not a major influence on their day-to-day decisions and interactions with children.

CHAPTER SIX

IT'S ALL RELATIONAL: IMPLICATIONS FOR THE FIELD

Teachers' experiences with evaluation policy in North Carolina were indicative of trends within a broader policy climate that favored a focus on teacher effectiveness and educational outputs (Ladd & Fiske, 2014). The impetus for widespread adaptation of teacher evaluation standards was initially spurred by federal messages about the need for evaluation reform that explicitly held teachers responsible for their students' learning. These messages were locally reinforced by North Carolina's introduction of an updated evaluation process, which was similar to the performance management plans put in place by several other RttT winners (U.S. Department of Education, 2010f). However, state alignment with federal objectives was not enough to ensure that evaluation would reach its full potential.

On one hand, North Carolina policymakers designed evaluation measures that echoed the federal agenda. Since the U.S. Department of Education believed that gathering more precise information about teacher performance would combat perceived lags in teacher quality and inequitable student achievement outcomes, RttT funding awarded to winning states was earmarked for implementing this solution. In North Carolina, discourse about the purpose of evaluation was similar to Secretary Duncan's lofty goals to increase student achievement by standardizing modern teaching expectations and improving instructional capacity. Duncan's push for greater transparency and accountability was also reflected in the state's efforts to centrally collect and analyze performance data. Furthermore, the individual

components of the official NC Teacher Evaluation Process were heavily focused on common systemic goals like continually driving measurable improvement.

Yet, while districts, schools, and teachers were willing to comply with evaluation procedures, evaluation implementation often fell short of its promise. Despite the stated purposes of the NC Teacher Evaluation Process including goals like fostering reflection and personal growth, it is primarily intended to “monitor” teachers and “serve as a measure of individual progress” (Public Schools of North Carolina, 2012a, p. 5). These emphases also tie evaluation feedback to consequences like being placed on probation or being denied career status. Although North Carolina’s evaluation procedures imply that the lowest performers will receive follow-up from administrators and mentors, this support is largely framed as a supervisory mechanism for checking up on teachers’ progress, instead of professional development. This makes supplementary goals like offering guidance and support to *improve* the effectiveness of teachers at *all* levels of performance less obvious within local implementation priorities.

Teachers’ Overall Impressions of Evaluation’s Influence

How, and to what extent, does teacher evaluation policy influence teachers’ conceptions of good practice and professional identity development? As was revealed in Chapter Five, the high importance placed on rigorous evaluation standards at the federal and state level did not fully align with local teachers’ own values, beliefs, and goals. Looking at teachers’ historical relationship with accountability policies in a post-NCLB era, research often reveals strong tensions within the fallout from such mandates that decrease teachers’ job satisfaction (Barrett, 2009; Berryhill, Linney, & Fromewick, 2001; Day, Flores & Viana, 2007; Labaree, 2011; Loeb, et al., 2005; Ma & MacMillan, 1999; Ravitch, 2012; Sachs,

2001). However, in this case study of teacher evaluation, participants' attitudes towards having their practice assessed were less definitive. The teachers' tendency to place teacher evaluation in the middle of the influence spectrum reflected an understanding of evaluation's potential benefits, a desire to receive more directed feedback, and a willingness to comply with state requirements. Yet, teachers also felt that evaluation fell flat when it came to making a concrete impact on their conceptions of good practice. In contrast to federal and state goals of evaluation policy impacting instructional quality, it was hard for them to identify discernable differences in their teaching that had stemmed directly from the evaluation process.

When teachers were asked to detail why evaluation was only somewhat influential, they listed several reasons why North Carolina's performance management measures often failed to live up to their potential. These reasons included 1) flaws in implementation, 2) little sustained connection between evaluation ratings and improvement strategies, especially for middle performers, and 3) the relatively average impact of evaluation as compared to other important influences on their teaching. First, implementation challenges that detracted from how well evaluation was being carried out drove the majority of mixed reviews about evaluation reform's ability to impact teachers' practice. Participants explained that the complexity of capturing snippets of live teaching practice in ways that are fair and comprehensive posed a logistical roadblock for administrators, who did not have the time to carry out evaluation for all of the teachers in a school while simultaneously aligning observation feedback with professional development efforts. In addition, as teachers who typically received proficient or better ratings, they could coast along without being pushed to improve, therefore restricting evaluation's intended purpose. In some ways, the average to

good performer suffers the most under a system in which the few truly terrible or exceptional teachers will either be dismissed or rewarded, while the mass of teachers in between those two extremes is left to continue along without much motivation or support to take their practice to the next level. Teachers in this middle space emphasized that it was important to provide teachers with feedback that would push them to grow and ensure that their students received the best education possible, instead of keeping teachers around who were “just here to collect a paycheck and go home”. However, the utility of the feedback they received was inconsistent, and execution and follow-through were lacking. Lastly, teachers also described a large variety of other influences on their teaching, including prior schooling, personal beliefs, and relationships with colleagues, which had longer-lasting impacts than they felt a single policy like evaluation ever could achieve. These relatively more important variables largely reflected the value that teachers assigned to sustained connections with prominent figures from their past and present educational communities (Cohen & Ball, 1990; Feiman-Nemser, 1983; Sloan, 2006; Spillane, Reiser, & Reimer, 2002).

Major Themes from Across Teachers’ Evaluation Experiences

Taking a comprehensive look at this study’s findings, a number of themes carried across distinguishing characteristics like teacher experience levels, prior background, and school setting. First, teachers’ personal orientations towards evaluation did not exist in a vacuum. School context and culture did matter when teachers were asked to consider how their working environment affected their approach to the classroom, as well as their attitudes about being evaluated. Some contexts for teaching and learning were more conducive to evaluation than others, either because they were better equipped to combat potential implementation challenges, or because they had pre-existing good will with teachers that

administrators could fall back on. If structures were already in place for channeling instructional guidance, evaluation flowed nicely into that norm. Throughout the data, teachers mentioned that when their schools created avenues for low-stakes, informal exchanges with administrators, these continual interactions contributed to the evolution of their practice.

Within their educational community, the importance of personal relationships was the most central to teachers' experiences with evaluation policy, as well as their conceptions of good practice and overall professional identity development. Throughout our conversations, teachers repeatedly mentioned inspirational individuals who had impacted their practice and continued to move their instruction forward. The foundational trust and respect that characterized these key relationships were the most commonly identified drivers of successful implementation of evaluation components like post-observation feedback delivery. In general, participants reported that the quality of these relationships either prompted genuine investment in evaluation as a professional growth tool or drove diminished views of evaluation as a chore that they needed to accommodate. Consequently, maintaining meaningful bonds with others within their school community was central to driving teachers' instruction forward on a variety of levels.

The Overarching Importance of Relationships

The central role that relationships with others played in teachers' lives and conceptions of good practice carried throughout their entire educational careers. Because this study asked teachers to describe influences on their practice and compare their impact to evaluation, the resulting data captured a trajectory of education-oriented relationship building that spanned from teachers' own schooling to their present-day employment. Participants

recognized that influential past teachers had served as their first instructional role models, and at times even compared unsuccessful educational experiences with classrooms where their desire to learn had been stimulated. They also described their best former teachers in socio-emotional terms, typically talking about the impact that they had made on their confidence or character, rather than their academic progress. When characterizing influential aspects of teacher training, participants were similarly inclined to pinpoint individuals like cooperating teachers instead of mentioning programmatic features like coursework. They had appreciated building connections with other educators during field-based training, and felt that the “essential” skills gained during opportunities for observation and reflection during those time periods had launched their careers in a positive direction. In addition, teachers reported that family members contributed through offering support, establishing their character, and serving as examples of ethics like hard work, dependability, and compassion. The beliefs they had developed about the role of education when they were younger, as well as their attitudes towards helping others, made them want to maintain a “positive outlook” and remain in a service-oriented profession. These formative experiences were centered on deeply rooted connections with people who gave their personal and professional lives meaning. Within these relationships, the personal and the professional were often intertwined, as evidenced by the numerous occasions when participants connected their feelings about these key individuals to their current impact on their teaching practice.

Once these teachers were at the head of their own classrooms, their reliance on meaningful connections with others only increased, as they soon realized that their work environment was only going to be as strong as their relationships with their colleagues. Whether they were turning to their colleagues for instructional support, sharing lesson plans,

or seeking encouragement, they relied on these relationships to help them overcome teaching's challenges and feel less "alone in all the struggles and triumphs". Having this type of collegiality within their work culture also made these teachers more likely to want to reach out to others "on the same playing field" and offer their own expertise as a resource, making collaborative connections. One teacher explained that it would be difficult for a generic evaluation rubric to have the same value as advice from an experienced colleague down the hall who she looked up to and respected enough to be able to say, "Hey, I'm struggling in this area. What do you do?" Her fellow teachers offered her quick, easy access to reliable guidance, which she had faith in, because she saw evidence of their accomplishments with children on a regular basis. Others reiterated that a "family-like" culture within their workplace enhanced their ability to develop their pedagogical capacity and "share data" within an environment where they could be "fully exposed" and feel "okay with that". Their colleagues had high value when they were serving as real life models of what good instruction looked like, and when, as another teacher related, they had a bad day and just needed to go sit in someone's rocking chair and vent.

Of course, establishing a certain comfort level with their administrators was another piece of the school culture puzzle. Participants immediately recognized that although their school's principal was just one individual, the power that principals had to affect their school environment was significant. As a result, teachers identified their interactions and relationships with administrators as highly relevant to teacher evaluation policy and professional identity development. Because evaluation ratings given by an administrator are at least somewhat subjective, believing in the general validity and accuracy of these ratings drives how much value teachers will assign to these judgments of their practice, and whether

or not they will “listen and take to heart what they say”. It was clear that in situations where, “There’s a little bit more driving force...because you respect and admire and have a great relationship with that person,” administrators had the potential to provide a new perspective, give useful advice, and even serve as inspiration. However, “if there’s an administrator that you just don’t have that same relationship, or you don’t feel their suggestions are coming from a positive place or trying to help you...” the evaluation process was not going to unfold as smoothly. A teacher gave an example of how in previous years, she had never taken the professional development planning component of the evaluation process very seriously, treating it as an administrative requirement that she would obediently complete but rarely reference or use. However, when her new principal spent time individually talking with her about it and seemed invested in her progress, she felt re-energized and “complete”. This teacher explained that it was a significant milestone in a series of good signs that this principal cared about her ongoing development and would partner with her to ensure that she succeeded. This was especially refreshing to her as an experienced teacher who might have otherwise seen herself as having hit a professional plateau. Instead, the new principal’s focus on relationship building drove positive feelings that propelled her forward even more than the specific advice she gleaned from the meeting.

Whatever their affinities for the other adults in their school looked like, the participants in this study saw teaching itself as an inherently relational task. On a practical level, they wanted to learn more about students’ backgrounds, interests, and preferences in order to inform their curricular planning and engage students in lessons. They made an effort to reach out to students and their families because such partnerships facilitated their instruction. However, they also noted that daily exchanges with students to guide their skill

development required constant personal interaction. Because teachers spend many hours each day with their students over the course of nine months, getting to know each other is an inevitable and important part of their job. As students enter school, they begin their transition from the relative isolation of home to having to cooperate and collaborate with other children on a regular basis. This means that PreK-3 teachers have to dedicate time within the instructional day to fostering social development, which several participants noted during conversation about what good teaching looked like, as well as when making suggestions about what else evaluation standards needed to encompass. Particularly in the early elementary grades, teachers saw their role as reaching beyond academics to include promoting good citizenship and exemplifying how to interact with each other within social environments. Across the findings, teachers emphasized that the familiar, everyday connections with students and their families were what motivated them to invest in improving their practice and continue to contribute to the surrounding community. They also hoped that evaluations of their practice would capture this part of their teaching, and provide them with feedback that felt as authentic as their dialogue with the children in their classrooms. However, if it did not pan out that way, teachers said, for instance, that as long as “my children want to be here, my children are successful, that’s all that matters” in the long run. Even though they wanted to perform well, as one teacher put it when describing what motivated her teaching, “I don’t do it for recognition. I do it because it’s my style, and I do it for my children, and for myself as a professional”. In this teacher’s school, a feeling that people “love each other, we care about each other”, combined with her dedication to her craft and her students, provided her with all the motivation she needed. Throughout this study, multiple participants noted that the reoccurring conversation about the role of positive

relationships with students, colleagues, administrators, and other community members signaled their central importance. Because these relational supports meant so much to teachers, it was difficult for a policy like evaluation to have an equally significant impact.

Evidence from Prior Research that Relationships Influence Teaching Practice and Policy Implementation

This study indicated that there are many relational and social inputs that affect a teacher's identity over the course of their professional identity development. For example, of 16 potential positive influences on their practice listed for participants to rate during the focus group activity, teachers selected an average of 8.6 factors to designate as either "important" or "very important". Many of the most frequently cited influences were person-driven, and involved interactions with their families, mentors, colleagues, administrators, and students. The evidence within this study that forming conceptions of good practice is a multi-faceted endeavor is confirmed by research on professional identity development that discusses how the multiple components of teachers' personas contribute to views of their role (Beijaard, Verloop, & Vermunt, 2000). Within socio-cultural institutions like schools, teachers' identities are shaped by the connections they make as members of their educational community (Gee, 2000). For these participants, ongoing connections with others fostered dialogue about what they should be doing in the classroom. Clearly, as teachers construct their own views towards their professional responsibilities, a variety of relationships help inform that process.

Confirming the importance of past relationships. Literature reminds us that the introduction of any new initiatives will have to make an unusually strong impact to spur any real difference in teachers' approaches to the classroom. This is partially because person-based, past influences have such a significant, lasting impact on teachers' perceptions of

what good practice looks like, which makes it hard for those established notions to be disrupted (Cohen & Ball, 1990; Feiman-Nemser, 2001; Lortie, 1975). The teachers in this study confirmed that experiences from decades ago stemming from their own experiences in school, their initial teacher training, and guidance from family and friends had made a considerable positive impact on their careers, and often helped shaped their personal beliefs about the purpose of working with students. The idea that teachers' own views of their classroom responsibilities, as well as personal values like work ethic or acceptance, could affect teachers' professional identity development is also echoed by an accumulation of generalized evidence about the sustained effort and time needed to enact lasting change in schools and shift teachers' attitudes, beliefs, and orientation towards instruction (Hargreaves & Fullan, 1992). As this study revealed, teachers' deeply rooted personal beliefs were formed by early experiences in their own lives and affected how they saw and worked with their students.

Confirming the importance of leaders. It is no surprise that once teachers are placed in schools, relationships with leadership personnel and other colleagues matter to teachers. Principals have long been identified as a key player in the success of their schools, as well as a major force behind establishing school culture (Leithwood, Steinback & Jantzi, 2002; Leithwood & Montgomery, 1982). Because principals set the tone for attitudes towards policies like evaluation, their impact is greater than simply enforcing state directives and ensuring that every teacher receives performance feedback. Evidence shows that school leaders are important drivers of stability and direction, which they can provide for teachers in situations like when a new initiative is introduced into the educational system (Hogg, 2001).

