Tennessee Educator Acceleration Model: Teacher Perceptions of One Policy Implementation

A Dissertation
Presented for the
Doctor of Philosophy Degree
The University of Tennessee Knoxville

Renee Melinda Moran
August 2013

Copyright 2013 by Renee Moran

All rights reserved.

DEDICATION

This manuscript is dedicated to my father, Ralph T. Rice. Your day-to-day support and great example of perseverance and work ethic have been paramount in helping me succeed.

ACKNOWLEDGMENTS

I am deeply indebted to my committee members, my family, and my colleagues for their guidance throughout the preparation of this dissertation. I would like to thank Dr. McGill-Franzen for her thoughtful and insightful commentary and support. Just like any wonderful teacher, you always pushed me a step further in my thinking, writing, and research and I am certainly a better researcher and instructor as a result. I would like to thank Dr. Allington for imparting his passion about reading and for providing important examples of scholarship that have furthered my own research. I would like to thank Dr. Rearden for providing me with important real life connections to our public schools as well as always being ready with a word of support. Finally, I would like to thank Dr. Brockett for introducing me to the field of adult learning and reminding me of the importance of this discipline in the area of teacher education.

I would also like to thank my friend and colleague, Monica Billen, for hours of conversation (sometimes work related and sometimes not!) and for being my writing partner. Thank you to my mom and dad for their endless support and willingness to be jacks-of-all trades: babysitters, editors, chauffeurs, cooks, and so much more! I am fully aware that I would never have been able to pursue this dream of mine without your love and hard work. Thank you to my husband, Tim, for your daily love and support for even my craziest of schemes. I can't imagine a better partner in life. Lastly, I would like to thank my two boys, Finn and Conor. You probably won't even remember this time in our lives, but just know that your laughter and kisses often provided me with the much needed respite in the midst of hard days.

ABSTRACT

Teacher effectiveness has become a national conversation and an issue that is debated in both public and educational arenas. Recently, the notion has arisen to quantify teacher effectiveness through the measurement of both teacher and student performance. This study focuses on one state's initial attempt to implement a policy reform that measures a teacher's performance both qualitatively, through scored classroom observations, and quantitatively, through student achievement scores. Ultimately, the idea is that these scores could then be used to make important decisions about salary, retention, and tenure. Using qualitative ethnographic research procedures within a framework of critical theory, I studied 8 first grade teachers and their experiences with Tennessee's Education Acceleration Model (TEAM). I specifically sought to understand the following: 1) the commonalities amongst teachers experiences with the TEAM evaluation, 2) the commonalities in teachers' perceptions of the influence of TEAM on literacy instruction, and 3) how the context of teachers work may complicate teacher perceptions of the influence of TEAM. Interviews, observations, artifacts, and public documents were the primary sources of data for this study.

Results indicated that the implications of policy implementation may be less cut and dry than policy makers might hope. While the hope might be that policy translates directly into change in classroom practice, this study demonstrated otherwise. The amount of change varied greatly from school to school and teacher to teacher and was highly dependent on teacher belief and context. For the most part, teachers tended not to disregard literacy practices they saw as valid just because policy required them to do so. These findings indicate that policy makers might be wise to consider teacher autonomy as well as teacher in-put. Likewise, there are

indications that appropriate professional development might assist in scaffolding a new policy implementation.

Table of Contents

Chapter 1. Introduction 1
Chapter Introduction 1
Statement of the Problem 2
Teacher Effectiveness 2
Purpose of the Study 9
Research Questions 9
Definition of Terms 9
Assumptions 11
Limitations 11
Delimitations 11
Positionality 11
Epistemic Orientation 12
Significance of the Study 13
Organization of the Study 13
Chapter 2. Literature Review 14
Chapter Introduction 14
Section I: Review of the History of Teacher Evaluation 14
Section II: Teacher Beliefs and Policy in a Top-Down Environment 28
Section III: Tennessee's TEAM Evaluation Model 42
Section IV: Critical Research Paradigm 56
Summary of Chapter 64
Chapter 3. Methodology 65
Chapter Introduction 65
Methodological Approach 65

Participant Selection 66
Access and Entry Procedures 67
School and Participant Descriptions 69
Data Collection Methods 77
Data Analysis Procedures 85
Summary of Chapter 93
Chapter 4. Analysis 94
Chapter Introduction 94
Teachers Want To Be Held Accountable 94
Initial Reactions to TEAM 96
Change 98
Triangulation of Beliefs, Daily Observed Teaching, and TEAM Evaluation 110
Administrator Role 116
Subjectivity 122
Scoring 125
Issues of Power 134
Final Issues with the TEAM Model and How the Model Can Be Improved 137
Considering Context 143
Summary of Chapter 148
Chapter 5. Discussion and Implications 150
Chapter Introduction 150
Discussion 150
Implications 155
Epilogue 160
References 164

Vita……180

List of Tables

Table 1: Observation Requirements for Apprentice Teachers……46

Table 2: Observation Requirements for Professionally Licensed Teachers……46

Table 3: Selected Districts, Schools, and Teachers and Corresponding Characteristics…….70

Table 4: Data Collected and Timeline…….78

Table 5: Observations- Literacy Components, Grouping, Time in Classroom……...84

Table 6: Triangulation of Beliefs, Daily Observed Teaching, and TEAM evaluation……..110

Table 7: Type of Administrator and Teacher Change……121

Table 8: Issues of Context Part 1……144

Table 8: Issues of Context Part 2……145

Chapter 1. Introduction

Chapter Introduction

The national conversation about teacher evaluation began with the 1983 publication of the National Commission on Excellence in Education's well-known document *A Nation at Risk*, which called for teachers to show evidence of their capability and for school systems to link salaries to performance. The document stated that "salary, promotion, tenure and retention decisions should be tied to an effective evaluation system that includes peer review so that superior teachers can be rewarded, average ones encouraged, and poor ones either improved or terminated" (National Commission on Excellence in Education, 1983, p. 35). This call for teacher improvement has recently moved into the realm of policy with the implementation of President Obama's 2009 enactment of Race to the Top. RTTT challenged states to raise the bar of teacher effectiveness through the implementation of more rigorous evaluation procedures, which are tied in part to student achievement.

As the state winners of Race to the Top funds begin to implement their new teacher evaluation models, it is vital that research be conducted on the models' effectiveness. Although politicians or administrators most often create educational policy, teachers are generally those responsible for carrying out the policy in the classroom. Therefore, when attempting to understand the effectiveness of a policy, it is important to grasp how policy translates into classroom practice. Through the combination of speaking with teachers about their instructional practices in relation to policy, observing classroom instruction, and examining evaluation and value added scores, I have offered a picture of how policy plays out in the lived experiences of educators and students.

Statement of the Problem

Experts in the educational community have come to recognize the essential nature of the teacher in the equation for student success (Nye, Konstantopoulou, & Hedges, 2004). Research has demonstrated that no other factor is as important in determining future academic progress as the quality of instruction a student receives from the teachers they encounter. As a result, scholars and policy makers have begun to consider how we might use this information to spur teacher improvement and, in turn, boost student achievement.

For policy makers and some researchers the answer lies in legislation that explicitly ties teacher performance to student outcomes and that encourages termination, pay, and tenure bestowal to be based on this data. While many view this connection as the only path towards accountability and achievement, others anguish that these policies may result in unfair judgments and consequences for some teachers.

Teacher Effectiveness

As policy reformers, the public, and those in the educational community have begun more vigorously to connect teacher quality to student achievement, there is growing momentum to describe the characteristics of effective teaching. Some scholars have come to believe that effective teaching is in fact the most important factor in student success. Allington (2002) argued that "it has become clearer that investing in good teaching—whether through making sound hiring decisions or planning effective professional development—is the most research-based strategy available" (p.740). As teacher effectiveness moves into the realm of policy and many states begin to attempt to quantify that effectiveness for high-stakes decisions such as tenure and promotion, it becomes paramount that a reliable definition of commendable teaching be formed. Nonetheless, this definition may be more elusive than policy makers would like to believe. In

this section, I will attempt to illustrate what the research has to say about effective teaching, particularly early literacy instruction.

In our current era, many see the pivotal, large-scale comparisons conducted in the 1960s as the beginning of the great debate on how best to instruct young students in literacy. Dubbed the "First Grade Studies," the hope was that researchers would settle once and for all the best method of reading instruction. While there was some evidence that basal instruction wasn't especially effective in comparison to other modes of instruction and that individual teachers were more significant than any particular program, the results were not definitive. Scholars had not answered the essential question of how best to instruct young readers (Bond & Dykstra, 1967; Pressley, Allington, Wharton-McDonald, Block, & Morrow, 2001).

The conundrum was fueled even further in subsequent decades, as scholars and educators debated the benefits and repercussions of the whole language versus the phonics movements. In 1967 Jean Chall argued for synthetic phonics instruction over the look-say approach. The debate continued with various scholars weighing in with their differing opinions on the matter. As we saw the entre of the pivotal report *A Nation at Risk*, which called for a greater focus on teacher effectiveness, no consensus on good teaching had been agreed upon. Currently, as President Obama's Race to the Top Measures are codified into legislation in several states, it is imperative that we consider the existing scholarly research on effective teaching in order to address the law's serious implications for teachers.

Issues of variability. Numerous research studies have demonstrated that students experience an extremely high variability in quality of instruction from classroom to classroom (Pianta & La Paro, 2003). This appears to be true regardless of program implementation or school effects. In their study of 103 kindergartens, Bryant, Clifford, and Feinberg (1991) found

that only 20% of the classrooms achieved the benchmark for developmentally appropriate instruction and that quality was irrelevant to school size, funding, and the level of teacher experience. In addition, high levels of irregularity were found amongst classrooms. Similar outcomes were noted in Meyer, Waldrop, Hasting, and Linn's 1993 study of 27 kindergarten classrooms in three school districts. Researchers again found a high level of variability, and they found that it directly correlated with student achievement. Particularly, Meyer et al. (1993) highlighted the significance of interchanges between teachers and students as well as the importance of instructional feedback for student achievement.

Studies such as these raised questions about issues of equal access for students in terms of a quality education. Conjointly, they demonstrated the relationship between classroom environment and student outcomes. Stuhlman and Pianta's 2009 study focused on educational quality in 820 first grade classrooms across 32 states. Their findings indicated that the most potent instruction, especially for younger children, involved both the interplay of high academic expectations and an affirmative emotional climate. However, the combination of these factors wasn't necessarily customary in the classrooms studied. They found that 31% of the classrooms exhibited a positive emotional climate with low academic demand, 23% showed high quality in both areas, 28% were mediocre, and 17% demonstrated low quality in terms of both academics and emotional climate. Perhaps the most pertinent finding of Stuhlman and Pianta's (2009) study was that high risk students were least likely to be placed in the classrooms that displayed high quality in both measured areas.