Confirming the importance of colleagues. Similarly, colleagues such as more experienced teachers often have guidance to offer teachers that can help them determine how to best work with their students and navigate new policy requirements (Feiman-Nemser, 2003). When teachers can build relationships with other educators who provide honest but caring feedback about their practice, they feel safe to try new instructional methods and reflect on their practice (Cochran-Smith & Lytle, 1999; Lord, 1994). As this study confirms, a teacher's relationships with other adults in the school building are a central part of his or her working environment. Participants saw the significance of these fellow educators as both affecting how evaluation would be executed and contributing to their broader growth and development. Their colleagues were important drivers of their overall impressions of their school culture and context.

Evaluation as Filtered Through School Context

In *Appendix A*, the path from initial evaluation policy implementation to influencing teachers' views of practice is shown as being filtered through school contexts. Evidence that context and working environment mattered proliferated participants' descriptions of how a broadly conceived state evaluation process was implemented at their schools. Veteran teachers who had been a part of efforts like developing district-level curriculum guides explained that they understood how original policy or program intention did not always translate from the written page to practice. As such, they could give several examples of times when widely distributed resources and tools that had been conceived with practitioner input and carefully considered by experts failed to have the desired impact once they were applied across a diversity of school environments. For instance, a teacher described a supplementary curriculum that was supposed to foster positive behavior and a better school

culture – but asked teachers to focus on establishing classroom rules and norms in January, instead of the logical choice of August or September. Perhaps there was some rationale for this odd chronology of lessons, but whatever the reasoning was, it was not clear to her on the ground level. In the same vein, the purpose of evaluation policy ran the risk of being misunderstood by both principals and teachers who were unable to fully connect stated objectives like using evaluation to foster professional development to the realities of their school environment. Because evaluation was subject to both implementation flaws and varied interpretations of its role in teachers’ lives, it ran the risk of becoming watered down. For instance, as one teacher described, a meeting with a principal about their ratings might mean simply hearing, “I signed it. Thanks for all you do!” and then moving on to the next requirement. Therefore, most participants saw teacher evaluation as only partially effective.

Context also mattered to these teachers because evaluation, an accountability measure, was being rolled out during a time in North Carolina’s political history that was generally not perceived as being friendly towards teachers. It was hard for them to view evaluation feedback as useful when it seemed more like just “one more thing” to “check off” that would not only take up time, but place teachers in the position of being unduly judged and criticized. While these teachers claimed that they tried to steer clear of negative news about local education, they also admitted that they did have concerns about how evaluation ratings, which only captured a glimpse of their classroom activity, might be used against them. In many cases, they trusted their school and their local community to be generally supportive, but could not completely shut out the rumors about using performance management to determine limited eligibility for bonuses or career status. Participants asserted that teachers were “smart people”, capable of solving instructional challenges, but

feared that the educational climate in North Carolina did not lend itself towards giving them credit for those abilities. Even if they were currently “safe”, in their present school situation, teachers made comments like, “I feel like there are probably administrators out there that do not give their teachers a voice and do not make them feel empowered”. Consequently, within this context, evaluation mechanisms that included many growth-oriented components like developmental rubrics, opportunities for teachers to document their practice, and required pre- and post-observation meetings with administrators, ended up being perceived as heavily focused on end results and numerical ratings. For example, state compilations of evaluation data are referred to as reports of “educator effectiveness”, emphasizing outcomes rather than the instructional process (Public Schools of North Carolina, 2014). In some schools, this ethos towards performance management permeated its implementation. In others, teachers who were relatively unconcerned about the direction of evaluation policy credited positive characteristics within their school environment for making performance management a more holistic endeavor. From across the data, these contextual boosts to evaluation implementation included a history of trust and respect between administrators and teachers, positive encouragement, regular opportunities to meet with administrators and coaches, extended knowledge of each others’ classrooms, close collaboration sustained over time, and structures like targeted professional development support that allowed evaluation feedback to better connect to instructional decisions.

Evidence from Prior Research that Context Influences Teaching Practice and Policy Implementation

Research indicates that the quality of implementation at school sites is key to determining the nature of the relationship between policy and practice (Cohen & Spillane, 1992; Crossley & Vulliamy, 1995; Fullan, 2001; Honig, 2006; Spillane, Reiser, & Reimer,

2002). As a result, it only makes sense that while baseline evaluation standards and procedures were in place across each of the three schools in this study, the differences in each school environment caused variation in evaluation experiences. When teachers felt invested in the evaluation process, and well connected to their administrator's goals for their progress, they were able to find utility in their performance being measured. In contrast, if they felt as if their school was merely enforcing evaluation as part of accountability measures designed to force teachers to prove their worth, implementation was less effective (Cohen & Spillane, 1992). In addition, this study's depictions of on-the-job influences as majorly affecting how teachers perceive their role in the classroom was affirmed by prior evidence that classroom experiences and interactions with others in the field significantly contribute to teachers' professional identity development (Harris & Sass, 2011). As teachers encountered new evaluation policies and standards within their school environment, they were forced to compare prior conceptions of teaching with current expectations (Spillane, Reiser, & Reimer, 2002). Even when their own views of teaching were not in conflict with evaluation's basic guidelines, the pressures imposed by the broader educational climate drove participants to see evaluation efforts as enforcing an emphasis on end results at the expense of attention to daily instructional processes.

Comparing Major Trends in Teachers' Experiences to Evaluation Rhetoric

In the following section, I compare teachers' views towards evaluating important job responsibilities to federal and state level discourse about the role of evaluation in establishing teaching expectations. I use this comparison to highlight disconnects between policy rhetoric and real-life teaching experiences. Major themes include teachers' emphasis on relationships

rather than business-like management strategies, as well as their inclination towards customization rather than uniform standardization of practice.

Relationships versus management. In contrast to teachers' focus on the relational aspects of teaching, federal level rhetoric surrounding teacher evaluation purposely tried to depersonalize such strategies and instead link their benefits to common sense, business-like reasoning. Because Secretary Duncan had to convince states to make major changes to their existing evaluation practices, he wanted to make those shifts seem like necessary next steps that would only enforce the positives of the current status quo, and rightfully eliminate the negatives. In describing an evaluation system that would either reward or rehabilitate teachers, he categorized teachers as either good or bad performers, whose primary responsibility was to "accelerate student achievement" (Duncan, 2010a, p. 1). Since "gut feelings" (Duncan, 2010f, p. 1) were no longer enough to guide teachers towards improvement, more rigorous documentation of their practice, as well as their results with students, was essential. Human connections between supervisors and their employees were all well and good – but this process needed to be systematized, so that "tracking" (Duncan, 2010g, p. 1) student growth and holding teachers accountable for learning would become standard operating procedure. That way, instruction could be driven by an accumulation of "assessments" and "longitudinal data" (Duncan, 2010h, p. 1). Furthermore, while Duncan did use emotional language to describe the inspiration that students could draw from teachers who provided the "vision" for a better future that would "change the course of a student's life" (Duncan, 2010d, p. 1), he emphasized that teachers needed to have "a single-minded focus on improving student learning" that had more to do with data than dialogue (Duncan,

2010h, p. 1). In his mind, “information” (Duncan, 2010g, p. 1; 2010h, p. 1) would drive student progress – not warm and fuzzy feelings.

Similarly, state-level evaluation guidelines issued by North Carolina provide one generic sense of what it means to be a quality teacher that includes some relational aspects, but makes it clear that the bottom line of using “data” (Public Schools of North Carolina, 2012a, p. 8) to increase student achievement is more important than goals like developing classroom community. In other words, if teachers are not found to be “effective” (p. 12) at boosting their students’ learning, as defined by a variety of growth mechanisms, their career status will be in danger. Even though the state currently assigns a reasonable weight to value-added measures and considers a range of data about a teacher’s practice, there was a perceived possibility of principals, districts, and state-level officials over relying on test scores as indicators of teaching success, rather than taking their personal knowledge of a teacher and their working environment into account. While it may seem logical to conclude that if teachers cannot fulfill this basic responsibility, they should not be in the classroom, teachers in this study argued that because teaching is a complex endeavor, it cannot be boiled down to a single assessment or even a series of ratings. In addition, although teachers in North Carolina receive credit for aspects of their practice like collaboration with colleagues, developing connections with diverse students, and serving on school-wide committees, language focused on driving learning outcomes and providing “evidence” (p. 5) of learning was more common than language used to describe the relational side of these efforts.

As a result, while teachers in this study liked that the state evaluation mechanism attempted to gain a more complete view of their practice, they knew that at the end of the day, what really mattered to schools were numerical depictions of their end results in the

classroom. While participants agreed there was a place for assessments of their work, they did not feel that the current mechanisms for evaluating teachers were going to lead to maximum growth opportunities, especially considering that policymakers and administrators were “still trying to figure out all that stuff”. Participants explained that the combination of tying “students’ performance to whether or not we even have a job”, state-level chatter about a proposed plan to only benefit the top “25% of teachers” and having to “prove ourselves” to administrators created further uncertainty and pressure, rather than fostering collaboration and teamwork. For instance, some North Carolina districts have made publicly searchable average evaluation results available by school, while other districts have plans to post individual evaluation information that participants feared would serve to “pit teachers against teachers”. Especially in light of the state’s recent efforts to use evaluation ratings to only offer extended contracts to top performers, teachers felt that a managerial, bottom-line approach to evaluation was becoming more dominant (Ladd & Fiske, 2014).

Customization versus uniform standardization. Teacher evaluation rhetoric struggles to balance large-scale standardization with involving individual districts and schools in implementation decisions. When Secretary Duncan discussed evaluation policy shake-ups, he tried to assure his audience that determining exactly how to measure teachers’ performance would be “a local decision” (Duncan, 2010g, p. 1). He described how schools formerly under his purview in Chicago were about to “transform a school culture” by working to together to decide how the “common mission” of boosting student learning should be accomplished (Duncan, 2010j, p.1). Within his descriptions of his preferred model of tracking teacher effectiveness, Duncan claimed that “local leadership can also choose from several different approaches to transition student achievement” (Duncan, 2010g, p. 1), taking

community input and their particular school contexts into account. His rhetoric painted a vision of district leaders being heavily involved in collaborating with union officials, teachers, and other stakeholders to figure out how to apply broad, federally driven standards. Similarly, Duncan described choices about how to reward, recognize, or penalize teachers according to their performance ratings as being “local” in nature. While guidelines for enforcing teaching standards would be provided, states and districts would be able to decide whether or not they should be linked to salary or staffing decisions. In addition, multiple data points like observation commentary would form a more complete picture of what teacher performance looks like within unique settings for teaching and learning. At the same time, Duncan asserted that schools across the country needed to adapt a uniform approach to teacher accountability that would provide “meaningful teacher evaluation systems” (Duncan, 2010a, p. 1) that directly relate teacher performance to “tracking student growth” (Duncan, 2010g, p. 1). Once again, making adaptation of this process seem smooth and simple, he did not mention any possible implementation glitches.

Along the same lines, throughout North Carolina’s evaluation guidelines, context was most clearly addressed within the standard that assesses a teacher’s ability to “establish a respectful environment for a diverse population of students” (Public Schools of North Carolina, 2012a, p. 9). However, the state’s attention to classroom-based contextual considerations, such as “different points of view”, “personality”, “special needs” and children’s “culture and background”, did not appear to extend to a consideration of how evaluation implementation might vary according to school setting or teachers’ own learning preferences (p. 9). Because the official NC Teacher Evaluation Process was designed for uniform distribution across the state’s districts, adhering to its procedures and processes is

simply presented as mandatory. Teachers could only hope that within their particular school environment, their principal would customize state guidelines to provide the best support possible, perhaps even finding a way to make feedback more inspiring, instead of simply having to “roll with it”. Their common refrain was one of valuing more frequent, instructionally focused, informal visits to their classrooms, followed up by individual face time that fostered regular conversation about relating professional development to their particular students’ needs. These low-pressure, context-specific experiences had more utility for them than instances when teachers reported, “I feel like they hear what we’re saying, but they’re not really listening”, because evaluation had become one more overwhelming burden to endure. In general, teachers thought the generic nature of the official evaluation process made it less valuable than more targeted forms of support.

Considerations for Policymakers and Practitioners

This study’s themes may prompt reflection from members of the education system who are involved in implementing recent updates to teacher evaluation systems. Although most teachers in North Carolina were not formally evaluated according to the official NC Teacher Evaluation Process until 2010-11, local schools are at the point where their initial implementation of teacher evaluation is ripe for review. This is especially critical at a juncture where observation rubrics have been uniformly unveiled, but recently updated components like value-added metrics are just being introduced to early elementary teachers. In addition, teachers have now been evaluated for two to three years in a row, providing them with a sense of how well the process has worked within their school context. Since these teachers can readily identify implementation factors that have contributed to evaluation’s efficacy, they have valuable insights to share. Participants’ unanimous characterization of

authentic, ongoing relationships as the building blocks for meaningful conversation about practice suggests that it would be useful to further explore the types of behaviors and structures that drive those vital connections. Consequently, this study has several implications for policymakers and administrators who want to improve teachers' evaluation experiences and better link the NC Teacher Evaluation Process to teachers' conceptions of good practice and professional identity development. Since the teachers in this study were receptive to instructional advice, but did not think that their evaluation comments were consistently useful, that leaves the door open for feedback structures to be improved. In addition, other teachers in the field might benefit from considering how to make the most of their evaluation experiences in collaboration with school leadership.

Recommendations for policymakers based on teachers' feedback. Policymakers and state agencies like North Carolina's Department of Public Instruction (DPI) are the primary forces behind enacting teacher evaluation across the state. For instance, DPI established uniform evaluation standards, along with guidelines for implementation and then trained principals and teachers on how to use them. Yet, teachers' reactions to those standards were only "somewhat" positive. As the findings of this study revealed, while teachers have not dismissed these baseline indicators of good practice, they would not classify them as being truly influential. The standards themselves were described as being too generic and open-ended, making it difficult to connect feedback generated from evaluation rubrics to explicit instructional decisions. This was especially true for early elementary teachers who worried that the general rubric did not entirely speak to their children's developmental stages. Broad terms like "global awareness" or "21st Century skills" need to be better defined for them, ideally by incorporating examples of how they

would apply to younger learners. Because of this, future editions of the NC Teacher Evaluation Process might be best enhanced by adding supplementary information specific to each age group. While on some level, administrators working across diverse schools need rating categories to remain applicable to a variety of settings, these categories also need to hold meaning for teachers in order to have a full impact. With that goal in mind, agencies like DPI could also better articulate how the evaluation process relates to potentially complementary requirements like following the Common Core State Standards. A lack of alignment between the multiple demands on teachers dilutes evaluation's effectiveness, as teachers do not know which direction to turn in first.

In addition, policymakers need to carefully consider the trade-offs when planning to make evaluation ratings a factor in high-stakes decisions. Teachers reported that they were more likely to incorporate feedback driven by relationships with administrators who they could trust to provide guidance without being overly evaluative or judgmental. The comfort level gained by being observed by an administrator who "knows me", and who presented evaluation as just one stepping stone within a teacher's professional trajectory, was invaluable. Because the current political climate in North Carolina is seen as unsupportive of teachers, the participants in this study were particularly sensitive to being "blamed" for circumstances that they were not entirely responsible for creating, and described a "big anxious feeling" about what might lie ahead. Given their belief that the majority of their colleagues were not going to perform poorly enough to warrant removal from the classroom, what purpose would distinguishing between levels of performance serve? These concerns highlighted the idea that if evaluation is seen as a tool for professional development, rather than as "punitive", it will have a better chance of resonating with teachers. Along those

lines, although North Carolina has taken a well-advised step by only making value-added growth one-sixth of a teacher's overall rating, and holding back on incorporating it in new teachers' reviews, many teachers are confused about what incorporating these models will mean for them. As a result, the pressure that already naturally results from being observed and judged may only intensify with the inclusion of test-based metrics. Furthermore, there are flaws within even the best value-added models that teachers themselves recognize as issues, such as the non-random nature of class assignment, the variable growth rates of children over time, and the potential downsides for teachers who take risks like moving to a different grade level or working with challenging students.

Overall, policymakers need to assess how evaluation resources are being spent, since if such policies are unable to make a real difference in how teachers approach the classroom, and any positive benefits are seen as “random”, evaluation in its current form may turn out to be an ill-advised investment. Teachers expressed that especially if they were steady performers, but continually maxing out at “accomplished”, there was only so much more insight that they could gain, and motivation that they could maintain, from trying to advance along the growth continuum. To strengthen evaluation's impact, policymakers should pay more attention to implementation factors within school contexts that can either boost evaluation's potential or diminish its impact. This study showed that ratings further lost their ability to drive change when they were not personalized to reflect each teacher's abilities within their context for learning. Incorporating lessons learned from the field, now that administrators and teachers have each had the chance to see how more rigorous evaluation has unfolded in their schools, could refresh initial training given to principals. District

personnel who supervise administrators might also play a part in increasing evaluators' capacity to deliver meaningful performance feedback.

Recommendations for administrators based on teachers' feedback. As the primary players in evaluation implementation, school leaders have a clear role in determining the success of the NC Teacher Evaluation Process. Teachers were able to give several examples of when administrators were able to provide them with useful feedback that they incorporated into their practice, as opposed to other instances when the evaluation process felt like more of an administrative task, or a “hoop to jump through” than a significant tool for growth and development. The most positive experiences with evaluation were the result of groundwork by administrators to develop trust over time. When administrators could remember details of their children's lives, or immediately recall that a classroom's dynamics included special needs students, they indicated to teachers that they knew their classrooms well enough to assess them. In addition, follow-up conferences became less intimidating when teachers felt that administrators were being authentically supportive, instead of taking a “gotcha” attitude towards evaluation. Even if teachers had a friendly relationship with their principal, when evaluation became a “game” of providing enough evidence to be able to justify ratings, such tasks were reduced to a mere fulfillment of requirements. Teachers recounted that when principals “shared” information with them and provided manageable, concrete chunks of advice, they then became less defensive. They liked to be pushed and challenged, but through productive questions about “the process or how I had gotten where I'd gotten” that would help them think through their instructional intentions. Recognizing strengths and highlighting them within the school community was another way to boost teachers' trust and make them more receptive to the flipside of hearing critiques of their

practice. Participants found that if these bonds could be sustained over time, the rushed nature of most evaluation ratings could at least be counteracted by those foundational interactions.

Because teachers find feedback to be the most relevant when it is immediate and specific, administrators should also approach evaluation from the position of an instructional leader, who can deliver guidance in real time. Schools that assigned administrators and coaches to particular teachers and charged them with making regular, informal visits to teachers' classrooms, followed by directed one-on-one conversation, had the best chance of making performance management meaningful. Since fulfilling all the stated purposes of the evaluation process and balancing "monitoring" teachers with mentoring them would be difficult for any one individual to execute, it made sense to divide and conquer so administrators could focus on the smaller group of teachers assigned to them. Although teachers at the one school in this study that incorporated peer observations reported some hesitation to formally rate their colleagues, especially if they did not have knowledge of the grade level that they were assigned to observe, allowing teachers to informally observe each other could be another way to increase dialogue about instructional practices. Because time limitations may still cause logistical issues, administrators might also consider whether other designated meetings would increase in value if they were used to address common areas in need of improvement. By linking the purpose, for example, of a professional development session, to other requirements like those within evaluation rubrics, administrators could deliver a more unified message. Such efforts would help connect the large number of mandates, policies, and programs that teachers are subject to each year. Otherwise, teachers reported that it was hard to see how their evaluation experience aligned with the already-

crowded marketplace of instructional ideas that they were being bombarded with on a daily basis.

Lastly, but perhaps most importantly, administrators should recognize that they have valuable sources of input about evaluation implementation at their fingertips. Since schools already contain experienced teachers who could offer ground-level perspectives towards the effectiveness of evaluation policy, it would be a shame to overlook their insights. Feedback might be the most productively channeled through designated leadership teams of teachers with an interest in informing implementation procedures. If these interested parties were well regarded by their colleagues, they could then serve as ambassadors for new ideas, consequently creating buy-in across their school. Participants in this study indicated that ideally, administrators would balance enforcing a policy with being flexible about exactly how to carry it out. A veteran teacher explained, “It would be nice if they just asked us what we thought sometimes” before introducing a new initiative. Once the hurdle of initial buy-in is cleared, leadership teams may also need to revisit implementation efforts to sustain a demonstrated purpose for policies like evaluation in the long-term. Teachers at one school in the study were able to say, for example, that at least their administrator had recognized that she needed to be more pro-active about explaining how the incorporation of value-added metrics would work. This at least gave teachers hope that, “the way that administration is already looking at it, and talking about it, and meeting with everybody about it, maybe it’ll make that transition a little more smooth”. Keeping teachers well informed during periods of change was vital to successful future implementation of new evaluation mechanisms within individual school contexts.

Recommendations for other teachers based on these teachers' feedback. This study may also have implications for other classroom teachers who are trying to figure out how to maximize evaluation feedback. Evaluation often seems like something that is happening to teachers, instead of with them. Although it is understandable that those in supervisory positions would have to avoid letting personal relationships and emotions get in the way of sound management, the teachers in this study still wanted to be able to engage with administrators to shape instructional decisions together. By leveraging positive connections with their evaluators, teachers could have conversations about their classroom that would build dialogue while taking their expertise into account. Because these teachers taught in schools that provided a decent amount of autonomy in the classroom, their experience working with children was typically respected enough to make their justification of pedagogical choices a smooth process. However, participants also provided examples of ways that they had been proactive about trying to maximize their evaluation experience, so they would not even find themselves in the position of having to treat evaluation solely like an exercise in documentation.

Teachers may have to take steps to enhance standard evaluation procedures by requesting more one-on-one time with administrators, asking for time to observe other teachers in action, inviting principals into the classroom to interact with students outside of required visits, and making concrete suggestions for ways to make evaluation implementation more effective. A teacher who had been in the classroom for a number of years explained that she still wanted to continue to grow. Therefore, it was important to remain open to others' viewpoints and push evaluators to give feedback that was directed enough to be useful, even if it meant pulling her out of her comfort zone. Because evaluation

implementation was subject to attitudinal responses, another teacher noted that although it had taken her a long time to reach that point, “if you are a person who needs approval...those evaluations tend to drive, or you know, can upset your emotions”, but “if you are one of those that just wants to be better because you want to be better for your students”, evaluations could be seen as part of continual development. Another participant agreed that it “all comes back to motivation”, meaning that the evaluation process would become whatever teachers made of it. Depending on whether teachers were willing to “dive in”, evaluation could either remain “just a piece of paper”, or something more instrumental.

Other teachers at a school being trained on peer observations informed administrators that they needed more details about how to apply some of the concepts within the state evaluation standards to younger grade levels. As a result, administrators spent time with them describing what fulfillment of each standard would look like in action during those years, which helped strengthen their focus and improve teachers’ attitudes about fulfilling those requirements. At another school, a teacher worked with administrators to link supplemental early childhood guidelines to basic observation rubrics, figuring out the answers to, “So, what's that gonna look like in our room?” Because building a relationship with their observers was so central to the success of evaluation feedback, teachers in this study recognized that they needed to invest time in getting to know principals, assistant principals, and coaches, just as they expected administrators to try and establish a strong professional relationship with them. They also needed to be specific about how to make evaluation a more meaningful process, by describing initiatives that did make an impact, which were generally “long-term” and collaborative, with people that they respected. While some contexts for teaching and learning were more conducive to these efforts than others,

teachers from across the study advised that establishing a productive give-and-take was worth their while.

Limitations to the Findings

While the particular teachers in this study were able to provide a glimpse into the lives of North Carolina teachers who are subject to the official state evaluation process, they were in many ways a unique subset of a larger population. For instance, because they tended to self-identify as average to strong performers, they were not as concerned about evaluation consequences. Participants pointed out that the evaluation experience may look quite different for brand-new or struggling teachers, who would either already be seeking more feedback in order to establish themselves within the field, or needing to pay attention to their principal's suggestions in order to maintain their position. As relatively seasoned and capable teachers, these participants instead were less likely to perceive each piece of feedback as crucial. While this also meant that they had a greater capacity for navigating policy mandates and determining which pieces to apply to their practice, this may have colored their perspective towards performance ratings. They also had average to strong relationships with their current principals and colleagues, even though some reported negative experiences in the past. This meant that they only exemplified the experiences of teachers without much conflict or tension within their relationships with influential fellow educators. Since 11 out of 12 participants taught on the PreK-2 level where standardized testing is not as intense, the broader effects of accountability-related administrative pressures may actually have been minimized within their accounts of evaluation's impact. If, for instance, this study had focused on upper elementary teachers with a longer and more intense history of navigating accountability pressures from standardized state testing, their attitudes

towards evaluation may have produced different results. Along those same lines, while implementation flaws were apparent throughout this study's data, these teachers provided a picture of standardized, baseline compliance with evaluation procedures that may not be fully representative of all schools across North Carolina. Overall, the similarities among this small sample of teachers suggest that further study of the impact of evaluation from across a broader range of teachers, grade levels, and school settings would be useful.

Future Directions for Research

Given that this study largely discusses evaluation's impact on PreK-3 teachers with an average of 12 years of experience, future directions for research might include closer examination of teachers in upper elementary grades where testing pressure is more prevalent, or a focus on novice teachers who are receiving official feedback about their instruction for the first time. In addition, trying to identify how evaluation might better help mid-career teachers further develop their instructional capacity instead of stagnating would continue address a deficiency within the current education system that several participants identified as problematic (Taylor & Tyler, 2011). In general, follow-up data about the initial success of promising but relatively small evaluation initiatives, such as those that effectively incorporate peer review or utilize a portfolio of data within a comprehensive system of providing instructional guidance would be useful to states trying to build their own evaluation capacities (Daley & Kim, 2010; Humphrey, et al, 2011).

Further information about evaluation could also be utilized to foster closer collaboration with key figures like policymakers or principals who want to improve the evaluation process. Because administrators play such a large part in evaluation implementation, future research designs that create dialogue between principals and teachers

might be one direction to consider. For example, if both administrator and teacher groups were interviewed separately, but then reviewed de-identified composite summaries of each other's data together, each group could then begin to address the challenges surrounding effective use of evaluation mechanisms. Regardless of how this exchange unfolded, gathering input from principals and combining it with teacher voices would enhance policymakers' understanding of what it is like to carry out evaluation requirements on the ground level. Along those lines, as more rigorous teacher evaluation becomes more established in schools, opportunities to ask teachers to reflect on their experiences with evaluation should increase, adding to the slim pool of data about this relatively new policy phenomenon.

Given that these data were only representative of the experiences of the dozen teacher participants in this study, who were demographically similar and all based out of North Carolina, this study does not claim to fully encapsulate what evaluation is like for teachers across the nation. However, capturing their stories does indicate that teachers have rich data to offer about how performance feedback may or may not drive their practice, as key street-level bureaucrats (Marshall & Gerstl-Pepin, 2005). Because teachers are directly impacted by evaluation policy, but also play a role in how well it will be received, their opinions of its instructional utility are instrumental. If the major takeaway from their experiences is that extensive evaluation measures were fairly ineffective ways to boost teachers' conceptions of good practice and professional identity development, then the efficacy of evaluation deserves future examination. This is especially true given the contrasting hype surrounding the supposed importance of evaluation on the federal and state level. The mismatches between teachers' tepid assessments of the impact of teacher evaluation on their professional lives and

the lofty stated objectives of official evaluation processes signals a need to reconsider whether or not such measures are worth investing in.

If the primary goals of evaluation are merely to categorize teachers according to their outputs and then use that information to drive short-term personnel action, then consistently applied monitoring mechanisms within the NC Teacher Evaluation Process may suffice. While early research indicates that components of evaluation processes like value-added scores are not reliable enough differentiators between teachers to warrant informing high-stakes staffing and salary decisions (Amrein-Beardsley & Collins, 2012; Baker et al., 2010; Darling-Hammond, Beardsley, Haertel, & Rothstein, 2012; Rothstein, 2010), combining their use with a number of other data points over time could help identify trends in performance. On the other hand, if the educational community wants teachers to truly invest in evaluation as a tool for ongoing professional development, policymakers will need to examine if evaluation is living up to its full promise and potential. When doing so, it is also important to note that while many Americans work in positions where regular performance review from a supervisor is a normal part of their job culture, for public schools in states like North Carolina, extensive standardized evaluation that explicitly holds teachers responsible for student learning is something new. This has made the transition from cursory performance management to more extensive evaluation ratings less smooth than in other sectors where such mechanisms have already been tested and refined. In many cases, schools quickly began to enact performance management without the benefit of administrator experience with delivering extensive feedback, the skill set to navigate the appropriate use of tools like value-added models, or long-standing investments in talent development. This is in contrast to the private sector, where “measurement of performance almost never depends on narrow

quantitative measures analogous to test scores in education” (Baker et al, 2010, p. 6), supervisors tend to build management experience with smaller numbers of employees over time, and performance ratings are rarely as high-stakes or widely broadcast as teacher evaluation results have recently become.