Several scholars have continued this conversation with studies that accentuated instruction for at risk readers. For example, Morrison and Connor (2002) observed that students who entered first grade at risk increased their decoding skills by the end of the year as a result of

explicit instruction on the part of their teachers. In contrast, no effect was noted for low risk readers. In their study of 1364 young readers, Hamre and Pianta (2005) found that at risk students who were placed in high quality classrooms achieved at the same rates as their peers. Conversely, at risk students who were placed in classrooms with low instructional support evidenced notably lower achievement scores than their low risk peers. Quality of emotional support was also a mitigating factor in these results. This may suggest that "focused literacy instruction, high quality feedback, and the engagement of students in discussion of academic concepts may be particularly important in facilitating achievement gains for children with fewer socioeconomic resources" (Hamre & Pianta, 2005, p. 961).

The implications for at risk readers in terms of classroom quality and teacher effectiveness are great. McGill-Franzen, Lanford, and Adams (2002) found high variability, particularly for at risk, emergent readers, in their study of 5 preschools. While some preschools effortlessly connected the literacy environments of home and school, others provided meager resources including less access to books as well as curriculum that was not as rigorous or as culturally congruent. Sadly, in this study those students from low-income families tended to be more likely to be recipients of the latter. McGill-Franzen et al. (2002) noted:

> All teachers initiated children into the literate practices and rituals of their respective communities …. Through interactions with their intimates they learn and share with their classroom community its beliefs about the uses and value of reading and writing, including larger political and social uses as well as personal aesthetic or everyday kinds of functions. (p. 449)

Thus, the literature demonstrated that classroom environments and good teaching truly matter in terms of the progress of students, particularly those who may be at risk for reading

difficulties. I would now like to take a closer look at studies that specifically traced what effective teaching looks like in the early literacy classroom.

Studies of effective literacy instruction. Pressley, Rankin, and Yokoi (1996) conducted a study that identified a sample of primary grades teachers across America who were classified as accomplished in literacy instruction and surveyed them in an attempt to glean what instructional behaviors they viewed as vital. The findings were telling and pointed to the belief that effective instruction may be much more complicated than previously considered. These teachers described their classrooms as being filled with print, books, and stories, with a particular focus on high quality literature. They viewed reading and writing as being integral to all subjects and advocated overlap of content areas. Fifty-four percent of the teachers identified themselves as whole language teachers, and 43% of the teachers identified themselves as partial whole language teachers. Nevertheless, it is important to note that they also advocated some isolated skill practice (an average of 13%) and skills taught within the context of real reading and writing experiences. The teachers reported explicit modeling as essential and engaged in modeling of daily reading, reading aloud, comprehension, and writing processes. Meaning making was emphasized over decoding. Students in these classrooms wrote and journaled frequently, and the teachers encouraged inventive spelling. Students also were instructed in a variety of grouping scenarios, including whole group, small group, and individual instruction. Lastly, the teachers advocated for attention to student motivation and emotional needs as an important aspect of effective teaching. It is also interesting to note what the teachers in the study avoided doing. In general, they refrained from employing ability-based reading and round-robin reading, dismissing both as ineffectual methods.

Ruddell's (1997) work on the attributes of powerful literacy instructors was based on the notion that student success relies heavily on the capabilities of their teachers. Ruddell found that effective teachers applied the following instructional strategies: monitoring students, providing regular feedback, displaying thorough content knowledge, and exhibiting close familiarity with the intricate processes of reading and writing. Teachers in this study simultaneously attended to the social and emotional needs of their students and were energetic, accessible, and understanding.

Wharton-McDonald, Pressley, and Hampston (1998) conducted a study in which they asked administrators and reading specialists to nominate exemplary first grade teachers. The researchers found several commonalities across the classrooms in terms of instruction. There was a conglomeration of skills work in conjunction with authentic reading and writing experiences. Students read and wrote independently on a regular basis and engaged in planning strategies. Group work and collaboration was encouraged. Finally, teachers were described as caring, and positive reinforcement abounded.

After many years of study of exemplary literacy instruction, Allington (2002) also noted presiding characteristics of effective teachers. He argued that great teachers have a better balance of reading and writing versus "stuff." For example, in some classrooms students read and wrote for as much as half of the day; in other classrooms students spent only 10% of the day engaged in literate activities. Effective teachers also gave their students ample expanses of time to read material on their level as well as to engage in productive dialogue with their peers. Additionally, little credence was given to commercial reading programs and low-level tasks. Rather, students were provided with longer assignments that overlapped into various subjects as well as choice in

their activities. Lastly, exemplary teachers tended to evaluate their students based on effort and progress rather than achievement scores.

Implications. Thus, we begin to see that defining effective literacy instruction is a layered and complicated process (Juel & Minden-Cupp, 2000). Nevertheless, effective instruction is an essential aspect when it comes to spurring student improvement. While reformers have focused on student achievement as the primary means of assessing performance, we see indications of the relevance of looking pointedly at classroom quality (Stuhlman & Pianta, 2009). Legislation such as No Child Left Behind has focused on teacher quality. However, state certification standards focus primarily on degree status. Pianta, Belsky, Houts, Morrison, and NICHD (2007) studied 1,000 students across 2,500 classrooms and found that quality of instruction didn't necessarily correlate with a teacher's experience or degree. In many cases, teachers were classified as highly qualified, but their instructional methods were lacking. The researchers found it "troubling that opportunities to learn in classrooms are unrelated to features intended to regulate such opportunities and that students most in need of high quality instruction are unlikely to experience it consistently" (Pianta et al., 2007, p. 1796).

In this section, I have traced some of the pivotal studies relating to teacher effectiveness. I would argue that based on the evidence it is vital that we consider the role of the teacher in terms of student progress. There may be, in fact, no other factor that is more potent in helping students succeed. Nevertheless, based on the complicated nature of effective teaching, we must proceed cautiously when contemplating how such a process can be evaluated and its effectiveness quantified.

Purpose of the Study

The purpose of this study is to examine how teachers perceive one new teacher evaluation policy implementation in Tennessee, the Tennessee Educator Acceleration Model (TEAM), and how those reactions translate into instructional decision making in the literacy classroom. The study was designed to explore the role of the teacher in policy implementation and how teachers view classroom practice as being positively or negatively impacted based on the policy.

Research Questions

The following research questions will be examined in this study:

1. What are the common elements in teachers' experiences with the TEAM evaluation model?
    a. What are the common elements in teachers' perceptions of the influence of TEAM on literacy instruction?
    b. How does the context of teachers work complicate their perceptions of the influence of TEAM on their work?

Definition of Terms

The following definitions are provided in order to clarify important terms throughout the study:

1. Critical Research Paradigm: epistemological view on research that argues that knowledge cannot be separated from the political structures inherent in our society.
2. A Nation at Risk: 1983 document presented by the National Commission on Excellence in Education that warned against the rising mediocrity in American schools.

3. NIET: National Institute for Excellence in Teaching; a website that provides detailed information to Tennessee school districts on how to implement the TEAM Model.
4. Problem framing: Coburn's (2001) notion that the way teachers and administrators frame a particular problem is indicative of the way a policy is carried out in terms of classroom practice.
5. Race to the Top: President Obama's immense grant program created by the Department of Education in 2009.
6. Sensemaking: "This view acknowledges that teachers are confronted with a variety of messages from multiple sources, of which policy is just one, and that teachers must find ways to translate these messages into workable, active classroom resources" (Coburn, 2001, p. 121).
7. TAP Model: The System for Student and Teacher Advancement; developed at Stanford University, a qualitative observation tool currently used in Tennessee for the purposes of teacher evaluation.
8. TEAM: Tennessee Educator Acceleration Model; the state of Tennessee's teacher evaluation model created in response to Race to the Top legislation.
9. TEAM Evaluation System Handbook: provides information on scoring, performance standards, TAP teaching standards, and conferencing for districts and evaluators.
10. Tennessee First to the Top Act: Tennessee state legislation based on the creation of TEAM, which requires that every certified teacher be annually evaluated based on both qualitative and quantitative data.
11. Value Added Assessment: a statistical system developed in 1993 by Dr. William Sanders and used to calculate teacher effect data.

Assumptions

This research was conducted with the assumption that observed student and teacher behavior may provide insights on internal thought processes. In addition, I assumed that the verbal interaction of the interview process allowed me, as a researcher, to form insights based on participants' commentary.

Limitations

The limitations of this qualitative study included a small sample size of 8-10 teachers, that negates the ability to generalize these findings to a larger population. The perspectives of the teachers in the study are unique and may not be representative of the larger educator population as a whole. Therefore, the reader will have to employ his or her own conjecture about the transferability of the results (Firestone, 1993). I focused particularly on first grade teachers in the area of literacy. Therefore, the study does not address other grade levels and subject areas.

Delimitations

I have had no personal contact with the selected participants. However, the study may be constrained by the following researcher imposed delimitation: Teachers were selected in part based on their proximity to me. This impacted the range of teachers and districts included in the study that may in turn have limited perspectives.

Positionality

As I undertake this qualitative study, it is essential that I note my own position and impact on the collection and analysis of the data. I am a white, thirty-four year old female who was raised in a middle class home environment in a very small town in the Appalachian Mountains of North Carolina. I acknowledge that these unchangeable attributes will always influence both my view of the data and the way I interpret it. In addition, it is important to note my former status as an

elementary school teacher and current status as a doctoral student and assistant professor specializing in reading education. Because of my experience, I may be more likely to sympathize with or criticize the teachers I speak with because I bring my own notions of what is right or wrong in terms of education and classroom practice. Additionally, my experience as a former teacher and now instructor of practicing educators may impact the way I view policy implementation. Because of my own experiences with "bandwagon" implementations, I tend to view the ability of policy to make great change without teacher buy-in and long-term sustainability with a skeptical eye. I also question the possibility of quantifying the complex and varied nature of instruction for the purpose of high stakes decisions. All of the above mentioned factors must be accounted for as I go about the interpretive process. My goal is to maintain reflexivity through the process of acknowledgement, transparency, and continual reflection.

Epistemic Orientation

Ultimately, I align myself with the constructivist paradigm because I believe in the notion that I cannot ever truly know or fully understand how someone else constructs reality. There is no one truth, but rather only the truths that I construct either individually or within a social group. In terms of this particular study on teacher evaluation, I apply the constructivist paradigm in that I believe that each interview will only shed light on how each participant experiences the policy change from his or her own view of reality. Ultimately, I can only learn about each of their perspectives, which are local and specific (Hatch, 2002.)

While I have been hesitant to align myself fully with a critical or feminist paradigm, I must also acknowledge my critical leanings. There is a small hope in the back of my mind that my work and those of others might have an impact on inequities that exist in the world of education and in doing so have a transformative effect. If the realities experienced by my participants are

shown to have a negative impact on students and classroom practice, the acknowledgment of these individual truths might rightly spur a positive social change.