For much of our school system’s history, teachers received very little individualized guidance beyond a cursory determination of satisfactory or unsatisfactory instruction (Weisberg, et al., 2009). While instructional coaches have helped fill that gap, their support is usually focused on providing resources to enhance certain content areas or curriculum, rather than assessing the entirety of a teacher’s pedagogical efforts. Similarly, although mentors are often assigned to novice teachers, most districts are not able to provide that level of support to teachers at all experience levels (Ingersoll & Strong, 2011). Perhaps rightly so, the majority of a school’s resources are directed towards providing services to children and fostering their growth, while the professional growth and job satisfaction of the teachers who drive instruction is given considerably less attention and therefore, less cohesive planning and financial backing. For example, across the nation, districts have traditionally only designated anywhere from one to eight percent of their budget for professional development activities (Miles, Odden, Fermanich, & Archibald, 2005). Therefore, management structures within schools are not as readily equipped with the capacity to deliver thorough, personalized feedback to each employee.

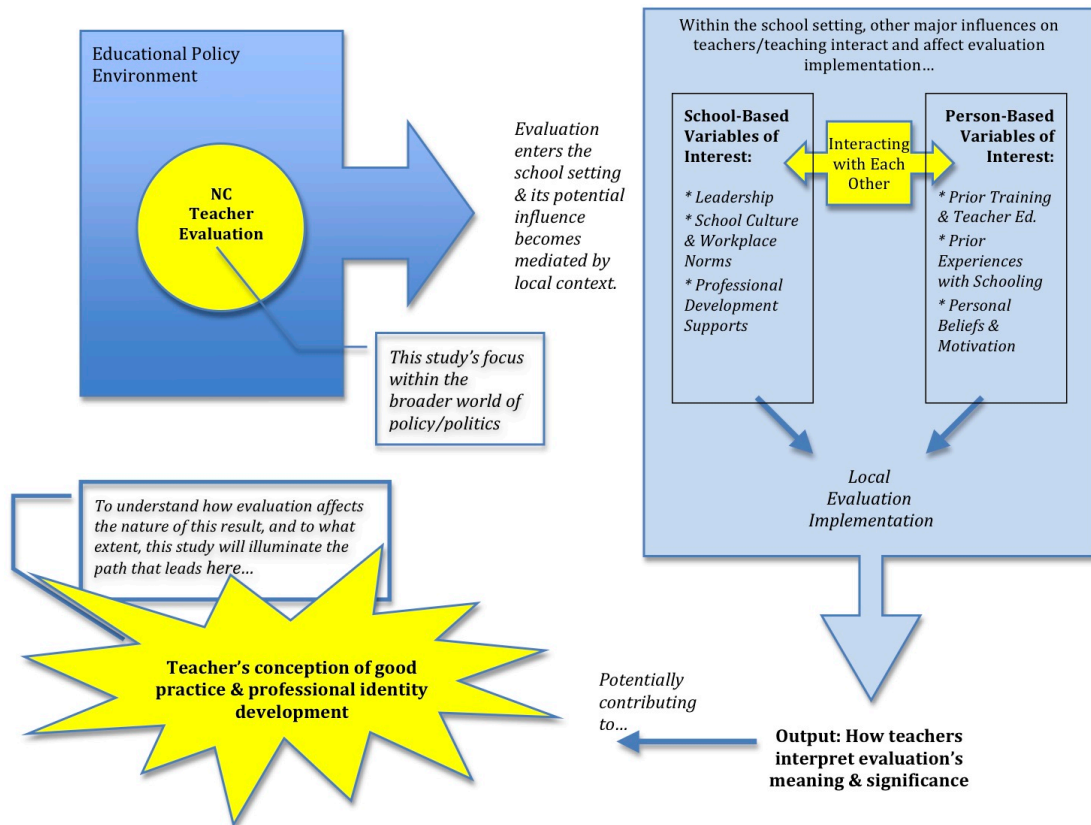
This means that at the very least, the structures that participants in this study identified as vital components of more effective evaluation implementation should be considered as models for other schools and districts that want to become better at delivering performance feedback. As researchers studying recent evaluation measures begin to

document the effects of implementation issues and inadequate feedback mechanisms, suggestions for richer alternatives are already starting to emerge (Hill & Grossman, 2013; Darling-Hammond, 2013; Kane & Staiger, 2012; Marzano, 2012; Papay, 2012). These suggestions align with many of the experiences of the participants in this study, which pointed towards strengthening observation procedures and making evaluation feedback a more specific and meaningful part of sustained coaching relationships. States and districts are particularly in need of additional data from schools that have figured out how to more directly link evaluation to other support measures that foster teachers' ongoing growth. As Hill and Grossman (2013) explain:

Policymakers must resist the urge to think that simply holding teachers accountable through evaluation systems will result in the changes for teaching that are required...Instead, policymakers must in engage in the kind of high-demand, high-support policies that both help teachers learn more about the kinds of instruction envisioned by new standards and to receive the feedback and professional development required to develop new knowledge and skills (p. 382).

This will promote widespread use of evaluation data to boost teachers' effectiveness, rather than stopping short after simply identifying areas where their practice needs improvement. Unless teacher evaluation can combine summative assessments of teachers' abilities with formative guidance, it has much less of a chance of making a significant impact on teachers' conceptions of good practice and professional identity development.

APPENDIX A: FRAMEWORK FOR EXAMINING TEACHER EVALUATION'S INFLUENCE



APPENDIX B: SURVEY QUESTIONS

Thank you for your participation in this study about teacher evaluation practices in North Carolina. The following survey questions will provide us with basic information about how evaluation is being implemented at your school, as well as some background information about your overall teaching experience.

<i>Researcher Will Fill In Your Study ID Code:</i>	
--	--

Basic Background Information:

# of Years Teaching (<u>just at your current school</u>):	
# of Years Teaching Total (career total, <u>at any school, anywhere</u>):	
Grade Currently Teaching:	<input type="checkbox"/> Kindergarten <input type="checkbox"/> 1 st Grade <input type="checkbox"/> 2 nd Grade <input type="checkbox"/> 3 rd Grade <input type="checkbox"/> 4 th Grade <input type="checkbox"/> 5 th Grade <input type="checkbox"/> Other
Gender:	<input type="checkbox"/> Male <input type="checkbox"/> Female
Race/Ethnicity:	
Highest Level of Education You Have Completed:	<input type="checkbox"/> High School <input type="checkbox"/> Associate's Degree <input type="checkbox"/> Bachelor's Degree <input type="checkbox"/> Some Coursework or Certification after Bachelor's <input type="checkbox"/> Master's Degree or Higher

Evaluation at Your Current School:

What tools does your school use to evaluate you? Please select all options that apply:

- | | |
|---|---|
| <input type="checkbox"/> Rubrics | <input type="checkbox"/> Formative benchmarks (that |
| <input type="checkbox"/> Observation guides | emphasizing ongoing development |
| <input type="checkbox"/> Test scores | vs. a final score) |
| <input type="checkbox"/> Peer feedback | <input type="checkbox"/> Other: |
| <input type="checkbox"/> Student feedback | _____ |
| <input type="checkbox"/> Parent/family feedback | _____ |

Does your school conduct formal teacher evaluations according to the North Carolina state guidelines, to the best of your knowledge?

- ☐ Yes – follows the state guidelines closely
- ☐ Yes – follows the state guidelines closely, but also incorporates other forms of evaluation chosen by my school
- ☐ Somewhat – does not follow through on all of state components
- ☐ No – does not use state guidelines
- ☐ Not sure

Do value-added scores based on student achievement progress currently factor into your performance ratings? (For instance, data that measures how much growth your students made over the course of the year.)

- | | |
|--|-----------------------------------|
| <input type="checkbox"/> Yes, currently do | <input type="checkbox"/> No |
| <input type="checkbox"/> Not yet, but will in the future | <input type="checkbox"/> Not sure |

Who conducts teacher evaluation observations at your school? Please select all options that apply:

- | | |
|---|---|
| <input type="checkbox"/> District personnel | <input type="checkbox"/> Other colleagues |
| <input type="checkbox"/> Principal | <input type="checkbox"/> Other: |
| <input type="checkbox"/> Another administrator (such as
assistant principal) | _____ |
| <input type="checkbox"/> Instructional coach | _____ |

How would you rate the overall importance of the current evaluation process to the administrators at your school?

☐ Very Important

☐ Somewhat
Important

☐ Not Very Important

Some pieces of the evaluation process are designed to provide ongoing feedback and development. However, some schools may do a better job than others of prioritizing giving feedback and providing support to help teachers grow. Please indicate how important these parts of the evaluation process are to your school?

	Important	Somewhat Important	Not Important	N/A
Feedback from ongoing formal observations throughout the year	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
End-of-year observation ratings	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Value-added scores	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Self-assessment/reflection	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Professional development planning	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Public reports of teacher performance	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pay-for-performance or bonus eligibility	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other <input type="text"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please rate the overall quality of the feedback that you have received as part of the evaluation process at your school:

☐ Excellent

☐ Average

☐ Poor

☐ Good

☐ Somewhat Poor

Your Personal Evaluation Experiences:

Earlier, you indicated how important various components of the evaluation process are to your school. Now we would like to know more about what you personally think of each of these components. Please indicate how important they are to you.

	Important	Somewhat Important	Not Important	N/A
Feedback from ongoing formal observations throughout the year	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
End-of-year observation ratings	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Value-added scores	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Self-assessment/reflection	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Professional development planning	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Public reports of teacher performance	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pay-for-performance or bonus eligibility	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other <input type="text"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

How important are your evaluation results to you overall?

- ☐ Very Important
 ☐ Somewhat Important
 ☐ Not Very Important

How would you describe your overall experiences with the current evaluation process?

- ☐ Very Positive
 ☐ Somewhat Positive
 ☐ Not Very Positive

Did you ever receive formal training on how the evaluation process would work?

- ☐ Yes – multiple times
 ☐ No – not at all
☐ Yes – once
 ☐ Not sure
☐ Was discussed more informally

The North Carolina Evaluation Process Rubric:

How familiar are you with the NC teacher evaluation observation rubric (i.e. the “standards”)?

- ☐ Very Familiar ☐ Somewhat Familiar ☐ Not Very Familiar

How many times have you been formally observed as part of the official evaluation process so far this year?

- ☐ 5 or more times ☐ 1-2 times
☐ 3-4 times ☐ 0/Not yet

Did you receive feedback after each formal observation?

- ☐ Always ☐ Sometimes ☐ Never
☐ Most of the Time ☐ Rarely ☐ N/A

How would you generally rate the helpfulness of the feedback you received after each formal observation?

- ☐ Very Helpful ☐ Somewhat Helpful ☐ Not Very Helpful

How many times have you been informally observed so far this year, in addition to your formal observations – for example, during walkthroughs?

- ☐ 5 or more times ☐ 1-2 times
☐ 3-4 times ☐ 0/Not yet

Did you receive feedback after each informal observation?

- ☐ Always ☐ Sometimes ☐ Never
☐ Most of the Time ☐ Rarely ☐ N/A

How would you generally rate the helpfulness of the feedback you received after each informal observation?

- ☐ Very Helpful ☐ Somewhat Helpful ☐ Not Very Helpful

Do you ever discuss the evaluation rubric, your observations, or administrator feedback with your colleagues?

- | | | |
|---|------------------------------------|--------------------------------|
| <input type="checkbox"/> Always | <input type="checkbox"/> Sometimes | <input type="checkbox"/> Never |
| <input type="checkbox"/> Most of the Time | <input type="checkbox"/> Rarely | <input type="checkbox"/> N/A |

Do evaluation standards contribute to your sense of what it means to be a good teacher?

- | | |
|---|---|
| <input type="checkbox"/> Significantly Contribute | <input type="checkbox"/> Hardly Contribute at |
| <input type="checkbox"/> Somewhat Contribute | All |

Do evaluation standards contribute to decisions you make about how to approach your teaching practice?

- | | | |
|--|--|--|
| <input type="checkbox"/> Significantly
Contribute | <input type="checkbox"/> Somewhat Contribute | <input type="checkbox"/> Hardly Contribute at
All |
|--|--|--|

Is there anything else that you would like to share about your evaluation experiences at this time?

APPENDIX C: FOCUS GROUP QUESTIONS & ACTIVITY

In-School Focus Group (with fellow Leadership team members - about 60 minutes)

Intro: Have consent forms ready to sign; explain that based on their initial survey answers, we would like to find out more about how evaluation is operating within their school context.

Mostly structured, but partially based on responses to initial survey, if participants are able to complete in advance of focus group session. If they are unable to complete it in advance, time will be allotted to fill out a paper copy of the survey at the start of the session.

Step 1 – Getting more details about evaluation within their context

- *How would you describe the new evaluation process?*
- *Can you provide more details about how this process was introduced to you?*
 - *Follow-up on implementation details: How is evaluation being carried out? Are there certain areas of the process that are being emphasized? How important is evaluation relative to other policies, etc. being implemented at your school?*
 - *Follow-up on school-based factors: What role does leadership play in how evaluation is carried out within your school? How does evaluation fit into other professional development opportunities offered by your school? Are the guidelines within the evaluation process a good fit with your school culture? Are there other ways that your school context shapes your evaluation experiences?*

Possible additional questions to gain detail, time & conversation direction permitting:

- Tell me about what a typical observation is like at your school. How long does it take to get that feedback? Is the advice you get specific to your lessons/customized to you?
- What parts of the rubric do your administrators focus on during your evaluation? (Leadership, diversity, content, etc.) What kinds of things would they like you to focus on?
- How well do you think the evaluation process works? Is it helpful to you? What makes it more/less successful? What role does your principal play in making evaluation *useful*?
- What do other teachers in your school think about evaluation?

Step 2 – Understanding how evaluation compares to other potential factors within the ecosystem of influences on their practice.

Now, I would like to find out more about evaluation's importance relative to other influences on your practice. I'm going to ask you to review this chart, which lists some school, person, and policy-based factors that might contribute to your teaching. School-based factors are the structural and organizational components of your school context, such as workplace norms,

leadership, professional development opportunities, and so on. Person-based factors are what you bring into the classroom with you, such as prior experience or training, or beliefs about teaching. Policy-based factors describe the broader educational context outside of your school, which includes the surrounding political environment (**See attached chart & further directions for filling out in Appendix C...I will walk them through this**).

After filling out the chart and identifying key positive & negative influences on their teaching...

- *What kinds of person-based influences (what you walk into the classroom with) did you rank as relatively important to you? Why are these influences so important? What kinds of influences are less important? (Sharing just a few examples with group)*
- *What kinds of school-based influences (within your work environment) did you rank as relatively important to you? Why are these influences so important? What kinds of influences are less important? (Sharing examples with group – focus here on bringing out commonalities within school context – for example, if they all say leadership – how so?)*
- *What kinds of policy-based influences (thinking more broadly about policy environment that surrounds teaching) did you rank as relatively important to you? Why are these influences so important? What kinds of influences are less important? (Sharing examples with group – focus here on bringing out relative importance of evaluation and how it relates to other policy trends)*

Note – when filling out the chart, make it clear that influences can be either positive or negative – for instance, maybe their PLC drives a lot of what they do in the classroom, even if they don't like it. Or, perhaps their principal is really great and would go under the positive leadership category, even though their district superintendent is draconian and would go in the negative section.