Significance of the Study

The notion of linking student achievement directly to teacher evaluation is a topic that has already been much debated in the literature but rarely enacted in the schools. In fact, the implementation of TEAM in the Tennessee schools marked one of the initial undertakings of this enterprise in our country. The implications for teachers in terms of job security and pay are great, and thus these undertakings cannot be taken lightly. It is imperative that scholars investigate and report the outcomes of such teacher evaluation implementations. It is my hope that this study will add to the base of research on teacher evaluation and the results will inform the educational community on the nature of literacy instructional decisions in relation to such a large policy initiative.

Organization of the Study

The purpose of this study is to glean how teachers may or may not alter their methods of literacy instruction based on the implementation of a new teacher evaluation model as well as how teachers perceive the policy implementation. In Chapter 2 I will review the existing literature on the history of teacher evaluation, teacher beliefs, and teacher responses to a top-down policy environment; the specifics of the Tennessee teacher evaluation model (TEAM); and the framework of critical theory. In Chapter 3 I will describe the methodology of my study, including the researcher's role, the population and setting, and data collection and analysis procedures. In Chapter 4 I will analyze the collected data, and in Chapter 5 I will provide thoughts for discussion as well as resulting implications.

Chapter 2. Literature Review

Chapter Introduction

In this chapter I review the scholarly literature in the following areas: history of teacher evaluation, teacher beliefs and the impact of a top-down policy environment, specifics of the Tennessee teacher evaluation model (TEAM), and the framework of critical theory. This review is based on a series of literature searches that I conducted using the online educational and psychological research databases of ERIC and Education Full Text, and using key search terms such as: teacher evaluation, educational policy, critical literacy, and TAP model. Additionally, I reviewed various handbooks and texts that provided insight and led me to further studies of teacher evaluation, policy, and critical literacy. Lastly, I identified and examined specific resources currently used in Tennessee to train and evaluate based on the TEAM model such as the NIET portal and Tennessee First to the Top Website.

Section I: Review of the History of Teacher Evaluation

Teacher evaluation in our country has a long and changing history that spans several decades of practice and research. At its root, teaching may be seen as important and open to scrutiny because, very simply, the public schools are just that—public institutions that receive public funding, and, as a result, many individuals often feel that they, as taxpayers, have a stake in the business of schooling and therefore believe their investment should yield positive results. (Danielson, 2010; Duke, 1995). On perhaps a deeper level, the hope may be that teacher evaluation can be the springboard for improved instruction, which will produce capable and successful students who will ultimately be productive, democratic citizens. I more closely examine how teacher evaluation has changed over the last few decades based on public, policy, and research interests.

Prior to 1980: The teacher, the basic skills generalist. Prior to 1980 there was an assumption in the educational and political communities that to be a teacher one must demonstrate some grasp of content matter as well as a basic skill set and general didactic abilities. Teachers could evidence these faculties by a score of proficient on a basic skills tests or by meeting the requirements on a standardized observation checklist evaluating general pedagogical abilities. Evaluators, or those conducting classroom observations, were not expected to have expertise in the subject they were observing because it didn't impact a teacher's effectiveness score (Shulman, 1987). For example, a teacher's ability to manage a classroom tended to be more pertinent than his or her ability to convey particular subject matter ideas or concepts. As Shulman articulated, teaching was considered little "more than personal style, artful communication [and] knowing some subject matter" (Shulman, 1987, p. 6).

The 1980s. The decade of the 1980s and the Reagan administration forever changed the educational conversation about the role and impact of teachers. A National Commission on Excellence in Education was appointed, and after a series of hearings, public meetings, and sessions by the committee, the report known as *A Nation at Risk* was published. The document was heavily laden with commanding rhetoric, and its authors warned, "the educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a Nation and a people" (National Commission on Excellence in Education, 1983, p. 9). The report likened the mediocrity in American education to an act of war and called for immediate action. Reagan attended to six reforms based on the report, one of which focused particularly on the advancement of good teaching, which was to be produced through pay and promotion based on merit (Cross, 2004).

A Nation at Risk along with reports from the Holmes Group (1986) and the Carnegie Task Force (1986) created a dialogue in America about how to increase the professionalization of teaching, with a particular focus on clarifying and increasing teacher performance standards. Shulman (1987) argued that the previous view on teaching, mainly that it was a generalized and basic skill, worked to trivialize teachers as well as to neglect the individual needs of students. He posited that policy makers had been too quick to grab hold of incomplete research-based descriptions of appropriate teaching, which were then converted to expected instructional competencies and ultimately items on classroom observation scales.

Various researchers during this period argued for the conceptualization of teaching to move away from the generic and toward the notion of teaching as transformative and reflective. Shulman (1987) maintained that the teaching knowledge base is actually a quite complex one that is comprised of the following: scholarship, knowledge of material and settings, understanding of applicable research, and grasp of content. From this viewpoint, conceptualization of the teaching knowledge base would then lead to a new understanding of how to achieve and implement quality instruction.

Fenstermacher (1986) cautioned against using this new information to create a prescribed method for teachers and argued rather that the purpose was to encourage them to use reason to perform educational tasks in a skillful manner. Likewise, Shulman (1987) aligned himself with the call for the professionalization of teaching and concurred with the need for rigor, standards, and change, but he also warned against indoctrination and uncompromising methodology. He noted,

> We have an obligation to raise standards in the interests of improvement and reform, but we must avoid the creation of rigid orthodoxies. We must achieve standards without

> standardization. We must be careful that the knowledge-based approach does not produce an overly technical image of teaching, a scientific enterprise that has lost its soul. (Shulman, 1987, p. 20)

As scholars, policy makers, and the American public began to reformulate their opinions about the role of teachers and their connection to student performance, new vigor was given to the creation of models that might improve teacher effectiveness. One of the most prominent of these attempts was Madeline Hunter's Instructional Theory into Practice Model. Hunter's plan called for teachers to be trained in the following: classroom management, appropriate directions and grouping, monitoring, test-taking strategies, motivation, retention, and the five-point lesson plan. The tenet was that when appropriately trained and evaluated teachers would improve their skill sets and student achievement would rise (Stallings & Krasavage, 1986).

Stallings and Krasavage (1986) conducted a study on the appropriateness of the Hunter model from 1982-1985 by examining two schools in the process of implementation. Using classroom observations, achievement test scores, and teacher interviews, the researchers hoped to glean the relationship between utilization of the model and improved teaching, student engagement, and student achievement. In terms of teacher behavior, they found significant positive change for the first two years of implementation. However, the reverse occurred in the final year of the study as scores dropped 11-12 points. Student engagement and achievement followed the same pattern. Students demonstrated significant improvement in engagement in 1983 and 1984 and a drop in engagement in 1985. Likewise, student achievement increased the two initial years of implementation, and in the final year scores decreased significantly in reading and non-significantly in math, leading to the conclusion that there is not a strong link between the Hunter model and student achievement (Slavin, 1986; Stallings & Krasavage, 1986).

This study is pivotal when contemplating the previously under-considered notion of program maintenance and momentum. Stallings and Krasavage (1986) contemplated a variety of reasons as to why the final year of implementation failed, including changes in the number of staff development days, review sessions, and group-focused meetings. Nonetheless, the study demonstrated that those in the educational community may have misjudged the difficulty in sustaining a program and teachers' interest in it. Interview data showed that some teachers felt that the Hunter model limited their creativity and independence and felt that many of the implementations were too prescribed. The researchers noted,

> Unless teachers internalize teaching strategies so that they feel comfortable using the strategies and teaching is made easier and more fun, teachers will stop using the strategies when the staff developers leave. Strategies that have been learned superficially will eventually be disregarded. (Stallings & Krasavage, 1986, p. 137)

Burden (1986) suggested three stages of teacher career development that may also need to be taken into consideration during program implementation. While teachers at Stage 1 may need more direct supervision, those at Stage 3 may benefit more from less directive clarification and more self-reflection. Based on these stages, teachers in the Stallings and Krasavage (1986) study may have experienced a disparity between their development needs and the trainings implemented.

The 1990s. As the decade of Reagan came to a close, the 1990s continued the previous notion of the importance of goals, standards, and accountability in the public schools. President Bush formed an education summit that included no educators as participants, and the leaders concluded that for too long those in education had ignored the issue of productivity while focusing on less consequential issues such as class size and budgets. The summit leaders called

for an answer to the following question: “Are students learning a year’s worth of education for every year of teaching?” (Cross, 2004, p. 94).

With that mindset as a backdrop, value-added assessment took center stage as a means to calculate teacher effect data. The value-added data system gained in popularity over the last two decades as a viable means of evaluating the performance of teachers, students, and schools. It was originally developed in 1993 by Dr. William Sanders, a statistics professor in the College of Agriculture at the University of Tennessee. The premise behind this methodology was that tracking student progress over a period of years more accurately detects a teacher’s impact on student learning. In the past, attainment-based accountability systems used raw achievement scores to calculate the progress of students. However, some members of the educational community begin to question the fairness of these models since they neglect to take other factors of inequality into consideration. For example, a school with large numbers of children in poverty who may have entered kindergarten with fewer academic experiences might never “catch up” with a school whose students entered kindergarten with higher scores despite an equal demonstration of progress along the way. Value-added assessment worked to adjust this bias by taking such factors into consideration. Schools that exceeded expected growth expectations had high value-added scores, while the converse was true for those that did not meet those expectations. Additionally, this statistical technique employed a system of measurement that considered each teacher’s current student test scores in comparison to those students’ scores in the previous year as well as to other students in that grade. The notion was that you could then isolate each teacher’s contribution to school achievement growth (Costello, Elson, & Schacter, 2008; Harris, 2010).

Collectively, the research community seemed to be in fairly consistent agreement that Sander's value-added model was a vast improvement over the attainment models of the past and that it provided a more equitable means of measuring growth, particularly for low-income school populations. Nonetheless, the agreement ended there as the research and educational communities debated fiercely about the appropriate uses for value-added assessment in relation to teacher evaluation policy (Papay, 2011). Contrasting uses of value-added have been proposed. Scholars have debated whether or not to analyze school versus teacher effects, make high stakes decisions and employ summative assessments or to apply low stakes uses and formative implications (Harris, 2010). In particular, many policy makers viewed value-added as a means of evaluating teacher performance and incorporating a performance pay model (Papay, 2011). Various researchers have questioned the reliability of using such measures for high stakes teacher evaluation (Rothstein, 2007; Rubin et al., 2004).

To further explore this issue, Papay (2011) conducted an important study of 663 teachers in an attempt to understand whether or not value-added assessment could be used as an accurate measure of teacher accountability. Using three different standardized reading tests (the Scholastic Reading Inventory, the Stanford Achievement Test, and the state reading test), Papay found that the impact for teachers would be quite varied based on which outcome of measurement was employed. For instance, in a performance pay model, upwards of one half of the teachers who would earn $7,000 in bonuses when the state test was applied as the criteria for measurement would have lost their bonus if the Scholastic Reading Inventory was used instead. On average, each teacher would have incurred a change of $2,178 in performance pay depending on the assessment used. In addition to test content, Papay (2011) noted that the appearance of

teacher effectiveness, or lack thereof, also may have been impacted by other factors such as the changing nature of student samples and test timing.

In light of these findings, Papay (2011) cautions policy makers against blindly using value-added data to evaluate individual teachers:

> While local administrators and policy makers must decide their tolerance for misclassification, a system that would identify nearly half of all teachers as 'high performing' on one test but not the other … is likely not sufficiently precise for rewarding or sanctioning teachers. (p. 188)

Other scholars concurred with Papay's conclusions. Bracey (2004) took issue with the value-added presentation of teacher effects as "independent, additive, and linear" (p. 2), and argued that teachers of similar caliber could receive different ratings based on the strength of their particular school system. Schochet and Chiang (2010) estimated that using just one year of value-added data could produce error rates of up to 35% in terms of rating teachers.

Nonetheless, Papay (2011) and those in his camp were not suggesting that value-added be abandoned altogether. Actually, they viewed value-added as a viable source of school effects and teacher performance for low-stakes decision making and program evaluation. They cautioned that the use of value-added assessment for high stakes decisions and performance pay may be a prejudiced method of interpretation.

2000-present. Since its creation in 1993, Sander's value-added model has gained in momentum and popularity. The first decade of the millennium assisted in this gain with the enactment of George W. Bush's No Child Left Behind, which pushed states to improve testing and accountability systems and also spotlighted the importance of teacher qualifications (Cross, 2004). The terms of this legislation required that all students be proficient in reading and math by

2014, and, as a result, focus on standards and testing became paramount in most schools (Ravitch, 2010).

While Bush's NCLB plan certainly furthered critique of teacher effectiveness, it has been President Obama's Race to the Top that has explicitly connected teacher performance and evaluation to student growth. This enactment has resulted in the implementation in several states of a model of evaluation that holds teachers responsible for their students' achievement on standardized tests and translates these scores into a score used in their evaluation. The Department of Education has required measures for "determining teachers' contributions to student learning be rigorous, between two specific points in time, and comparable across classrooms" (Adamczyk, 2011, p. 27). For many politicians and policy makers the answer lies in an equation that will determine the success of each teacher based in part on his or her test scores. Race to the Top, the most immense grant program created in the history of the Department of Education, provided a total of 4.35 billion dollars divided amongst participating states. As a prerequisite to funding, states were required to implement a system of teacher evaluation that, in part, ties student growth to teacher performance. According to the plan these evaluation scores could and should be used to make personnel decisions regarding tenure, dismissal, and compensation (Smarick, 2011). Even the National Education Association, which has historically avoided efforts to link teacher accountability to standardized test scores, has joined the debate with a policy statement presented to their governing body addressing the use of "valid, reliable, high-quality standardized tests of student learning in combination with multiple other measures for evaluating teachers" (Sawchuck, 2011, p. 1).

For states that have received a portion of the 4 billion dollars allocated from President Obama's Race to the Top Program, the pressure is especially high because many of them laid out

in their applications plans to revamp their teacher evaluation programs. Some states, such as Georgia and Delaware, have asked for extensions to complete their plans that for the most part tie teacher advancement, including tenure, to high ratings (Cavanagh, 2011).

Other states are jumping head first into the overhaul of their programs. Washington, D.C., for example, has instituted a new evaluation system called *Impact* as a result of Race to the Top. It was projected that 200 to 600 of the city's 4,200 teachers will lose their jobs in the first year as a result of the new system (Otterman, 2011).

In New York, state education officials responded to criticisms of a measure that would allow 20% of teacher evaluation to be based on standardized test scores by increasing that number to 40 %. The governor, Andrew Cuomo, recently wrote, "We must not squander the opportunity to set the right course and make New York a leader in evaluating performance in our education system" (Otterman, 2011, p. 17).

Tennessee, a state that was in the first round of the Race to the Top competition and received $500 million in funds, passed legislation requiring teacher evaluation to be linked to student achievement for the 2011-2012 school year (Heitin, 2011). After three months of field tests, the state implemented the Teacher Advancement Program (TAP), which originates from the National Institute for Excellence in Teaching in Santa Monica, California. As a result of the law, teacher evaluations will be based on the following: 50% on observations, 35% on student growth measures including value added data, and 15% on other measures of student achievement (Heitin, 2011). For teachers who teach in a non-tested subject or grade, their score will be based on school wide results on state tests in reading and math.

Cautionary notes. There is certainly wide agreement amongst scholars and policy makers that the current teacher evaluation system in America is not working (Harris, 2010). As

Clark (1993) pointed out, many teachers continued to receive acceptable marks on their evaluations when it seems that they employed "little signs of creativity, academic or professional growth, or gains in student achievement" (p. 3). On the other hand, teachers engaged in exemplary practice often received similar marks. In an investigation of the Chicago City Schools, The New Teacher Project found that 87% of schools did not issue one unsatisfactory rating between 2003 and 2006.

In addition, in a survey of the largest school districts in the nation, Loup, Garland, Ellett, and Rugutt (1996) found that the results of teacher evaluations were rarely put to use despite research available that demonstrated the usefulness of these materials for staff development purposes. Rather, districts tended to file away the results and neglected to acknowledge successful performance (Colby, Bradshaw, & Joyner, 2002).

Some researchers have questioned whether or not teacher evaluation can effectively do what many view as its primary purpose: improve instructional practice by changing teacher behavior. Improvement may occur if teachers believe that a certain method is appropriate and if they feel a sense of efficacy in its implementation (Wise et al., 1985). Unfortunately, the majority of teacher evaluation systems neglect to tackle the issue of teacher beliefs and change in instructional practice. Fenstermacher (1978) noted that we must take heed of this issue if we are to alter classroom methods that may require addressing what individuals know and can inwardly validate rather than prescribing behavioral change.

Another problem that has been noted in the literature revolves around the issue of evaluator competence. Evaluators (normally administrators) are often generalists and may lack specific content knowledge to evaluate appropriately a teacher's performance in a particular area

of expertise. In addition, principals admitted that they spent little time evaluating competent teachers, and thus these teachers often received very little feedback (Wise et al., 1985).

The issue has also arisen that several short classroom observations may not be the most fair and accurate way of assessing a teachers performance over time. Rather, a more holistic assortment of data that combines a long-term compilation of student work, lesson plans, and other performance data should be employed (Wise et al., 1985).

Regardless of the issues of the current system, data abound that demonstrate that good teachers matter in terms of student success rates. For example, in his longitudinal study of first and fourth grade teachers in six states, Allington (2002) found that exemplary teaching was much more effective in ensuring student success than any curricular approach or program and ultimately was the "most research-based strategy available" (p. 740).

Thus, teacher improvement may be a tenable means to boost student achievement. Nevertheless, it seems we must be especially cautious when considering what types of teacher evaluation systems are most appropriate in connecting these two phenomena. As I noted earlier, Papay (2011) demonstrated the tenuous relationship between the type of test employed and value-added scores used for teacher effects. McGill-Franzen and Allington (2006) also pointed to widespread contamination in accountability systems, including the impacts of summer reading loss, in-grade retention, test preparation, and testing accommodations, as sources that ultimately sabotage the dependability of student achievement scores. The question remains: Is this test on which we are basing such high-stakes decisions for teachers, such as tenure, promotion, and pay, truly a reliable source of information?

Valli, Croninger, and Walters (2007) further complicated the issue with their study that asked: "Just who is doing the teaching?" They collected data from a large and very diverse

school system over the 2003-2004 school year. Their report indicated that it may be more difficult than previously considered to isolate an individual teacher's effects based on the influx of different individuals who work with students over the course of a school year. In the study, 17 of the 18 schools engaged in some type of simultaneous instruction during which multiple adults in the classroom setting influenced what students learn. Sixteen of the 18 schools employed supplemental instruction, which means that students received additional instruction outside of the regular classroom. This finding aligned with that of Allington and Cunningham (1996), who found that Title I teachers report spending at least one-half of instructional time in pull-out environments. These studies pointed to the notion that it may be more difficult than policymakers imagined to accurately assess the responsibility of an individual teacher in imparting particular knowledge to his or her students. In contrast, it seems that a variety of actors have a role in the advancement of student achievement (Croninger & Valli, 2009).

What we know that works. Wise et al. (1985) posited that when teacher evaluation is implemented properly, it can spur positive change in terms of teacher improvement. They, along with other researchers, have recommended that before selecting a system of evaluation, a district should carefully consider what they want the evaluation to accomplish as well as how their school system and community view the purpose of teaching. In this manner, a district must decide on a theoretical conception of teaching. For example, two contrasting theoretical views would be the conception of teaching as labor versus teaching as art. In the labor model, teaching would be routinized, and lesson plans and student performance would be scrutinized to conclude whether or not teachers are following appropriate concrete criteria (Mitchell & Kerchner, 1983). In an opposing perspective, such as teaching as art, teaching may be seen as unorthodox and based on the individual personality traits of the teachers. Ultimately, these researchers believe it

would be inappropriate for a district to choose an evaluation system that is misaligned to the common sentiment of the members of their respective community. The district should also consider what purpose the evaluation is to fulfill. Is it solely for instructional improvement or also for purposes of termination? Regardless, school systems should never assume that because an evaluation system was successful in another district that it necessarily will be in their own without first carefully reviewing the needs and purposes of their own community (Wise et al., 1985).

The purpose of the evaluation should not only be clear to all parties involved in the process; it should also be agreed upon by all (Colby, Bradshaw, & Joyner, 2002). Pembroke and Goedert (1982) contended that teachers must see evaluation as fair and objective. This viewpoint advocated for teachers to be a part of the decision making process in terms of the creation and implementation of the evaluation. In their survey of teacher evaluation practices in thirty-two school districts, Wise et al. (1985) found that when a school moved away from an "adversarial" approach toward a more "participatory" approach teachers viewed both their rights and obligations as increasing. In turn, this increased teachers' sense of obligation or responsibility and along with teacher participation, enhanced the quality of the evaluation system. It is also essential that educators view the evaluation process as worthwhile and as an avenue for improvement of practice. Fuller, Wood, Rapoport, and Dornbusch (1982) referred to this as performance efficacy in which teachers have some authority over assessment of their work. Danielson (2010) advised moving away from the traditional models in which the evaluator observes, makes judgments, and then passes the advice on to the teacher. Rather, she advocated for a more participatory process in which the teacher analyzes the evaluator's notes and then reaches his or her own conclusions about the strengths and weaknesses of the lesson. Afterwards

the observer and teacher engage in a productive discussion about the lesson. The criteria for the purpose of evaluating should be both meaningful and reliable, and school districts must be willing to allocate both time and resources to ensure proper implementation (Colby et al., 2002).