Step 3 – Understanding how major influences on their practice connect to each other.

Many influences on your teaching may be connected. For instance, if you had prior opportunities for collaboration (either positive or negative!) within your teacher training, that may affect how you view current opportunities for collaboration within your school. Take a minute to think about the relationships between the most powerful influences on your approach to teaching.

- *Would anyone like to share an example of how some of the major influences on your teaching practice are connected to each other?*
- *Which influences are closely aligned or in sync with each other? Which empower you/work in combination to drive you forward? (Positive relationship)*
- *Which are in conflict/cause roadblocks or tension? (Negative relationship)*

- How does evaluation fit into this picture (or not)? Which of these other influences do you think most affects how evaluation policy is carried out?

Focus Group Activity

Teachers have many **positive** influences on their practice. Some of those potential influences are listed in the chart below. First, review the listed options and decide if there is anything you would like to add to the list. If so, add it to one of the boxes where there is room for you to write in an “other” option. Then, in the space provided below each listed influence, briefly describe what it means to you. For example, under “Family background”, someone might write, “Come from a long line of teachers”. Lastly, use the scale to indicate the relative importance of each of these positive influences on your own teaching career.

Potential <u>Positive</u> Influences on Your Practice		0 (not important at all)	1 (not very important)	2 (somewhat important)	3 (important)	4 (very important)
Person-based influences (what you walk into the classroom with)	<i>Family background:</i>					
	<i>The years you spent in school as a student:</i>					
	<i>Your race, ethnic background, culture:</i>					
	<i>Personal beliefs & motivation:</i>					
	<i>Prior training & teacher education:</i>					
	<i>Other person-based factor:</i>					
	<i>Other person-based factor:</i>					
	<i>Other person-based factor:</i>					

Potential <u>Positive</u> Influences on Your Practice		0 (not important at all)	1 (not very important)	2 (somewhat important)	3 (important)	4 (very important)
School-based influences (factors within your school context or workplace)	<i>Leadership:</i>					
	<i>Relationships with colleagues/mentors:</i>					
	<i>Who your students are (demographics, preferences, etc.):</i>					
	<i>School culture/norms & overall feel of working environment:</i>					
	<i>Professional development sessions:</i>					
	<i>PLC's/opportunitie s for collaboration:</i>					
	<i>Other school-based factor:</i>					
	<i>Other school-based factor:</i>					
	<i>Other school-based factor:</i>					

Potential <u>Positive</u> Influences on Your Practice		0 (not important at all)	1 (not very important)	2 (somewhat important)	3 (important)	4 (very important)
Policy-based Influences (within the surrounding political environment)	<i>Teacher evaluation expectations:</i>					
	<i>Other performance-based/accountability policies:</i>					
	<i>Other policies in general:</i>					
	<i>General state of education within NC (priorities, funding, etc.):</i>					
	<i>How teachers are portrayed in the media to the public:</i>					
	<i>Other policy-based factor:</i>					
	<i>Other policy-based factor:</i>					
	<i>Other policy-based factor:</i>					

Once you have completed all three sections of this chart, please star the 2-3 positive influences that you consider to be the most prominent, or central to your experience.

Teachers also have many **negative** influences on their practice. Some of those potential influences are listed in the chart below. First, review the listed options and decide if there is anything you would like to add to the list. If so, add it to one of the boxes where there is room for you to write in an “other” option. Then, in the space provided below each listed influence, briefly describe what it means to you. For example, under “Family background”, someone might write, “People in my family look down on teaching – they are all doctors”. Lastly, use the scale to indicate the relative importance of each of these negative influences on your own teaching career.

Potential <u>Negative</u> Influences on Your Practice		0 (not important at all)	1 (not very important)	2 (somewhat important)	3 (important)	4 (very important)
Person-based influences (what you walk into the classroom with)	<i>Family background:</i>					
	<i>The years you spent in school as a student:</i>					
	<i>Your race, ethnic background, culture:</i>					
	<i>Personal beliefs & motivation:</i>					
	<i>Prior training & teacher education:</i>					
	<i>Other person-based factor:</i>					
	<i>Other person-based factor:</i>					
	<i>Other person-based factor:</i>					

Potential <u>Negative</u> Influences on Your Practice		0 (not important at all)	1 (not very important)	2 (somewhat important)	3 (important)	4 (very important)
School-based influences (factors within your school context or workplace)	<i>Leadership:</i>					
	<i>Colleagues/mentors:</i>					
	<i>Who your students are (demographics, preferences, etc.):</i>					
	<i>School culture/norms & overall feel of working environment:</i>					
	<i>Professional development sessions:</i>					
	<i>PLC's/opportunities for collaboration:</i>					
	<i>Other school-based factor:</i>					
	<i>Other school-based factor:</i>					
	<i>Other school-based factor:</i>					

Potential <u>Negative</u> Influences on Your Practice		0 (not important at all)	1 (not very important)	2 (somewhat important)	3 (important)	4 (very important)
Policy-based Influences (within the surrounding political environment)	<i>Teacher evaluation expectations:</i>					
	<i>Other performance-based/accountability policies:</i>					
	<i>Other policies in general:</i>					
	<i>General state of education within NC (priorities, funding, etc.):</i>					
	<i>How teachers are portrayed in the media to the public:</i>					
	<i>Other policy-based factor:</i>					
	<i>Other policy-based factor:</i>					
	<i>Other policy-based factor:</i>					

Once you have completed all three sections of this chart, please star the 2-3 negative influences that you consider to be the most prominent, or central to your experience.

APPENDIX D: INTERVIEW QUESTIONS

Follow-up Interview (about 60 minutes)

Semi-structured, partially based on responses to initial survey and focus group activities:

Intro:

- Explain overall purpose of study and participation requirements:
 - Thank you very much for being a part of this study! As you know, I am documenting teacher experiences with the new evaluation policies in North Carolina that have been put into practice over the past few years. I am interested in hearing more about what your personal experiences have been like within your particular school, what you think of the evaluation rubric and guidelines, and how the expectations outlined within the rubric compare to your own perceptions of your job responsibilities.
 - I plan to interview about a dozen teachers to build detailed case study examples of evaluation experiences. This interview that will allow me to ask some more specific questions about your survey and focus group responses and discuss the current policy a bit further.
- Ask if there are any questions at the start.

Follow-up topic #1: Comparing Evaluation Language to Own Sense of Professional Identity

Overall topic points – reiterating their sense of teaching vs. evaluation (may be in sync, unconnected, or in opposition):

- **How would you describe your job to others?**
 - What does it mean to be a teacher?
 - What would you say your main job responsibilities are?
- **What is really important to you as a teacher?**
 - What are you trying to accomplish within your classroom? What do you believe you should be striving for?
 - How would you describe good practice?
 - Can you give an example of a goal you have set for yourself as a teacher?
 - How do you personally measure your teaching success? How do you know if you have met your goals?
- **What/Who do you think has influenced your ideas about what teachers should do?**
 - How do you decide how to go about your job? How do you decide how to approach your students? Who/what shapes your practice?
 - Looking back at chart from focus group, you indicated that X and X were the two most significant *school-based positive* influences on your practice. Can you tell me more about why these influences have made such an impact? How do they inform your practice?

- Looking back at chart from focus group, you indicated that X and X were the two most significant *school-based negative* influences on your practice. Can you tell me more about why these influences have made such an impact? How do they inform your practice?
 - Ask same for top policy and person-based positive/negative influences on their practice, as indicated during earlier focus group activity
 - Is your personal view of how to approach your job in sync with your school's expectations? In sync with what you think educational policy makers would like you to accomplish?
- **How does teacher evaluation fit into this picture?**
 - How much do you think the evaluation process contributes to who you are as a teacher? (If at all – relatively speaking)
 - May also connect to previous focus group/survey answers to similar questioning depending on prior responses
 - Is there anything about the evaluation process that has been particularly helpful to your professional growth? Anything that has been particularly detrimental?
 - Earlier, you indicated that evaluation had a *low/average/high* importance in your professional life. Can you tell me a little more about its relative importance?
 - Again, also referring back to answers from earlier data from survey/focus group questions

Follow-up topic #2: Experiences Being Observed

Overall topic points:

- **Describe further what evaluation observations have been like for you.** (Follow up on answers from focus-group)
 - What is it like to be observed by your supervisor/administrator?
 - How many times do you get observed per year?
 - Do you know when you will be observed in advance?
 - How long does your supervisor typically observe you?
 - Is there time designated for follow-up conversation? If so, what do you usually talk about in those meetings? Do you get specific instructional guidance?
- **What do you think administrators are looking for?** (Related to own personal evaluation experiences – what the policy is asking administrators to do, how they are carrying it out)
 - What does your administrator tend to focus on during evaluations?
 - What kinds of feedback have you gotten? Have there been particular areas of strength? Areas you are focusing on improving?

- Are there specific sections of the evaluation that get more emphasis than others?
 - *Will also follow up on this when we refer to evaluation process guidebook later in interview in follow-up topic #4*
- **Which implementation factors make being evaluated a more/less successful experience?** (some follow-up from focus group, but goal is to move on from here to think about connection between context, policy, and personal expectations).
 - How well do you think your school implements the evaluation process?
 - What role does your administrator play in evaluation implementation?
 - Do your colleagues influence your perception of evaluation goals? Do you have opportunities to connect with colleagues to discuss what you are working on improving?
 - Is the general environment of your school receptive to evaluation? What is the overall feel or sentiment? How much do you think the particular work culture within your school affected your view of evaluation policies?
 - Within your particular school environment, is evaluation linked to professional development?
 - If so, how?
 - What kind of professional development have you received based on your evaluation ratings?

Follow-up topic #3: Attitudes Towards Evaluation

Overall topic points:

- **Continued focus on helpfulness/usefulness of process & feedback: How do you feel about the evaluation process?**
 - Do you feel pressured by the evaluations? Supported? Constrained? Emboldened? Are you nervous about being observed?
- **Effects on practice: How helpful are evaluation ratings in improving your practice?**
 - Are there other supports that you find more/less helpful, relatively speaking?
 - How well are you able to connect evaluation expectations to daily decisions about teaching?
 - Can you give an example of a time when you made a decision about teaching that was linked to evaluation expectations?
- **Formative vs. summative emphasis: Do you feel that the evaluation process is being used as a tool for professional growth?**
 - How would you describe evaluation's purpose?
 - Do you feel as if evaluation is an ongoing process, contributing to your continual professional development?
 - What effect do evaluation ratings have on your view towards teaching?
 - What part of the evaluation process is the most important to you?
 - How much does the rubric matter as compared to student test scores? How do you feel about value-added measures becoming part of the evaluation process?

- How well do you feel the ratings you receive reflect your teaching abilities and your practice?
- **Importance compared to other ways of being judged – i.e. student test scores: Do you feel that the evaluation process is the key way that your performance is being evaluated within your school setting?**
 - Is there any type of pressure associated with evaluation ratings? How does this type of pressure compare to other pressures you may feel on the job?
 - How does pressure to perform on teacher evaluations compare to pressure to produce higher student achievement? Is there a relationship between the two?

Follow-up topic #4: Discussion of Rubric Structure

Overall topic points:

- **Talking about specific sections of rubric – which are being emphasized?** (*Looking at overview of purpose of evaluation, major teaching responsibilities, rubric headings*)
 - How would you describe your general impression of what the state wants to see in teachers, and what a successful teacher should do – as compared to your own teaching objectives?
 - What messages does this policy send about what your teaching responsibilities are and how success in those areas will be measured?

Detailed questions about rubric structure:

- **How does this policy define good teaching?**
 - What do you think the rubric is telling you about what it takes to become a successful teacher?
- **How would you define good teaching?**
 - What do you see yourself as responsible for?
 - How do you know you have fulfilled your job responsibilities? (*Following up on earlier answers to clarify and confirm*)
- **Do you appreciate these guidelines, and/or find them to be beneficial?**
 - Are they a factor in making instructional decisions? What purpose do you think they serve? How do they compare to your own goals/approaches to the classroom?
- **How much discretion do you feel you have in determining *how* to meet these objectives?**

APPENDIX E: DATA COLLECTION TIMELINE

Data Collection Element	Description	Purpose/Connection to Analysis	Further Rationale
<i>Recruitment of participant teachers</i> (September-October 2013; began after IRB approval and buy-in from principals)	Drew from teacher “leadership team” from prior research project, spanning three NC partner schools; initial request was to participate in school-specific focus groups	Allowed for comparison between teacher experiences with evaluation policy at three distinct North Carolina elementary schools	Data from these groups provided insight into a “best-case scenario” of how teachers who are primed to discuss their practice might interpret & perceive evaluation feedback; teachers all had some commonalities, but varied in terms of teaching experience, background, grade level across PreK-3, implementation experiences, and person-based motivators
<i>Initial survey</i> (Directly before focus group)	Brief survey was distributed to all participating teachers prior to initial focus group (original plan was to distribute electronically, but logistically it ended up working better to provide in person)	Tool for collecting demographic info like teaching experience; created a basic understanding of how evaluation is implemented within their schools, and its relative importance; provided foundation for future conversation	Provided common basis for assessing how evaluation has been executed & how attitudes towards the evaluation process compare across schools
<i>School-specific focus groups</i> (October-December 2013)	School-specific, focused on understanding context for implementation	Began by asking teachers to describe their evaluation experiences. Then, teachers identified the range of factors that contribute to their professional personal/professional identity development and subsequent perceptions of practice. Lastly, a structured activity was designed to compare the relative	Allowed teachers to talk about their personal experiences within their particular school, but also provided a common frame for identifying important influences on practice relative to evaluation; shows how school context can mediate these relationships; identifies

		influence of these factors, particularly in comparison to the evaluation objectives being presented within their school, as well as whether these influences have a positive/negative impact (see Appendix C, Focus Group Activities)	implementation variables particular to their school context that appear across their “group narrative”
<i>Initial Analysis of Survey, Focus Group, and Focus Group Activity Data</i> (November 2013-January 2014)	Looked across survey, focus group, and focus group activity data.	Goal was to identify both broader themes and specific examples from each teacher’s prior data to follow up with teachers about during upcoming interview interviews.	Guided follow-up questions about individual interpretation and internalization of evaluation standards and related performance management processes.
<i>Individual interviews</i> (December 2013-February 2014)	With most teachers – spoke with whoever was available and willing to participate, in an effort to do as much member checking and follow-up as possible.	During interview, more specifically discussed NC process together. Confirmed and elaborated on themes from across evaluation experiences to be used to build larger NC case - using Yin’s (2009) view of being able to make theoretical generalizations by utilizing case types to demonstrate what could happen given circumstances (p. 38). Also asked about own conceptions of good practice/teaching.	See interview protocols in Appendix D – questions guided them describing evaluation process’s role in their lives and providing more details about themes within prior data analysis. Member checking – triangulation of multiple data sources. Thinking about how individual experiences connect within and across school contexts. What do these experiences imply about evaluation effects? How could they inform future implementation efforts?
<i>Analysis of common themes incorporating interview insights</i>	Connected all prior data points with individual interviews.	Brought data/common themes together across schools to gain a picture	Looked to see if their priorities for good practice were