Research has also demonstrated that the evaluator matters and that teachers view evaluators with insights on curriculum as more valuable to their developmental process (Atkins, 1996). In conjunction with this notion of the importance of an evaluator's knowledge base is the significance of effective feedback as a component in the implementation of an advantageous evaluation system. Teachers perceived the quality and nature of the feedback provided by the evaluator to be an invaluable part of their growth (Colby et al., 2002). The feedback resulting from teacher evaluations has been shown to improve communication between teachers and administrators, and thus it enhanced teachers' satisfaction with the implementation of an evaluation system (Wise et al, 1985).

Concluding thoughts. In this section, I have traced the history of the last three decades of teacher evaluation. I focused particularly on how the push for accountability and measures of teacher effect impact the lens with which policymakers view teacher performance.

Section II: Teacher Beliefs and Policy in a Top-Down Environment

In the last few decades of American educational history, there has been a notable shift from local control of school policy to federalization of decision-making in our schools (Allington, 2002). The first half of the 20th century was characterized by a largely hands-off approach by both the state and federal governments, however the enactment of the Elementary and Secondary Education Act (ESEA) Title I altered this (McGill-Franzen, 2000). In the 1960s President Johnson viewed the legislation as a means to filter money into districts with large numbers of low-income, high-need students. For many the enactment symbolized the

encroachment of the federal government upon the jurisdiction of the local school districts (Cross, 2004). Congressman John William lamented, “The needy are being used as a wedge to open the floodgates and you may be absolutely certain that the flood of Federal control is ready to sweep the land” (Sundquist, 1968, p. 215).

ESEA was followed in the 1970s by another important piece of legislation, PL 94-142, which required disabled students to have access to public education. The law resulted in immense increases in the number of special-education students. Originally, the idea was that the federal government would share in the monetary responsibility for disabled students, but ultimately the legislation was underfunded, which left school districts scrambling to allocate their resources appropriately (Cross, 2004; McGill-Franzen, 2000).

While ESEA and PL 94-142 certainly increased the federal government’s role in education, on the whole the focus was on funding rather than accountability. This focus changed in the 1980s with the notion of systemic reform, which promoted the view that not only is the government capable of improving the educational system, they are responsible for supplying authorizations and incentives to further policy (McGill-Franzen, 2000).

While Shulman (1983) noted that many federal mandates such as ESEA and PL 94-142 stemmed from the desire to protect powerless groups, it is foolhardy to believe that the process is so simple as the issuance of a directive that then results in change. He argued:

> We will have to examine what other policies may conflict with that in question, what bathwater of unintended outcomes may accompany the baby of this remedy, and what undesirable effects may develop because of the manner in which the innovation, desirable though it may be in principle, has been imposed unilaterally on members of a profession. (Shulman, 1983, p. 494)

McGill-Franzen (2000) concurred, declaring that the complexity of teaching lends itself to unpredictability in terms of policy outcomes. Based on these interpretations, I more closely examine assumptions about policy in a top-down environment as well as teacher beliefs and teachers' capacity for change. Lastly, I trace several seminal studies that considered the outcomes of policy implementation in schools and future directions that the mounting knowledge base on this issue may lead.

Assumptions about top-down policy. In a top-down policy environment on the surface, the implications are clear. Those in authority create and/or implement a particular educational policy. Teachers are informed or instructed as to how to execute the policy correctly. Teachers change their classroom practice to fit the policy, and a change in student behavior and achievement should occur. In reality the success of policy implementation is far more layered and complicated. When these complications are not taken into consideration, policy makers and administrators may be puzzled when well-intentioned reform efforts fail (Toll, 2001).

Despite the assumptions made about the implications of policy, research has suggested that even with official frameworks, curricula, and mandates these measures often fail behind the closed doors of classrooms for both intentional and unintentional reasons (Cuban, 1995). Various ideas have been posited as explanations for the deficiencies of many policy reforms. In a 1984 study, Cuthbert interviewed classroom teachers about the implementation of federal and state policy initiatives in concordance with career, gifted, and special education. Her findings pointed to teachers' lack of understanding of the policy implementations, which resulted in the absence of proper classroom utilization.

Wise (1987) posited that the distance of policy makers from the actual classroom environment may impact the success of reform efforts. In concurrence, Eisenhart, Cuthbert,

Shrum, and Harding (2001) noted that the lived world of policy makers, which revolves around approval from the public, budgetary requirements, and bureaucratic implications, may be far removed from the lived worlds of teachers and students who focus on curriculum and day-to-day classroom social interaction. These researchers argued that this disconnect may make the road from policy creation to policy fulfillment a long and arduous one.

For Toll (2001), the lack of success of policy reform had much to do with the differing discourses employed by policy makers and teachers. She contended that the two groups simply weren't speaking the same language and that this division impeded policy implementation. In her study, she compared two opposing sources of documentation—interviews conducted with teachers and a policy document entitled the "National Reading Panel Report"—to demonstrate the contrasting language of teachers and policy makers. In multiple interviews with classroom teachers, Toll (2001) found that they based instructional decisions primarily on engagement with students and concern for children's affect. In addition, teachers reported the value of having control over their own choices in terms of classroom decision-making. In contrast, her analysis of the National Reading Panel Report emphasized discourse that focused little on the elements that were valued highly by the teachers in the study (Toll, 2001). Rather, the discourse in the document suggested a heavier emphasis on the following: the necessity of objective research to inform teacher judgment, the essential nature of student on-task behavior, and the notion of teachers in the passive role of consumers of information. In essence, reform failure may be directly related to competing discourse between these two differing groups (Toll, 2001).

Perhaps the most notable element in the research, when trying to understand the success or failure of top-down policy reform, was the relationship of top-down policy to teacher beliefs. Research has demonstrated that even well-meaning policy plans fail when teacher beliefs are not

taken into account because these beliefs are at the core of instructional decision making (Clark & Peterson, 1986; Eisenhart et al., 2001). It seems that whether or not policy becomes a viable part of classroom practice depends heavily on individual teacher beliefs. In fact, whether purposely or not, educators ignored or amended policy implications in concordance with their beliefs (Eisenhart et al., 2001). Despite a wealth of research on the influence of teacher beliefs, most school improvement efforts continue to focus on changing only the behavior of educators rather than working on both beliefs and behaviors (Guerra & Nelson, 2009). Guerra and Nelson (2009) cited this omission as the primary rationalization for lack of change in educational outcomes particularly for students who are diverse in terms of language, economics, or culture. In accordance with this line of thinking, I more closely examine the issue of teacher beliefs and the implications for classroom instruction.

Teacher beliefs. Generally, beliefs are perceptions held by particular individuals in regards to their world view. According to Begum (2012), "beliefs are psychologically held understandings, premises or propositions about the world that are felt to be true" (p. 16). Beliefs are commonly fixed and are not easily changed by outside influences (Nespor, 1987; Ogan-Bekiroglue & Akkoc, 2009). In addition, beliefs don't necessarily have to be endorsed by others for an individual to hold fast to them (Haney & McArthur, 2002). According to Helle and Murtonen (2001) beliefs are comprised of a fusion between each person's subjective experiences and the learning and knowledge they acquire throughout their lifetime.

Green (1971) distinguished between three differing elements of belief that included the following: primary beliefs, derivative beliefs, and core beliefs. From this view, primary beliefs are unaffected by other beliefs and determined from each person's manner of operating.

Derivative beliefs are the direct result of primary beliefs. Core beliefs are the least adaptable to change and influenced by an individual's personality.

Kagan (1990) emphasized that within the framework of belief it is important to distinguish specifically between general beliefs and teacher beliefs. Kagan (1990) defined teacher beliefs as "the highly personal ways in which a teacher understands classrooms, students, the nature of learning, the teacher's role in the classroom and the goals of education" (p. 423). Teachers' belief systems may be both subjective and objective, and they are often informed by a variety of life experiences. In this way, individuals who have chosen teaching as their career path have formed many layered beliefs about education before they ever enter their first education course (Lortie, 1975). According to Richardson (1996), these life experiences are just one component of what makes up a teacher's belief system. It is also impacted by what each individual encountered in his or her own experiences with formal or informal schooling or learning. Lastly, beliefs are informed by the pedagogical knowledge that they have acquired or continue to acquire through new information (Richardson, 1996). Begum (2012) added to Richardson's list when noting that beliefs are also influenced by teachers' own experiences in the classroom in terms of what works and what doesn't as well the traditions embedded in their own school communities. Additionally, teacher beliefs may be affected by individual personality traits, demographic characteristics and attributes, and level of education (Decker & Rimm-Kaufman, 2008).

As the research shows, a teacher's belief system is multi-layered and complex and almost always influences classroom practice. As noted by Richards and Lockhart (1994), "teachers' beliefs and values serve as the background to much of their decision making and actions, and hence constitute what has been termed the culture of teaching" (p. 30). In other words, a

teacher's belief system will guide what he or she does and says in the classroom and what he or she deem important or inconsequential in terms of classroom instruction (Begum, 2012).

In this way, the intersection between the problems of policy reform and teacher belief becomes apparent. For instance, if a policy implementation or curricular framework does not align with a teacher's belief system, then he or she may be less likely to implement effectively the change in the classroom. Furthermore, teachers may change the way they teach to match more closely with what they believe about appropriate instruction. Likewise, a teacher may spend more instructional time on a classroom activity that he or she views as credible in terms of furthering student learning (Powers, Zippay, & Butler, 2006). In this manner, beliefs will always influence the way programs are carried out and, in many cases, student achievement and investment.

> What teachers teach may be determined by an authority separate from the teacher, but the way the teachers interpret the curriculum or the syllabus, the way they enact the curriculum in the classroom context, is strongly influenced by their belief regarding what and how the students should be taught. (Begum, 2012, p. 17)

Teacher beliefs or world views may impinge upon a teacher's willingness to embrace a particular policy from the outset. For example, an educator's world view may lean on the cyclical characteristics of some teaching modifications in which a person believes that there is nothing new to be learned (Buehl & Fives, 2009). In their study on teacher beliefs, Buehl and Fives (2009) found that teachers had very different views on knowledge construction and that these views impacted whether an individual embraced or rejected a particular program or policy. For example, some participants in the study believed that teaching would change because knowledge is constantly changing. They attributed new knowledge to advancements in science,

technology, and student learning. On the other end of the spectrum were participants who viewed knowledge as stagnant. These individuals' beliefs could very well impact the fidelity of a policy implementation. According to Buehl and Fives (2009), if

> individuals view actual teaching experience as the only legitimate source of teaching knowledge, or if they believe that there is nothing new in teaching, they will be less likely to attend to, process, or implement the new information or techniques that are presented through workshops, courses, or readings. (p. 402)

One might begin to wonder why policy makers, with such a wealth of resources on the importance of the relationship between teacher beliefs and classroom instruction available, have neglected to address this component of the equation. It may be that policy makers lean on the old mantra that if practice changes, then teacher beliefs will eventually follow. It could be that their agenda for change lies more heavily in the arena of students and families rather than teachers (Guerra & Nelson, 2009). Alternately, it could be an even more simple answer—teacher beliefs are just not an easy issue to tackle. As researchers such as Ertmer (2005) have made clear, beliefs are deeply ingrained and are not easily altered by means of persuasion. Certainly, singular workshops and badly executed professional development won't spark change and may even increase resistance (Guerra & Nelson, 2009). Accordingly, I ask: can beliefs change, and, if so, how?

Can teacher beliefs change? As the literature has shown, beliefs can be quite entrenched and are a rooted phenomena that often exist beyond the influence of persuasion or group influence. Based on this notion, it is quite reasonable to wonder if beliefs can be addressed at all when considering educational change. Nonetheless, while beliefs may be difficult to alter, this difficulty does not mean that it is impossible to change a person's beliefs. Nespor (1987)

contended that the process of change is not incurred through dispute or rationality but rather through a process referred to as a Gestalt shift. There is also some agreement in the research community that this conversion process must be preceded by an individual's questioning of his or her existing belief system (Ertmer, 2005; Posner, Strike, Hewson, & Gertzog, 1982). This manner of questioning often occurs when a person is presented with new information and begins to examine his or her own views in light of new knowledge. According to Kagan (1990), this notion has implications for professional development. He noted that professional development must "require [teachers] to make their pre-existing personal beliefs explicit; it must challenge the adequacy of those beliefs; and it must give novices extended opportunities to examine, elaborate, and integrate new information into their existing belief systems" (Kagan, 1990, p. 77).

In conjunction with confrontation of prior beliefs, researchers have identified three additional strategies that may facilitate change in teacher beliefs. These strategies include: personal experience, vicarious experience, and socio-cultural influence. The assumption behind the first strategy, personal experience, is that because experience assisted in the creation of belief it might also facilitate change (Cuban, 1986). In this model, teachers would be given the opportunity to explore new strategies over the course of multiple personal experiences (Ertmer, 2005). For example, a teacher might be introduced to a novel approach and then try it out in a supportive environment.

The second strategy, vicarious experience, leans on the notion that observation can be a compelling and worthwhile tool to facilitate change (Ertmer, 2005; Schunk, 2000). In this framework, teachers would be given multiple opportunities to view someone else modeling a new method. Elmore, Peterson, and McCarthy (1996) argued that "teachers' practices are

unlikely to change without some exposure to what teaching actually looks like when it's being done differently" (p. 241).

Lastly, it seems that socio-cultural influences may have a significant effect on the alteration of teacher beliefs. According to this theory, the manner in which teachers engage in their world is always heavily influenced not only by their daily classroom interactions but also by the views and assumptions of those around them. This may give weight to the importance of participation in dialogue or professional communities with peers in order to support one another in the acquisition and processing of new procedures, approaches, and thought processes (Ertmer, 2005).

Hence, we see that the process of belief change is not a simple one, but it is not an impossible one either. As Nespor (1987) illustrated, we cannot expect teachers to simply and wholly abandon their belief systems; rather, we should be aware the possibility may exist for gradually displacing old beliefs with new ones.

Studies of policy implementation. Several studies have been conducted that examined how policy implementation may change as it filters down from the federal and state governments to individual school districts and ultimately individual classrooms. In this section, I will focus particularly on studies by Spillane (1999) and Coburn (2001) that support the argument that policy implementation may be more layered and complicated than previously thought.

In his study, Spillane (1999) attended to the relationship between state and local policy in Michigan. Through a combination of qualitative and quantitative methods he examined how teachers viewed, internalized (or did not), and enacted (or did not) policy reforms. The results of this study indicated that while all teachers claimed to have applied the reforms, observations demonstrated something very different. In fact, classroom change varied from non-existent to

modest to profound, and there appeared to be little relationship between the use of innovative materials and a desire for change. Rather, Spillane (1999) argued that the success of policy implementation relied heavily on each teacher's zone of enactment, which he defines as the "space where reform initiatives are encountered by the world of practitioners and practice … [where] teachers notice, construe, construct, and operationalize the instructional ideas advocated by reformers" (p. 144). Based on this, he argued that the individual resources of the teacher matter, including prior knowledge, disposition, and beliefs, and that enactment zones can serve as mediating forces between policy implementation and classroom practice.

Coburn's (2001) work extended this notion of attention to teacher effect in terms of policy implementation by consideration of how teachers choose to implement or disregard a particular mandate. She contended that while some researchers suggest that policy influences teachers' work, it is more probable that the reverse occurs and "teachers interpret, adapt, and even transform reforms as they put them into place" (p. 145). Coburn studied a California school system during the late 1990s in the midst of the state's sweeping reading accountability reforms to better understand these phenomena. The results of the study demonstrated that teachers often made sense of new policy through the process of collective sensemaking, in other words, through both formal and informal conversations and interactions with their peers. In this manner, messages from policy makers were reconstructed and then either attended to or disregarded. Coburn (2001) noted:

> From a policy maker's perspective it may seem that schools and districts in reconstructing and reinterpreting policy messages are subverting the intent of policy or thwarting implementation …. But another way to look at it is that this sensemaking is both necessary and unavoidable. (p. 153).

This view acknowledges that teachers are confronted with a variety of messages from multiple sources and that they must find ways to translate these messages into workable, active classroom resources.

Coburn (2006) built on the issue of sensemaking with an examination of problem framing as well. In a continuation of a look at the policy implementations in California, she found that the way teachers and administrators framed a particular problem often impacted the way a policy was carried out in terms of classroom practice. For example, through informal and formal dialogue as well as principal influence, one California school reframed a policy implementation to focus on comprehension instruction rather than increased phonics as was originally intended in the state legislation. As a result, the faculty worked to direct classroom instruction as well as professional development toward comprehension rather than toward the original intention of the policy.

Coburn's work may point towards framing (2006) and collective sensemaking (2001) as having potentiality for successful reform efforts. As of yet, policy makers have not acknowledged the potential of stimulating collaboration in schools, but collaboration could be a viable means of making change. Nonetheless, based on Coburn's (2001, 2006) results, it is essential that these interactions are not stilted or contrived because that will simply work to undermine the effort of the reform.

Spillane's (1996) work also reminded demonstrated that local leadership may be an impactful force in terms of influencing the way teachers enact policy. In case studies of two districts, he explored how state policy influenced each in terms of changing reading instruction. Although the influence of school districts has been largely ignored in regards to policy change, the opposite may actually be true. Both districts studied contrived their own reading policies,

which were based on a variety of sources in addition to the state reading policy. In fact, one district whose leaders were in opposition to the state policy created adverse instructional guidelines for teachers in that system. The result was that teachers in this district received an abundance of conflicting messages in regards to how to implement reading instruction. Adding another layer to this conundrum was the finding that the school administrators also impacted which messages teachers received.

Spillane's (1996) findings pointed to the necessity of not discounting the impact of district leaders in the implementation of policy. Administrators have the ability to either undermine or promote state and federal policy simply by the power and choice of dissemination of information to teachers.

The studies I highlighted call for attention to the issues of collaboration, district control, and teaching readiness and knowledge when considering policy implementation. Ultimately, teachers are the final "policy brokers" (Spillane, 1999, p. 144), and their beliefs and capacity for change must be considered. In the following concluding section I touch briefly on possible future directions to take into account when considering how to effectuate policy.

Future directions. Carlone, Haun-Frank, and Kimmel (2010) identified schools as locales of constant struggle in which teachers experience "tensions between their everyday, local practices and larger social, historical, and political structures of schooling" (p. 947). From their stance, rigid policies undermine professionalism, impede quality instruction, and erase teacher autonomy. On the other hand, other researchers contended that a solely bottom up approach will not provide adequate support or community for teachers (Weathers, 2011). From this view, there is a clear role for leadership in supporting teacher improvement and growth. Eisenhart et al.,

(2001) concurred, contending that policy makers shouldn't simply take a hands-off, laissez-faire approach. Consequently, I ask: Is there a middle ground?

In a 1983 study Heath demonstrated that teachers were more likely to take ownership and react positively to policy when it aligned with their beliefs and when they had a say in the details of implementation. Teachers in Heath's (1983) study were given the freedom to design a variety of innovative classroom and community interventions, and the results included greater teacher satisfaction as well as improved student and family involvement. When stricter policy was imposed that did not align with teacher beliefs, teacher creativity plummeted. Therefore, it seems that policy may have a better chance of succeeding when teachers are actually a part of the research or implementation process (Hall, 2005).

Research also suggested that policy may be better implemented when teachers receive appropriate support and scaffolding throughout the process (Theirot & Tice, 2009). Wedman and Robinson (1988) found that educators benefitted most from exposure to new methods when they were given opportunities to observe new strategies, implement them, and then engage in discussion afterwards.

It seems that teachers who have a greater sense of agency or control over their own actions tend to be happier and more devoted in their profession. This autonomy can even lead to higher levels of achievement and motivation amongst students as well as a more beneficial learning environment (Malmberg & Hagger, 2009). As Allington (2002) pointed out:

> Sufficient autonomy is absolutely essential for teachers to accept individual professional responsibility for student outcomes—autonomy in making decisions about how, what, and when children will be taught. If someone else tells you to follow their plan, any failure becomes a failure of their plan, not yours. (p. 35)

Shulman (1983), likewise, has worked through this conundrum in an attempt to identify how policy can protect the powerless yet maintain autonomy. He proposed a loose coupling of the two in which the educator is presented with a variety of well-thought-out directives yet essentially has the dominion to accept or reject these directives. In addition, based on the ever-broadening wealth of knowledge about teachers' influence on policy, it may be wise to include them in the creation of these policies (Shulman, 1983). Regardless, the research I presented in this review suggests that a view of policy implementation as a simple equation of mandate = change is outdated and counterproductive. As Jennings (1996) noted, "the worth of a policy is in what teachers learn from it" (p. 108). Policy makers must begin to consider the multifaceted layers of policy and the ripples of effect that will undoubtedly result from it.

Section III: Tennessee's TEAM Evaluation Model

Method of Review. To better understand Tennessee's new teacher evaluation model, TEAM, I began by reviewing the NIET (National Institute for Excellence in Teaching) Best Practices Center Portal. This website provides school districts detailed information on how to implement the TEAM Model. On this website I was able to view sample lessons in a variety of subject areas and grades as well as corresponding pre- and post-conferences. In addition, training modules were provided in the domains of instruction, environment, planning, coaching, and post-conferencing. I was also able to download and review the TEAM Evaluation System Handbook, which provided detailed information and applicable forms on the following: scoring, performance standards, TAP teaching standards, conferencing, and research.

After reviewing the NIET portal, I spent time examining the TAP (The System for Teacher and Student Advancement) website since TEAM is based in part on the TAP model. This website provided information on research conducted to support the model as well as some

insights on value-added, legislation, and outcomes. In addition, I looked at Tennessee's First to the Top website which provided updates on the implementation process.

Lastly, I searched for current research on the TAP model. I found a few studies that reported on the results of TAP in various school systems. Not surprisingly, I was unable to find, as of yet, any studies based on the TEAM model.