(January-February 2014)		of how evaluation functions across NC. Compared evaluation's impact to explicit examples of other influences on teachers' practice and professional identity development.	aligned with evaluation language and objectives. Also tried to determine which influences on professional identity development emerged as central to this process.
-------------------------	--	---	--

APPENDIX F: CODING DETAILS FOR PART TWO OF FINDINGS

MAIN CODING	SUBCODING	DESCRIPTION	REASON
ATT (General attitudes towards evaluation)	<i>Further broke down into + and - feelings</i>	General feelings towards evaluation's utility and value, overall nature of evaluation experience	Stemmed from main research question about evaluation's relative influence
	<i>Doesn't Matter</i>	Data about overall impressions indicating that evaluation didn't "really matter"	Attitudinal pattern that emerged after initial analysis of survey, focus group & interview data
	<i>Connection</i>	Mentions of a lack of personal connection/meaning	"
	<i>Pressure</i>	Mentions of negative feelings like pressure, specifically in association with evaluation & being rated/observed	Patterns of negative language emerged from across focus group & interview data (i.e. "anxious"; "stressed")
	<i>Administrative</i>	Mentions of evaluation as an administrative task to be completed	Patterns emerged from across focus group & interview data characterizing views towards evaluation
	<i>Performance</i>	Mentions of performance-oriented motivation; desire to meet expectations, wanting to be praised	Patterns of motivation emerged from across focus groups/interviews
	<i>Evaluative</i>	Depictions of the evaluation process that were formative vs. summative	Stemmed from initial research impetus to understand how evaluation's purpose was being perceived
	<i>Experience</i>	Mentions of how experience level affected view towards being evaluated	Multiple teachers brought up experience variable; was a follow-up topic with either very experienced or newer teachers during interviews

IMP (Implementation)		Any mention of how evaluation policy was being carried out on school level; logistical procedures	Stemmed from research sub-questions about effects of school context; also, emerging theme across initial scan of focus group & interview data
	<i>General Leadership</i>	Mainly focused on principals but also includes other administrators like assistant principals	“
	<i>Turnover</i>	Mentions of leadership that particularly related to turnover and its effect on evaluation implementation	
	<i>Alignment</i>	How well evaluation was aligned with other professional development supports	“
	<i>Time/Logistics</i>	Any mentions of logistical barriers to implementation like limited time	Emerged as a pattern across all schools in study
	<i>Uncertainty</i>	Any mentions of areas of evaluation implementation that were unclear; also in terms of how evaluation ratings would be used, now or in future	Linked to discussion of broader educational climate in NC; Saw patterns of language (i.e. “I guess, or “I don’t know” or “maybe”)
	<i>Context/Culture</i>	Mentions of when particular school setting affected implementation; also related to overall school culture & community	In follow-up interviews, participants were asked to further describe how evaluation was being carried out in their school, based on earlier mentions during focus group
	<i>Positive</i>	Example of positive implementation strategies	These stemmed from examples teachers gave of what worked

INF (Influence)		Mentions of positive influences on practice/teaching	Emerged from central examination of professional identity development process; also linked to focus group activity where participants were asked to rate relative importance of positive influences on practice & give specific examples
	<i>Past role models</i>	From family, prior schooling, teacher training, past mentors	“
	<i>Personal beliefs/motivation</i>	Personal beliefs, values & convictions (not always exclusively related to teaching, but applied within classroom context)	“
	<i>Relationships with colleagues/mentors</i>	Influential co-workers who offered support within school setting	“
	<i>Evaluation</i>	Comparisons of evaluation to other influences; how evaluation was characterized in terms of impact on practice	Related to general impressions of evaluation's value
	<i>Negative</i>	Influences that were more like pressures; characterized as negative in contrast to positive supports	Teachers gave specific examples of these pressures (generally policy-based) during individual interviews & focus group follow-up
GOOD (Definitions of good practice)		Definitions of good teaching: what good teaching looks like, ideal teaching behaviors, what participants strive for with their students, what they think they should be	Specifically asked teachers to give more examples of this during follow-up interviews, to compare to policy discourse that

		accountable for	includes performance standards & teacher expectations, ideal behaviors, etc.
	<i>Long-term</i>	Definitions of good practice linked to long-range goals and views of student success	Emerged from trends in teacher responses during individual interviews about this topic.
	<i>Socio-emotional</i>	Definitions of good practice linked to building students' socio-emotional skills.	“
	<i>Relationships</i>	Definitions of good practice linked to building relationships with both individual students and class as a whole	“

REFERENCES

- Achinstein, B., Ogawa, R. T., & Speiglmán, A. (2004). Are we creating separate and unequal tracks of teachers? The effects of state policy, local conditions, and teacher characteristics on new teacher socialization. *American Educational Research Journal*, 41(3), 557-603.
- Adams, P. & Holland, P. (2002). Through the horns of a dilemma between instructional supervision and the summative evaluation of teaching [Electronic version]. *International Journal of Leadership in Education*, 5, 227-247.
- Aguirre, J., & Speer, N. M. (2000). Examining the relationship between beliefs and goals in teacher practice. *Journal of Mathematical Behaviour*, 18(3), 327-356.
- Alternative Salary Plans/Charlotte-Mecklenburg Schools, H546, House DRH801-LE-93 (03/15), General Assembly of North Carolina, 2011.
- Amrein, A.L. & Berliner, D.C. (2002, March 28). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives*, 10(18).
- Amrein-Beardsley, A., & Collins, C. (2012). The SAS education value-added assessment system (SAS® EVAAS®) in the Houston Independent School District (HISD): Intended and unintended consequences. *Education Policy Analysis Archives*, 20(12).
- Apple, M. (1996). Power, meaning, and identity: Critical sociology of education in the United States. *British Journal of Sociology of Education*, 17(2), 125-144.
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36(5), 258-267.
- Baker, E. L., Barton, P.E., Darling-Hammond, L., Haertel, E., Ladd, H.F., Linn, R.L., Ravitch, D., Rothstein, R., Shavelson, R.J., & Sheppard, L.A. (2010). *Problems with the use of student test scores to evaluate teachers*. EPI Briefing Paper (Press release). Retrieved from http://epi.3cdn.net/b9667271ee6c154195_t9m6ijj8k.pdf
- Ball, D.L. (2011, February 20). Let's radically improve teacher education (and stop fighting about it). *The Chronicle of Higher Education*. Retrieved from <http://chronicle.com/article/Lets-Radically-Improve/126428/>.
- Ball, D. L., & McDiarmid, G. W. (1990). The subject matter preparation of teachers. In R. Houston (Ed.), *Handbook of research on teacher education* (pp. 437-449). New York: Macmillan.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191-215.

- Barrett, B.D. (2009). No Child Left Behind and the assault on teachers' professional practices and identities. *Teaching and Teacher Education*, Vol. 25, 1018-1025.
- Bartell, C. A. (2001). Bridging the disconnect between policy and practice in teacher education. *Teacher Education Quarterly*, 28(1), 189-198.
- Bathey, D. & Franke, M. L. (2008). Transforming identities: Understanding teachers across professional development and classroom practice. *Teacher Education Quarterly*, 35(3), 127.
- Beijaard, D., Verloop, N., & Vermunt, J. D. (2000). Teachers' perceptions of professional identity: An exploratory study from a personal knowledge perspective. *Teaching and Teacher Education*, 16(7), 749-764.
- Beijaard, D., Meijer, P., & Verloop, N. (2004). Reconsidering research on teachers' professional identity. *Teaching and Teacher Education*, 20(2), 107-128.
- Berger, P. & Luckmann, T. (1966). *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. Garden City, NY: Anchor Books.
- Berryhill, J., Linney, J. A., & Fromewick, J. (2009). The effects of education accountability on teachers: Are policies too-stress provoking for their own good? *International Journal of Education Policy and Leadership*, 4(5), 1-14.
- Blumer, H. (1954). What is wrong with social theory? *American Sociological Review*, 19(1), 3-10. Retrieved from <http://cooley.libarts.wsu.edu/schwartz/pdf/Blumer.pdf>
- Borthwick, E., Coades, S., Sennette, J., & Touhey, A. (2010). *History of Teacher Evaluation Models in American Public Education*. Retrieved from <http://a100educationalpolicy.pbworks.com/f/TEM+History+Section.pdf>
- Bourdieu, P. (1990). *The logic of practice*. Stanford, CA.: Stanford University Press.
- Bourdieu, P. (1991). *Language and symbolic power*. Cambridge, MA: Harvard University Press.
- Brandt, R. M. (1995). Teacher Evaluation for Career Ladder and Incentive Pay Programs. In D.L. Duke (Ed.), *Teacher evaluation policy: From accountability to professional development* (pp. 13-33). Albany: State University of New York Press.
- Charlotte-Mecklenburg Schools (2009, November 5). *Strategic plan 2014: Teaching our way to the top*. Retrieved from www.cms.k12.nc.us
- Charlotte Observer. (2011, May 26). Time, care needed on teacher pay changes (Editorial). *The Charlotte Observer*. Retrieved from www.charlotteobserver.com.

- Clotfelter, C.T., Glennie, E., Ladd, H.F. & Vigdor, J.L. (2008). Would Higher Salaries Keep Teachers in High-Poverty Schools? Evidence from a Policy Intervention in North Carolina. *Journal of Public Economics* 92.5-6: 1352-1370.
- Clotfelter, C., Ladd, H., & Vigdor, J. (2010). *Teacher mobility, school segregation, and pay-based policies to level the playing field* (CALDER Working Paper 44). Washington, DC: The Urban Institute. Retrieved from www.urban.org.
- Coburn, C.E. & Stein, M.K. (2006). Communities of practice theory and the role of teacher professional community in policy implementation. In M. I. Honig (Ed.), *Defining the field of education policy implementation* (pp. 25-46). Albany, NY: SUNY Press.
- Cochran-Smith, M., & Lytle, S. L. (1990). Teacher research and research on teaching: The issues that divide. *Educational Researcher*, 19 (2), 2-11.
- Cochran-Smith, M., & Lytle, S. (1999). Relationships of knowledge and practice: Teacher learning in communities. In P. Pearson & A. Iran-Nejad (Eds.), *Review of Research in Education*, 24, (pp. 249-307). Washington, D.C. American Educational Research Association.
- Coffey, A. & Atkinson, P. (1996) *Making Sense of Qualitative Data Analysis: Complementary Strategies*. Thousand Oaks CA: Sage.
- Cohen, D. K., & Ball, D. L. (1990). Policy and practice: An overview. *Educational Evaluation and Policy Analysis*, 12(3), 347-353.
- Cohen, J., McCabe, E., Michelli, N., & Pickeral, T. (2009). School climate: Research, policy, practice, and teacher education. *Teachers College Record*, 111(1), 180 – 213.
- Cohen, D.K., Moffitt, S.L., & Goldin, S. (2007). Policy and practice: The dilemma. *American Journal of Education*, 113(4): 515-548.
- Cohen, D.K. & Spillane, J.P. (1992). Policy and practice: The relations between governance and instruction. *Review of Research in Education*, 18, 3-49.
- Collins, P.H. (2000). *Black feminist thought: Knowledge, consciousness, and the politics of empowerment*. New York: Routledge.
- Conley, S. & Glasman, N. (2008). Fear, the school organization, and teacher evaluation. *Educational Policy*, 22(1) 63-85.
- Crossley, M., & Vulliamy, G. (1995, September). *Issues and trends in qualitative research: Potential for developing countries*. Paper presented at the Oxford International Conference on Education and Development: Globalisation and Learning, New College, Oxford, UK.

- Danielson, C. (2001). New trends in teacher evaluation. *Educational Leadership*, 58(5), 12-15.
- Danielson, C. (2007). *Enhancing professional practice: A framework for teaching* (2nd edition). Alexandria: Association for Supervision & Curriculum Development.
- Danielson, C. (2010). Evaluations that help teachers learn. *Educational Leadership*, 68(4), 35-39.
- Danielson, C., & McGreal, T. L. (2000). Teacher evaluation to enhance professional learning. Princeton, NJ: Educational Testing Service.
- Daley, G. & Kim, L. (2010). *A teacher evaluation system that works* (National Institute for Excellence in Teaching Working Paper). Retrieved from http://www.tapsystem.org/publications/wp_eval.pdf.
- Darling-Hammond, L. (1990). Teacher professionalism: Why and how? In A. Lieberman (Ed.), *Schools as collaborative cultures: Creating the future now*. New York, NY: The Falmer Press.
- Darling-Hammond, L. (2000). How teacher education matters. *Journal of Teacher Education*, 51(3), 166-173.
- Darling-Hammond, L. (2012). *Creating a comprehensive system for evaluating and supporting effective teaching*. Stanford, CA: Stanford Center for Opportunity Policy in Education.
- Darling-Hammond, L. (2013). *Getting teacher evaluation right*. New York: Teachers College Press.
- Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan*, 93(6), 8-15.
- Darling-Hammond, L., Berry, B., & Thoreson, A. (2001). Does teacher certification matter? Evaluating the evidence. *Educational Evaluation and Policy Analysis*, 23(1), 57-77.
- 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.
- Day, C., Elliot, B., & Kington, A. (2005). Reform, standards and teacher identity: Challenges of sustaining commitment. *Teaching and Teacher Education*, 21(5), 563-577.
- Day, C., Flores, M. A., & Viana, I. (2007). Effects of national policies on teachers' sense of professionalism: Findings from an empirical study in Portugal and in England. *European Journal of Teacher Education*, 30(3), 249-265.