What is the TEAM evaluation model? When President Obama enacted Race to the Top it became the most immense grant program created in the history of the Department Education. The program provided a total of 4.35 billion dollars divided amongst participating states. In response to receipt of funding, one component required such states implement a system of teacher evaluation that ties student growth to teacher performance. According to the plan, these evaluation scores can and should be used to make personnel decisions regarding tenure, dismissal, and compensation (Smarick, 2011).

Forty-one states as well as the District of Columbia submitted Race to the Top applications. Most states agreed to implement both measures of student growth as well as annual teacher evaluations. However, only about half of the states that entered the competition agreed to incorporate student growth measures in teacher evaluation. Tennessee was one of these states, and, along with Delaware, was chosen as a first-round winner of Race to the Top by the Department of Education (Smarick, 2011). As a result, Tennessee was awarded 501 million dollars in funding and began the process of creating a teacher evaluation model that would meet the requirements laid out by Race to the Top and the Department of Education (TN First to the Top Website, 2012).

Tennessee's plan, which has been dubbed "Tennessee First to the Top," was based on the creation of the Tennessee Educator Acceleration Model (TEAM). As a compulsory part of

Tennessee's First to the Top Act, every certified teacher is evaluated annually based on multiple measures. These multiple measures are a combination of both quantitative and qualitative data. The qualitative element applies the TAP model (The System for Student and Teacher Advancement), an observation tool that uses student growth measures as a component of teacher evaluation, as a foundation. Tennessee largely based this decision on research that linked student achievement improvement to the use of the TAP model. Districts are required to employ the TAP performance rubric as a means of evaluating teacher performance unless a district chooses to execute another form of teacher evaluation. In this case, the evaluation tool has to be approved by the state of Tennessee (NIET website, 2012).

The TAP model rubric addresses fifty percent of the evaluation component using data from classroom observations. Thirty-five percent of the multiple measures are based on student growth. All educators employed in a grade or subject tested annually apply individual teacher-effect data to their evaluation score. Teachers employed in non-tested grades or subjects are given a score based on school-wide value-added data. The remaining fifteen percent of a teacher's evaluation score is derived from student achievement data. The achievement data has to be in compliance with the selections deemed appropriate by the state board (NIET website, 2012).

The three measures of student growth, student achievement, and observation data are then combined into a single rating for each individual teacher. The rating is computed using a web-based platform provided by the state. Teachers are scored on a scale ranging from one to five with the following specifications: (1) significantly below expectations, (2) below expectations, (3) at expectations, (4) above expectations, and (5) significantly above expectations. Again, per

federal guidelines, each state is compelled to use these data to make decisions regarding tenure, professional development needs, and removal from employment (NIET website, 2012).

***Who will be responsible for conducting observations and making scoring decisions*?** Those considered qualified to conduct the observations, which translate into a score for the qualitative component of the evaluation, include the following: principals, assistant principals, and instructional leaders. Instructional leaders were not explicitly defined. However, each evaluator is required to become successfully certified in the Tennessee Educator Acceleration Model through several means. Evaluators are required to attend a four day certification training as well as perform favorably on the NIET online certification exam. The exam consists of two parts. First, each potential evaluator reviews an online lesson and decides on an appropriate score for that teacher. In order to pass the test, his or her score has to range within one point above or below the yardstick total rubric score provided by the online system. In addition, the evaluator has to score within one point for at least seventy-five percent of the rubric indicators. Lastly, the applicant is required to answer a minimum of six out of eight multiple-choice questions correctly to receive certification. If all of the above components are successfully completed, the evaluator obtains a certificate of completion that entitles him or her to conduct observations and administer scores. This certification is valid for a single year (NIET website, 2012).

A closer look at the qualitative component of the model. As stated previously, under the First to the Top Act, each teacher in Tennessee is evaluated and administered a rating based on data from observations of his or her classroom teaching. The number of required observations varied based on years of teaching experience. The following charts illustrate the observation requirements for both apprentice and professionally licensed teachers (NIET website, 2012).

Table 1: Observation Requirements for Apprentice Teachers

Suggested Sequence	Type	Length	Rubric	Pre-Conference	Post-Conference Type
First	Announced	15 min	Planning Rubric	Yes	Informal
Second	Unannounced	Lesson + 15 min	Instruction Rubric and Environment Rubric	No	Formal
Third	Unannounced	Lesson	Instruction Rubric	No	Formal
Fourth	Announced	Lesson	Instruction Rubric and Planning Rubric + Environment Rubric	Yes	Formal

Table 2: Observation Requirements for Professionally Licensed Teachers

Suggested Sequence	Type	Length	Rubric	Pre-Conference	Post-Conference Type
First	Announced	Lesson + 15 min	Instruction Rubric and Planning Rubric	Yes	Formal
Second	Unannounced	Lesson + 15 min	Instruction Rubric and Planning Rubric	No	Formal

Each teacher is assessed based on the teaching skills, knowledge, and professionalism performance standards that have been divided into four domains: instruction, environment, planning, and professionalism. Evaluators are required to note a teacher's ability to address the performance indicators described under each domain. Only descriptions of performance are

given at the scoring levels of 1, 3, and 5, but evaluators are able to choose to assign scores of 2 or 4 based on their determinations (NIET website, 2012).

- **Domain 1: Instruction.** Under the domain of instruction, evaluators will be looking specifically for the following twelve performance indicators: standards and objectives, motivating students, presenting instructional content, lesson structure and pacing, activities and materials, questioning, academic feedback, grouping students, teacher content knowledge, teacher knowledge of students, thinking, and problem solving. Based on performance in these areas, teachers are scored significantly above expectations (5), at expectations (3), or significantly below expectations (1), unless an evaluator chooses to administer a score of (2) or (4) (NIET website, 2012).

Standards and objectives. Standards and objectives are considered the foundation of all the performance indicators because of the belief that unless a teacher is "clear about what he or she wants students to know and be able to do as a result of the lesson, the balance of the lesson cannot be properly developed or implemented" (TEAM Handbook, p. 21). Inherent in this notion is the idea that every teacher and every student will understand the objectives for every lesson. To receive exemplary status in this area, teachers have to ensure that the standard is communicated to students in an understandable manner in both written and oral formats multiple times throughout the course of the lesson. The objectives must be clear, measurable, and observable. Sub-objectives are required and they are used to reflect on previous learning, instruct students on a new sub skill, or teach a process that is supportive of the primary standard (NIET website, 2012).

The rubric states that the objectives should always work to help students make connections. These connections might be to another content area that they have previously been instructed on or to an experience from the students' lives. The objectives should always be connected to state standards and relay the message of challenging expectations for all students. In addition, all students should be highly involved with the objectives on a daily basis. In other words, they must be able to locate the objective, read it, and be able to reference it. This might involve talking about the standard with a classmate or recording it at the top of their paper (NIET website, 2012).

Lastly, to ensure exemplary performance on this indicator, teachers must ensure that the majority of students are able to demonstrate mastery of the objective. Verbal or written assessments should be employed to determine if this mastery has been achieved (NIET website, 2012).

Motivating students. The TAP rubric examines each individual teacher's ability to motivate his or her students. According to this rubric exemplary teachers ensure that instruction is pertinent to students on a personal level, encourage inquisitiveness, and acknowledge accomplishments. Thus, evaluators should be looking for lessons in which students connect material to their lives in a relevant manner. In addition, they should anticipate teacher instruction that initiates student-led learning and questioning. Teachers with high scores in the area of motivation should also be skilled in bolstering student effort. This might involve encouraging students to take chances and providing verbal approval or other feedback (NIET website, 2012).

Presenting instructional content. This particular area of the rubric focuses on the manner in which content is relayed. The indicator looks first at an educator's ability to incorporate appropriate visuals in their lesson to facilitate student learning. These visuals should "assist the

learner in making connections with prior learning and in clarifying newly acquired concepts" (TEAM Handbook, p. 29). The teacher should also support learners in the organization of new information. This might involve providing the following reinforcements: examples, illustrations, analogies, and labels (NIET website, 2012).

Additionally, the instructional content indicator requires that teachers have the ability to model appropriately their expectations for student performance. To do so they must have a firm grasp on the presumed outcome as well as be able to provide appropriate steps and examples for students to be able to accomplish that outcome (NIET website, 2012).

Lastly, the instructor must be careful not to include any information in the lesson that is not applicable or relevant. According to the TAP rubric, this can be achieved by returning repeatedly to the lesson's objective (NIET website, 2012).

Lesson structure and pacing. This indicator addresses a teacher's ability to appropriately deal with time and format in regards to instruction. The lesson should follow a coherent progression in which events unfold in a logical manner. It should begin promptly and flow naturally from one activity to the next. Transitions should be expedient, with no lost instructional time. Teachers must incorporate a careful balance between a brisk pace and accounting for the needs of all learners. In other words, it must be acknowledged that students learn at varying rates and may need to be provided with differing amounts of time to complete the required assignment. No student should experience "down time" while waiting on others to finish a task (NIET website, 2012).

Activities and materials. This element of the TAP rubric takes into account a teacher's ability to address the specific needs and interests of his or her students when selecting activities and materials. Commendable activities and materials will address the objective of the lesson as

well as encourage higher-order thinking and reflection. Students should be able to make connections to their own lives and find them relevant and worthwhile. In addition, evaluators will be looking for teachers who are able to successfully incorporate multimedia and technology and to access materials found outside the realm of the school curriculum. The TEAM handbook notes that these requirements may be better facilitated by providing students with several choices to meet the objective (NIET website, 2012).

Questioning. According to the evaluation handbook, "questioning is an art form that reveals a great deal about a teacher's effectiveness" (TEAM Handbook, p. 38). Thus, in order to receive an exemplary rating on this component, teachers must demonstrate the ability to ask assorted and appropriate questions that address a mixture of: knowledge and comprehension, application and analysis, and creation and evaluation. Questions should be deliberate and align directly with the instructional focus of the lesson, and teachers should employ an appropriate amount of wait time. TAP suggests utilizing Bloom's Taxonomy as a guide. In addition, teachers should assist students in the formation of their own questions in an attempt to foster inquiry (NIET website, 2012).

Academic feedback. This indicator deals with the importance of proper feedback to support student achievement. Evaluators will be looking for both oral and written feedback that is consistent and appropriately targeted. Specifically, evaluators will be utilizing a checklist regarding the following characteristics in regard to feedback: links to the objective; spurs student thinking; and is specific, timely, and varied. The information teachers glean from student feedback should be employed to change instruction when necessary (NIET website, 2012).

Grouping students. This component of the rubric addresses how students are instructionally grouped during a particular lesson and how each student will be held responsible

for the completion of assigned work. Regardless of the kind of grouping used, methods of grouping should further student learning. When in groups, students should be clear about their obligations to themselves and to their group members. Lastly, groups should provide circumstances in which students may engage in reflection and goal setting (NIET website, 2012).