- Dee, T.S. & Keys, B.J. (2004) Does Merit Pay Reward Good Teachers? Evidence from a Randomized Experiment. *Journal of Policy Analysis and Management*, Vol. 23, No. 3, 471-488.
- Delpit, L. (2006). *Other People's Children: Cultural Conflict in the Classroom (2nd Edition)*. New York, NY: The New Press.
- Denzin, N. K., & Lincoln, Y. S. (1998). The discipline and practice of qualitative research. *Handbook of Qualitative Research*, 2, 1-28.
- Derrington, M. (2011). Changes in Teacher Evaluation: Implications for the Principal's Work. *Delta Kappa Gamma Bulletin*, 77(3), 51-54. Retrieved from EBSCOhost.
- Desimone, L., Smith, T. M., & Phillips, K. J. R. (2007). Does policy influence mathematics and science teachers' participation in professional development? *Teachers College Record*, 109(5), 1086-1122.
- Dufour, R. (2004). What is a professional learning community? *Educational Leadership*, 61(8), 6-11.
- Duke, D. L. (1995). The Move to Reform Teacher Evaluation. In D.L. Duke (Ed.), *Teacher evaluation policy: From accountability to professional development* (pp. 1-11). Albany: State University of New York Press.
- Duncan, A. (2010a, March 29). *U.S. Secretary of Education Arne Duncan's statement on Race to the Top Phase I winners*. Retrieved from <http://www.ed.gov/news/speeches/us-secretary-education-arne-duncans-statement-race-top-phase-1-winners>
- Duncan, A. (2010b, April 9). *The well-rounded curriculum: Secretary Arne Duncan's remarks at the Arts Education Partnership National Forum*. Retrieved from <http://www.ed.gov/news/speeches/well-rounded-curriculum>
- Duncan, A. (2010c, April 30). *Success as the norm: Secretary Arne Duncan's remarks at the KIPP Annual Dinner*. Retrieved from <https://www2.ed.gov/news/speeches/2010/04/04302010.html>
- Duncan, A. (2010d, May 19). *The lasting impact of teachers: Commencement remarks of Secretary Arne Duncan at Lesley University*. Retrieved from <https://www.ed.gov/news/speeches/lasting-impact-teachers-commencement-remarks-secretary-arne-duncan-lesley-university>
- Duncan, A. (2010e, June 9). *Education and destiny: Secretary Arne Duncan's remarks at the 96th Annual Hampton University Ministers' Conference in Hampton, Virginia*. Retrieved from <http://www.ed.gov/news/speeches/education-and-destiny-secretary->

[arne-duncans-remarks-96th-annual-hampton-university-mi](#)

- Duncan, A. (2010f, June 29). *Education research: Charting the course for reform — Remarks by Secretary Arne Duncan at the Institutes for Education Research Conference*. Retrieved from <http://www.ed.gov/news/speeches/education-research-charting-course-reform-remarks-secretary-arne-duncan-institutes-edu>
- Duncan, A. (2010g, July 14). *Equity and education reform: Secretary Arne Duncan's remarks at the Annual Meeting of the National Association for the Advancement of Colored People (NAACP)*. Retrieved from <http://www.ed.gov/news/speeches/equity-and-education-reform-secretary-arne-duncans-remarks-annual-meeting-national-ass>
- Duncan, A. (2010h, July 16). *The win-win Solution: Secretary Arne Duncan's remarks at the Council of Chief State School Officers and State Higher Education Executive Officers' 2010 Joint Summer Meeting*. Retrieved from <https://www.ed.gov/news/speeches/win-win-solution-secretary-arne-duncans-remarks-council-chief-state-school-officers-an>
- Duncan, A. (2010i, July 27). *Secretary Arne Duncan's remarks at the National Urban League Centennial Conference*. Retrieved from <http://www.ed.gov/news/speeches/secretary-arne-duncans-remarks-national-urban-league-centennial-conference>
- Duncan, A. (2010j, July 27). *Secretary Arne Duncan's remarks at the National Press Club*. Retrieved from <http://www.ed.gov/news/speeches/quiet-revolution-secretary-arne-duncans-remarks-national-press-club>
- Duncan, A. (2010k, July 28). *Unleashing the power of data for school reform: Secretary Arne Duncan's Remarks at the STATS DC 2010 Data Conference*. Retrieved from <https://www.ed.gov/news/speeches/unleashing-power-data-school-reform-secretary-arne-duncans-remarks-stats-dc-2010-data->
- Duncan, A. (2010l, August 25). *Secretary Arne Duncan's remarks at the Statehouse Convention Center in Little Rock, Arkansas*. Retrieved from <http://www.ed.gov/news/speeches/secretary-arne-duncans-remarks-statehouse-convention-center-little-rock-arkansas>
- Eckert, J. M., & Dabrowski, J. (2010). Should value-added measures be used for performance pay? *Phi Delta Kappan*, 91(8), 88-92. Retrieved from EBSCOhost.
- Ellett, C. D. & Teddlie, C. (2003). Teacher evaluation, teacher effectiveness, and school effectiveness: Perspectives from the USA. *Journal of Personnel Evaluation in Education*, 17(1), 101-128.
- Elmore, R. (2002). *Bridging the gap between standards and achievement: The imperative for professional development in education*. Washington, DC: The Albert Shanker

Institute.

- Elmore, R.F. & McLaughlin, M.W. (1988). *Steady work: policy, practice, and the reform of American education*. (Report No. 3574-NIE/RC). Washington, DC: National Institute of Education.
- Ermeling, B. (2010). Tracing the effects of teacher inquiry on classroom practice. *Teaching and Teacher Education*, 26(3), 377-388.
- Fairclough, N. (1995). *Critical Discourse Analysis*. London: Longman.
- Fairclough, N. (2005). Discourse analysis in organization studies: The case for critical realism. *Organization Studies*, 26(6), 915-939.
- Fairclough, N. (2013). Critical discourse analysis and critical policy studies. *Critical Policy Studies*, 7(2), 177-197.
- Feiman-Nemser, S. (1983). Learning to teach. In L. Shulman & G. Sykes (Eds.), *Handbook of teaching and policy* (pp. 150-170). New York: Longman.
- Feiman-Nemser, S. (2001). From preparation to practice: Designing a continuum to strengthen and sustain teaching. *Teachers College Record*, 103 (6), 1013-1055.
- Feiman-Nemser, S. (2003, May). What new teachers need to learn, *Educational Leadership*, 25-29.
- Figlio, D. N. & Kenny, L. W. (2007). Individual teacher incentives and student performance. *Journal of Public Economics*, 91(5-6), 901-914.
- Finnigan, K. S., & Gross, B. (2007). Do Accountability Policy Sanctions Influence Teacher Motivation?: Lessons from Chicago's Low-performing Schools. *American Educational Research Journal*, 44(3): 594-629.
- Flores, M. A., & Day, C. (2006). Contexts which shape and reshape new teachers' identities: A multi-perspective study. *Teaching and Teacher Education*, 22(2), 219-232.
- Franzak, J. K. (2002). Developing a teacher identity: The impact of critical friends practice on the student teacher. *English Education*, 34(4), 256-61.
- Fullan, M. (2001). *The new meaning of educational change (3rd ed.)*. New York: Teachers College Press.
- Gage, N. L. (1972). *Teacher effectiveness and teacher education: The search for a scientific basis*. Palo Alto, CA: Pacific Books.
- Gee, J. P. (2000). Identity as an analytic lens for research in education. *Review of Research in*

- Education*, 25, 99-125.
- Gee, J. (2004). Discourse analysis: What makes it critical? In R. Rogers, Ed., *New directions in critical discourse analysis: The role of language and learning in social transformation* (pp. 19-50). Mahwah, NJ: Erlbaum.
- Gee, J. (2005). *An introduction to discourse analysis: theory and method* (2nd ed.). New York, NY: Routledge.
- Gibson, S., & Dembo, M. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology*, 76, 569-582.
- Gitlin, A., & Smyth, S. (1989). *Teacher evaluation: Educative alternatives*. London: Falmer Press.
- Glazerman, S., Loeb, S., Goldhaber, D., Staiger, D., Raudenbush, S., & Whitehurst, G.J. (2011). *Passing muster: Evaluating teacher evaluation systems*. Washington, DC: The Brookings Institution.
- Glesne, C. (2006). *Becoming qualitative researchers: An introduction*. Boston: Pearson, Allyn & Bacon.
- Gorski, P. C. (2010). Unlearning deficit ideology and the scornful gaze: Thoughts on authenticating the class discourse in education. Retrieved from <http://www.edchange.org/publications/deficit-ideology-scornful-gaze.pdf>
- Goldhaber, D. (2009). *Teacher pay reforms*. Washington, DC: Center for American Progress. Retrieved from http://www.americanprogress.org/wp-content/uploads/issues/2006/12/pdf/teacher_pay_report.pdf
- Goodson, I. (1992). *Studying teachers' lives*. London: Routledge.
- Gordon, R., Kane, T. J., & Staiger, D. O. (2006, April). *Identifying effective teachers using performance on the job*. (The Hamilton Project, Discussion Paper 2006-01). Washington, DC: The Brookings Institution. Retrieved from http://www3.brookings.edu/views/papers/200604hamilton_1.pdf
- Grier, T.B. & Holcombe, A.A. (2008). Mission Possible: Recruiting and retaining teachers. *Educational Leadership*, 65(7), 25-30. Retrieved from EBSCOhost.
- Haniford, L.C. (2009). Tracing one teacher candidate's discursive identity work. *Teaching and Teacher Education*, 26, 987-996.
- Hanushek, E. (1992). The trade-off between child quantity and quality. *Journal of Political Economy*, Vol. 100, 84-117.

- Hardy, I., & Lingard, B. (2008). Teacher professional development as an effect of policy and practice: A Bourdieuan analysis. *Journal of Education Policy*, 23(1), 63-80.
- Hargreaves, A. (1992). Time and teachers' work: An analysis of the intensification thesis. *Teachers College Record*, 94 (1), 87-108.
- Hargreaves, A., & Dawe, R. (1990). Paths of professional development: Contrived collegiality, collaborative culture, and the case of peer coaching. *Teaching and Teacher Education*, 6(3), 227-241.
- Hargreaves, A. & Fullan, M. G (1992). *Understanding Teacher Development*. London: Teachers College Press.
- Harris, D. & Sass, T. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics*, 95(7-8), 798-812.
- Harris, D., Sass, T., & Semykina, A. (2010). Value-Added Models and the Measurement of Teacher Productivity [CALDER Working Paper No. 54]. Retrieved from <http://www.urban.org/UploadedPDF/1001508-Measurement-of-Teacher-Productivity.pdf>
- Haycock, K. (1998). Good teaching matters: How well-qualified teachers can close the gap. *Thinking K-16*, 3(2). Washington, DC: The Education Trust.
- Hazi, H. M. & Rucinski, D. A. (2009). Teacher Evaluation as Policy Target: Viable Reform Venue or Just Another Tap Dance? *ERS Spectrum*, 27(3), 31-40. Retrieved from http://www.edreadysearch.org/content/657/preview-spec_su_09_teacheval.pdf
- Hill, H.C. & Grossman, P. (2013). Learning from teacher observations: Challenges and opportunities posed by new teacher evaluation systems. *Harvard Educational Review*, 83(2), 371-384.
- Hogg, M. A. (2001). A social identity theory of leadership. *Personality and Social Psychology Review*, 5(3), 184-200.
- Hogg, M. A., & Terry, D. J. 2000. Social identity and self-categorization processes in organizational contexts. *Academy of Management Review*, 25: 121-140.
- Holland, D.C., Lachicotte, W.S., Skinner, D., & Cain, C. (1998). *Identity and agency in cultural worlds*. Cambridge, MA: Harvard University Press.
- Honig, M. (Ed.) (2006). *New directions in education: Policy implementation*. Albany, NY: State University of New York Press.
- Hope, W., & Pigford, A. (2002). The principal's role in educational policy implementation. *Contemporary Education*, 72(1), 44-47.

- Humphrey, D. C., Koppich, J. E., Bland, J. A., & Bosetti, K. R. (2011). *Peer review: Getting serious about teachers' support and evaluation*. Menlo Park, CA: SRI International.
- Ingersoll, R. M., & Kralik, J. M. (2004). *The impact of mentoring on teacher retention: What the research says*. Denver, CO: Education Commission of the States.
- Ingersoll, R. & Strong, M. (2011). The impact of induction and mentoring programs for beginning teachers: A critical review of the research. *Review of Education Research*, 81(2), 201-233.
- Jacob, B.A. (2005). Accountability, incentives and behavior: the impact of high-stakes testing in the Chicago Public Schools. *Journal of Public Economics*, 89, 761–796
- Johnson, J. (2010, Sept. 29). *Race to the Top has unique role to play in reforming schools for the future*. [Web log]. Retrieved from <http://www.ed.gov/blog/2010/09/race-to-the-top-has-unique-role-to-play-in-reforming-schools-for-the-future/>
- 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. Research Paper. MET Project. Seattle, WA: Bill & Melinda Gates Foundation. Retrieved from <http://files.eric.ed.gov/fulltext/ED540959.pdf>
- Kane, T. J., & Staiger, D. O. (2008). *Estimating teacher impacts on student achievement: An experimental evaluation* (NBER Working Paper 14607). Retrieved from <http://www.nber.org/papers/w14607.pdf>.
- Kane, T. J., & Staiger, D. O. (2012). Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains. MET Project. Seattle, WA: Bill & Melinda Gates Foundation. Retrieved from http://www.metproject.org/downloads/MET_Gathering_Feedback_Research_Paper.pdf
- Keith, M. (2008). Public sociology? Between heroic immersion and critical distance: Personal reflections on academic engagement with political life. *Critical Social Policy*, 28(3): 320-334.
- Koedel, C., & Betts, J. R. (2011). Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein critique. *Education*, 6(1), 18-42.
- Kimball, S., & Milanowski, A. (2009). Examining teacher evaluation validity and leadership decision making within a standards-based evaluation system. *Educational Administration Quarterly*, 45(1), 34-70.
- Klassen, R. M., & Chiu, M. M. (2010). Effects on teachers' self-efficacy and job satisfaction:

- Teacher gender, years of experience, and job stress. *Journal of Educational Psychology*, 102, 741–756.
- Koretz, D. (2002). Limitations in the use of achievement tests as measures of educators' productivity. *The Journal of Human Resources*, 37(4), 752-777.
- Korthagen, F. (2004) In search of the essence of a good teacher: Towards a more holistic approach in teacher education, *Teaching and Teacher Education*, 20, 77- 97.
- Labaree D.F. (2011). Targeting teachers. *Dissent*, 58(3), 9-14.
- Ladd, H. & Fiske, T. (2014, March 6). *A guide to what happened to public education in North Carolina*. Retrieved from <http://dianeravitch.net/2014/03/06/ladd-and-fiske-a-guide-to-what-happened-to-public-education-in-north-carolina/>
- Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal*, 32, 465-491.
- Lasky, S. (2005). A sociocultural approach to understanding teacher identity, agency and professional vulnerability in a context of secondary school reform. *Teaching and Teacher Education*, 21(8), 899-916.
- Leithwood, K. & Montgomery, D. (1982). The role of the elementary principal in program improvement, *Review of Educational Research*, 52(3), pp. 309–339.
- Leithwood, K., Steinbach, R., & Jantzi, D. (2002). School leadership and teachers' motivation to implement accountability policies. *Educational Administration Quarterly*, 38(1), 94-119.
- Lipsky, M. (2010). *Street-level bureaucracy: Dilemmas of the individual in public service*. New York: Russell Sage Foundation.
- Little, O., Goe, L., & Bell, C. (2009). *A practical guide to evaluating teacher effectiveness*. Washington, D.C.: National Comprehensive Center for Teaching Quality. Retrieved from <http://www.tqsource.org/publications/practicalGuide.pdf>
- Liu, C. W. (2011). *The implementation of teacher evaluation for professional development in primary education in Taiwan*. Universal-Publishers.
- Loeb, S., Darling-Hammond, L., & Luczak, J. (2005). How teaching conditions predict teacher turnover in California schools. *Peabody Journal of Education*, 80(3), 44–70.
- Loeb, S. & Miller, L.C. (2006). *A review of state teacher policies: What are they, what are their effects, and what are their implications for school finance?* Retrieved from [http://cepa.stanford.edu/sites/default/files/11-Loeb-Miller\(3-07\).pdf](http://cepa.stanford.edu/sites/default/files/11-Loeb-Miller(3-07).pdf)

- Lord, B. (1994). Teachers' professional development: Critical collegueship and the role of professional communities. In N. Cobb (Ed.), *The future of education: Perspectives on national standards in America* (pp. 175-204). New York: College Entrance Examination Board.
- Lortie, D. (1975). *Schoolteacher: A sociological study*. Chicago: The University of Chicago Press.
- Ma, X., & MacMillan, R. B. (1999). Influences of workplace conditions on teachers' job satisfaction. *The Journal of Educational Research*, 93(1), 39-47.
- Markley, T. (2004). Defining the Effective Teacher: Current Arguments in Education. *Essays in Education*, 11 (3), 1-14. Retrieved from <http://www.usca.edu/essays/vol112004/markey.pdf>
- Marshall, C., & Gerstl-Pepin, C. (2005). *Re-framing educational politics for social justice*. Boston: Pearson Education.
- Marzano, R. J., Frontier, T., & Livingston, D. (2011). *Effective supervision: Supporting the art and science of teaching*. Alexandria, VA: Association for Supervision & Curriculum Development.
- Medley, D. M., Coker, H., & Soar, R. S. (1984). *Measurement-based evaluation of teacher performance*. Longman: New York.
- Merseth, K., Sommer, J. & Dickstein, S. (2008). Bridging Worlds: Changes in Personal and Professional Identities of Pre-Service Urban Teachers. *Teacher Education Quarterly*, 35(3):89-108.
- Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. *Peabody Journal of Education*, 79(4), 33-53.
- Miles, K.H., Odden, A., Fermanich, M., & Archibald, S. (2004). Inside the black box of school district spending on professional development in education: Lessons from comparing five urban districts. *Journal of Education Finance*, 30(1), 1-26. Retrieved from <http://www.financeprojectinfo.org/Publications/insidetheblackbox.pdf>
- New York State Union of Teachers. (2012, February 16). *NYSUT says teacher evaluation agreement is 'good for students and fair to teachers'*. Retrieved from http://www.nysut.org/mediareleases_17505.htm
- No Child Left Behind Act (2001). 20 U.S.C.A. § 1119. Retrieved from <http://www2.ed.gov/policy/elsec/leg/esea02/pg2.html#sec1119>
- Oertwig, S., & Gillanders, C. (2012). *FirstSchool: Improving the School Experience of*

- Young Boys of Color [PowerPoint]. Chapel Hill, NC: FirstSchool.
- Olsen, B., & Sexton, D. (2009). Threat rigidity, school reform, and how teachers view their work inside current education policy contexts. *American Educational Research Journal*, 46(1), 9-44.
- Ozga, J. (2000). *Policy research in educational settings: Contested terrain*. Buckingham: Open University Press.
- Palmer, D., & Snodgrass Rangel, V. (2011). High stakes accountability and policy implementation: teacher decision making in bilingual classrooms in Texas. *Educational Policy*, 25(4), 614 -647.
- Papay, J.P. (2012). Refocusing the debate: Assessing the purposes and tools of teacher evaluation. *Harvard Educational Review*, 82(1): 123-141.
- Peirce, C.S. (1979). *Collected papers*. Cambridge, MA: Belknap.
- Peterson, C. H. (1982). *A century's growth in teacher evaluation in the United States*. Vantage Press, New York,
- Peterson, P., & Comeaux, M. (1990, March 1). Evaluating the Systems: Teachers' Perspectives on Teacher Evaluation. *Educational Evaluation and Policy Analysis*, 12(1), 3-24.
- Pianta, R. C., LaParo, K. M., & Hamre, B. K. (2008). Classroom Assessment Scoring System (CLASS) Pre-K Version. Baltimore, MD: Brookes Publishing.
- Podgursky, M. and Springer, M.G. (2007). Teacher performance pay: A review. *Journal of Policy Analysis and Management*, 26(4).
- Popham, W. J. (1999). Where large-scale assessment is heading and why It shouldn't. *Educational Measurement: Issues and Practice*, 18(3), 13- 17.
- Popham, W. J. (2003). Test better, teach better: The instructional role of assessment. *Association for Supervision & Curriculum Development*, Alexandria, Va. Retrieved from <http://site.ebrary.com.libproxy.lib.unc.edu/lib/uncch/docDetail.action?docID=10047157>
- Popkewitz, T.S. (1998). *Struggling for the soul: The politics of schooling and the construction of the teacher*. New York, NY: Teachers College Press.
- Public Schools of North Carolina. (2012a). *North Carolina Teacher Evaluation Process*. Retrieved from <http://www.ncpublicschools.org/docs/profdev/training/teacher/teacher-eval.pdf>

- Public Schools of North Carolina. (2012b). The NC Teacher Evaluation Process [PowerPoint]. Retrieved from <http://www.ncpublicschools.org/effectiveness-model/ncees/teachers/>
- Public Schools of North Carolina (2013). Education first: NC school report cards. Retrieved from <http://www.ncreportcards.org/src/>
- Public Schools of North Carolina. (2014). Educator effectiveness model: NC educator evaluation system. Retrieved from <http://www.ncpublicschools.org/effectiveness-model/ncees/>
- Ravitch, D. (2012, February 21). *No Student Left Untested*. Retrieved from <http://www.nybooks.com/blogs/nyrblog/2012/feb/21/no-student-left-untested/>
- Ray, A., Bowman, B. & Robbins, J. (2006). Preparing Early Childhood Teachers To Successfully Educate All Children: The Contribution of Four-Year Undergraduate Teacher Preparation Programs. *A Project of the Initiative on Race, Class and Culture in Early Childhood, Final Report to the Foundation for Child Development*, New York, New York.
- Rice, J. K. (2010). *The impact of teacher experience: Examining the evidence and policy implications*. [Brief No. 11]. Washington, DC: National Center for Analysis of Longitudinal Data in Education Research. Retrieved from <http://www.urban.org/uploadedpdf/1001455-impact-teacher-experience.pdf>
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417–458.
- Rockoff, J. E., & Speroni, C. (2010). Subjective and objective evaluations of teacher effectiveness. *The American Economic Review*, 100(2), 261-261.
- Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *Quarterly Journal of Economics*, 125(1), 175-214.
- Rousmaniere, K. (1997). *City Teachers. Teaching and School Reform in Historical Perspective*. New York, Teachers College Press.
- Rowan, B. Comparing teachers' work with work in other occupations: Notes on the professional status of teaching. *Educational Researcher*, 23(6), 4-17, 21.
- Sachs, J. (2001). Teacher professional identity: Competing discourses, competing outcomes. *Journal of Educational Policy*, 16(2), 149-161.
- Sanders, W. L. (2000). Annual CREATE Jason Millman Memorial Lecture: "Value-added assessment from student achievement data: Opportunities and hurdles". *Journal of Personnel Evaluation in Education*, 14(4), 329-39.

- Scribner, J. P. (1999). Professional development: Untangling the influence of work context on teacher learning. *Educational Administration Quarterly*, 38, 238-266.
- Shulman, L.S., & Shulman, J.H. (2004). How and what teachers learn: a shifting perspective. *Curriculum Studies*, 36(2), 257-271.
- Singleton, G.E. & Linton, C. (2005). *Courageous conversations about race: A field guide for achieving equity in schools*. Sherman Oaks, CA: Corwin Press.
- Sloan, K. (2006). Teacher identity and agency in school worlds: Beyond the all-good/all-bad discourse on accountability-explicit curriculum policies. *Curriculum Inquiry*, 36(2), 119-152.
- Smit, B. (2005). Teachers, local knowledge, and policy implementation A qualitative policy-practice inquiry. *Education and Urban Society*, 37(3), 292-306.
- Smith, T.M., & Rowley, K.J. (2005). Enhancing commitment or tightening control: The function of teacher professional development in an era of accountability. *Educational Policy*, 19(1), 126-154.
- Spillane, J. (2004). *Standards deviation: How schools misunderstand educational policy*. Cambridge, MA: Harvard University Press.
- 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.
- Stake, R. E. (1978). The case study method in social inquiry. *Educational Researcher*, 7(2), 5-8.
- Stake, R. (1995). *The art of case research*. Thousand Oaks, CA: Sage Publications.
- Stronge, J. H. (2006). Teacher evaluation and school improvement: Improving the educational landscape. In J. H. Stronge (Ed.), *Evaluating teaching: A guide to current thinking and best practice* (pp. 1-23, 2nd ed.). Thousand Oaks, CA: Corwin.
- Stronge, J. H., & Tucker, P. D. (2003). *Handbook on teacher evaluation: Assessing and improving performance*. Larchmont, NY: Eye On Education.
- StudentsFirst. (2013). Policy agenda: Evaluate teachers based on evidence. Retrieved from <http://www.studentsfirst.org/policy-agenda/entry/teacher-evaluations-and-the-collective-bargaining-process>
- Sullivan, G. P. (2006). *The impact of high stakes testing on curriculum, teaching, and*

- learning*. Blacksburg: Virginia Polytechnic Institute and State University
- Sunderman, G. L., Tracey, C. A., Kim, J., & Orfield, G. (2004). Listening to teachers: Classroom realities and No Child Left Behind. *The Civil Rights Project*. Cambridge, MA: Harvard University.
- Sutherland, L., Howard, S., & Markauskaite, L. (2010). Professional identity creation: Examining the development of beginning preservice teachers' understanding of their work as teachers. *Teaching and Teacher Education*, 26(3), 455-465.
- Sykes, G. (1990). Fostering teacher professionalism in schools. In R.F. Elmore (Ed.), *Restructuring schools: The next generation of educational reform* (pp. 59-66). San Francisco: Jossey Bass.
- The New Teacher Project. (2010). *Teacher Evaluation 2.0*. New York: Author. Retrieved from <http://tntp.org/ideas-and-innovations/view/teacher-evaluation-2.0>
- Taylor, S. (2004). Researching educational policy and change in 'new times': using critical discourse analysis. *Journal of Education Policy*, 19(4), 433-451.
- Taylor, E. S., & Tyler, J. H. (2011). *The effect of evaluation on performance: Evidence from student achievement data of mid-career teachers* [NBER Working Paper No. 16877]. Cambridge, MA: National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w16877>
- U.S. Department of Education (2009). *Race to the Top Program Executive Summary*. Retrieved from <http://www2.ed.gov/programs/racetothetop/executive-summary.pdf>
- U.S. Department of Education (2010a, January). *States' Applications Phase One (North Carolina)*. Retrieved from <https://www2.ed.gov/programs/racetothetop/phase1-applications/north-carolina.pdf>
- U.S. Department of Education (2010b, March 16). *Race to the Top Technical Review Form (Phase One - North Carolina)*. Retrieved from <http://www2.ed.gov/programs/racetothetop/phase1-applications/comments/north-carolina.pdf>
- U.S. Department of Education (2010c, March 29). *Delaware and Tennessee win first Race to the Top grants*. Retrieved from <http://www2.ed.gov/news/pressreleases/2010/03/03292010.html>
- U.S. Department of Education (2010d, June). *States' Applications Phase Two (North Carolina)*. Retrieved from <https://www2.ed.gov/programs/racetothetop/phase2-applications/north-carolina.pdf>
- U.S. Department of Education (2010e, August 10). *Race to the Top Technical Review Form*

- (North Carolina Phase Two). Retrieved from <http://www2.ed.gov/programs/racetothetop/phase2-applications/comments/north-carolina.pdf>
- U.S. Department of Education (2010f, August 24). *Nine states and District of Columbia win second round Race to the Top grants*. Retrieved from <http://www.ed.gov/news/press-releases/nine-states-and-district-columbia-win-second-round-race-top-grants>
- U.S. Department of Education (2012). *Waivers granted under No Child Left Behind*. Retrieved from <http://www2.ed.gov/nclb/freedom/local/flexibility/waiverletters/index.html>
- Urrieta, L. (2007). Figured worlds and education: An introduction to the special issue. *The Urban Review*, 39(2), 107-116.
- Valencia, R. R. (Ed.). (1997). *The evolution of deficit thinking: Educational thought and practice*. Washington, DC: Falmer Press.
- Vasquez, C. & Urzua, A. (2009). Reported speech and reported mental states in mentoring meetings: exploring novice teacher identities. *Research on Language and Social Interaction*, 42, 1-19.
- Wagner, L. (2013, November 20). *Teachers worried and confused over new contract system*. Retrieved from <http://www.ncpolicywatch.com/2013/11/20/teachers-worried-and-confused-over-new-contract-system/#sthash.GU5ZOCWI.dpuf>
- Wang, J., Odell, S., & Schwille, S. (2008). Effects of teacher induction on beginner teachers' teaching: A critical review of the literature. *Journal of Teacher Education*, 59(2), 132-152.
- Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. The New Teacher Project. Retrieved from: <http://widgeteffect.org/downloads/TheWidgetEffect.pdf>
- White, L. G. (1994). Policy analysis as discourse. *Journal of Policy Analysis and Management*, 13(3), 506-525.
- White, J., & Lowenthal, P. R. (2009, Spring). The cyclical rhetoric of educational reform and the rationalization of a failed zeitgeist. *eJournal of Education Policy*. Retrieved from <https://www4.nau.edu/cee/jep/journals.aspx?id=282>
- Woodside-Jiron, H. (2004). Language, power, and participation: Using critical discourse analysis to make sense of public policy. In R. Rogers (Ed.), *New directions in critical discourse analysis: The role of language and learning in social transformation* (pp. 173-206). Mahwah, NJ: Erlbaum.

- Wortham, S. (2005). *Learning identity: The joint emergence of social identification and academic learning*. New York: Cambridge University Press.
- Yan, C., & He, C. (2012). Bridging the implementation gap: an ethnographic study of English teachers' implementation of the curriculum reform in China. *Ethnography and Education*, 7(1), 1-19.
- Yin, R. K. (2009). *Case study research: Design and methods* (4th ed.). Thousand Oaks, CA: Sage.
- Young, M. (1999). Multifocal educational policy research: Towards a method for enhancing traditional educational policy studies. *American Educational Research Journal*, 36(4), 677-714.
- Zeichner, K., & Gore, J. (1990). Teacher socialization. In W. Robert Houston (Ed.), *Handbook of research on teacher education*, 329-348. New York: Macmillan.
- Zeichner, K., Grant, C., Gay, G., Gillette, M., Valli, L., & Villagas, A.M. (1998). A research informed vision of good practice in multicultural teacher education: Design principles. *Theory into Practice*, 37(2), 163-171.
- Zembylas, M. (2003). Interrogating “teacher identity”: Emotion, resistance, and self formation. *Educational Theory*, 53(1), 107-127.