Teacher content knowledge. This indicator focuses on whether or not a teacher has a firm grasp on the content and material of the subject he or she is teaching. Evaluators will be looking primarily for teachers who can demonstrate superior content knowledge as well as the ability to help their students begin to comprehend this knowledge. Teachers should help their students connect new content to prior learning and to take the time to cover new concepts with appropriate depth (NIET website, 2012).

Teacher knowledge of students. In this component of the TAP rubric, teachers are required to adequately display knowledge of their students' cognitive abilities as well as interests. This is directly related to how well a teacher knows his or her students and the teacher's ability to motivate them. Teacher knowledge can be enhanced by anticipating problems students may have with a particular activity, taking differing interests and cultures into account when planning, and providing differentiated instruction (NIET website, 2012).

Thinking. Teachers who are rated exemplary in this category will apply four categories of thinking on a regular basis. The first type is analytical thinking in which "students analyze, compare and contrast, evaluate, and explain information" (TEAM Handbook, p. 49). In practical thinking, the second type, teachers will assist students in connecting their classroom learning to real world experiences. Thirdly, skilled teachers will foster creative thinking. In applying creative thinking, students will use their imaginations and develop dexterity. The final addressed type of thinking is research-based thinking in which students are taught to locate new knowledge

via research and make thoughtful decisions about the validity and usefulness of that research (NIET website, 2012).

In addition to modeling and fostering the four types of thinking, outstanding teachers will provide contexts in which students can develop new thoughts and develop different possibilities. Evaluators will also be looking to teachers to encourage the application of a variety of perspectives when solving problems. Lastly, students should be active in reflecting on their own thinking and scrutinizing it when appropriate (NIET website, 2012).

Problem solving. According to the TAP rubric, "developing multiple skills in problem solving enriches the learner's ability to manage complex tasks and higher levels of learning" (TEAM Handbook, p. 54). There are nine descriptors under the subcategory of problem solving that are thought to reinforce the skill of problem solving: abstraction, categorization, drawing conclusions and justifying solutions, predicting outcomes, observing and experimenting, improving solutions, identifying relevant and irrelevant information, generating ideas, and creating and designing. Abstraction involves excluding one or more properties in order to focus on the others. In the area of categorization, students will learn to look closely at information before going through the processes of classifying and sorting. Descriptor three, drawing conclusions, is based on three levels of conclusions in which students form determinations through means such as objectivity. When applying predicting outcomes, students will not only make predictions but also test the validity of those predictions. In observing and experimenting, students will apply the steps of the scientific method. For descriptor six, improving solutions, students will brainstorm possible improvements or modifications for a particular solution. In descriptor seven, students practice identifying the differences between important information and irrelevant information. Generating ideas, descriptor eight, involves helping students search for

analogies by working through possible solutions. Finally, when employing descriptor nine, creating and designing, students will develop an idea in the form of a problem or experiment for another student to resolve (NIET website, 2012).

Domain 2: Planning. This domain addresses the following three indicators: instructional plans, student work, and assessment. The suggestion in terms of instructional plans is that administrators evaluate their teachers' lesson plans on a regular basis and offer feedback. Specifically, lesson plans should align to state standards, build on prior knowledge, and provide evidence of differentiation. Student work should be purposeful and useful and ultimately enhance learning. Pre- and post-tests should be incorporated regularly. Assessment should also align with the content standards. It should also be able to be clearly measured. Teachers should provide written evidence within their plans that documents how assessment advances forthcoming lessons (NIET website, 2012).

Domain 3: Environment. In one of the conducted observations evaluators will be focusing the teacher's classroom environment. A teacher who performs well in this area will excel at managing student behavior. In these classrooms, students are routinely on task. Disruptions are infrequent and are dealt with promptly. The environment will also be organized and welcoming with student work on display. The culture of the classroom will be one of respect and caring in which all individuals, both teachers and students, demonstrate regard for differences (NIET website, 2012).

Domain 4: Professionalism. Each teacher will be evaluated the last six weeks of the school year using the Professionalism Rating Report. The rubric addresses the following ten indicators: professional development opportunities, implementing new strategies, personal learning plan, self-assessment, improving performance, using student achievement data,

supporting school activities and events, adherence to personnel policies, safe and orderly learning environment, and contribution to school community. Each teacher should provide evidence of each indicator. After conferencing with the teacher, the evaluator will average the score on the report into the qualitative component of the evaluation (NIET website, 2012).

Relevant research. The Milken Family Foundation created the TAP model in the 1990s in an attempt to improve the quality of classroom teachers. The theory behind the model is that a combination of tracking performance and incentivizing success can improve student achievement as well as teacher attrition. The model has increased in popularity with the implementation of the Department of Education's Teacher Incentive Fund, which provides grants to some districts that implement performance-based compensation (Glazerman & Seifullah, 2010).

Unfortunately, as of yet research that addresses the effectiveness of TAP is limited. The bulk of the current research has been conducted by the National Institute for Excellence in Teaching (NIET), the agency that oversees TAP (Glazerman & Seifullah, 2010). NIET's most recent study examined evaluation data from the 2006-2007 and 2007-2008 school years. They looked at a sample of value-added scores from 1,830 teacher records and analyzed evaluation scores and retention rates from a sample of 7,377 teacher records from 2004-2009. Based on the results of the study, NIET has claimed the following findings: (1) TAP evaluations support differentiated feedback for teachers. NIET argued that while most evaluation systems consistently score teachers at the highest scoring levels, teachers in TAP systems have a variety of scores and averaged a 3.5 out of 5. (2) There is a strong relationship between teacher evaluations and value-added student achievement scores. In other words, those who participate in the TAP model show an increase in student achievement scores over time. Researchers noted that this increase was in spite of prior achievement results or socio-economic status. (3) TAP

teachers become more capable the longer they are a part of the model. This finding was based on the instructional scores of 650 teachers over a period of two years. (4) Schools employing the TAP model are better able to retain quality teachers and have a higher turnover rate of poorer quality teachers. According to NIET, with each one-point increase in score in one year, a teacher's probability of remaining at a particular school increases by 87%. On the other hand, the inverse is true for teachers with lower ratings (NIET website, 2012; Daley & Kim, 2010).

Glazerman and Seifullah (2010) questioned these results, arguing that in NIET's studies "the comparison schools were chosen as a convenience sample and may not be representative of the outcomes that would have been realized in the TAP schools had the TAP schools not adopted the program" (p. 1). They conducted their own study on the TAP model over a two-year period in the Chicago Public Schools. They found no evidence that the model improved student achievement. There was no difference between TAP and non-TAP schools in terms of math and reading scores on the Illinois Standards Achievement Test. In addition, they found no indication that retention rates for teachers were higher in TAP schools in comparison to non-TAP schools (Glazerman & Seifullah, 2010).

An additional study on the TAP model was conducted by Schacter and Thum in 2005. They studied five elementary schools in Arizona that had recently implemented the program and evaluated TAP based on the following three measures: student achievement data, teacher satisfaction and attitudinal data, and implementation data. They found that participating schools had significant gains in reading, math, and language achievement in comparison to non-participating schools. Also, 74% of teachers in TAP schools reported positive attitudes and job satisfaction (Schacter & Thum, 2005).

Despite positive results, Schacter and Thum (2005) acknowledged a few mitigating factors that may impact the success of the model. For example, they noted that in the most successful schools TAP was teacher approved, meaning there had been strong teacher buy-in before implementation of the model. Furthermore, teacher surveys were administered before teachers received their performance-based compensation. Data showed that while teachers supported professional development and collegiality there was much less support in terms of compensation and accountability (Schater & Thum, 2005). In general, it seems that school reform engendered more success when administrative leadership was strong and teachers showed support for the program (Berends, Chun, Schuyler, Stockly, & Briggs, 2002).

Ultimately, it seems that further research is needed to corroborate the effectiveness of the TAP model, especially from individuals who are not closely associated with its creation. Likewise, research will be necessary to evaluate the success of Tennessee's Teacher Educator Acceleration Model as it is in its first year of implementation.

Section IV: Critical Research Paradigm

Many individuals in the educational, political, and public arenas see schools as ideal environments for reform and teachers as vital agents in the process of change (Schmidt, Armstrong, & Everett, 2007). Morgan (1997) cautioned against complacency and routine, or, what she refers to as "the rituals of catechism" (p. 65), which she claims have been relied upon by conventional teachers for decades. Likewise, Harste (2000) reminded educators to be wary of our own absoluteness, of being sure that we are right, because this absoluteness can lead to a feeling of justification. Suddenly, educators feel justified in mandated implementations of a certain reading program or testing series. Thus, Harste (2000) suggested pursuing new avenues with caution and always with the assumption that our own knowledge may be flawed in some

way. This may be the only means to ensure that personal experiences don't become a barricade to novel pathways. Critical literacy may serve as an avenue for exploring new knowledge both for students and teachers through its call to "disrupt the commonplace" by forcing each to take a second look at what was perceived as an absolute truth (Lewison, Flint, & Van Sluys, 2002, p. 382).

With this view as a backdrop, I will use this section to do the following: (1) examine the tenet of the critical research paradigm including perspectives, role of the researcher and participants, and contentious issues; (2) look closely at what research says about issues of power; and (3) note critical paradigm's connection to education and trace examples of literacy studies that have applied the critical methodology.

Critical research paradigm. From the view of the critical researcher, the world in which all reside is one heavily laden with social structures that are both historically-based and always currently relevant because they constantly impact the lives and choices of everyone in society (Hatch, 2002). The notion is that these structures work to create a hierarchy that positions particular individuals over others, which results in distinctions in regard to how people or groups are treated based on race, gender, and social class. Epistemologically speaking, researchers who subscribe to this paradigm believe knowledge cannot be separated from the political; thus, the initial step in the process is to acknowledge it. Hatch (2002) noted that these historically-situated structures "are in the absence of insight as limiting and confining as if they were real" (p. 16). From this stance the question is not: is it political? It *is* political, and the question is: are you going to acknowledge it or ignore it? As Fine (1994) argued, "all research projects are (and should be) political … researchers who represent themselves as detached only camouflage their deepest, most privileged interests" (p. 15).

While some critical researchers have begun to shy away from the "strict economic determinism of Marxism" (Rogers, Malancharuvil-Berkes, Mosley, Hui, & Joseph, 2005, p. 367), for most, the conviction remains that the role should be one in which the researcher recognizes the power of partiality and domination in our world. Therefore, for many critical theorists the goal is not only to reveal structures of oppression that are exacerbated by those in power but to raise awareness and, in some instances, even transform the dominated community (Glesne, 2011). This may mean not simply delineating *what is* through observation and documentation but imagining *what could be*.

Carspecken (1996) maintained that critical research requires the following: 1) research for cultural and social criticism, 2) opposition to all manner of injustice, 3) research employed to divulge oppression and in turn transform it, and 4) acknowledgement that prevailing research contributes to inequality. Ultimately, researchers must develop a comprehension of the interplay between fact and value and then work against unfairness *via* disclosure and critique. The researched in the process must be willing to be participatory and to engage in dialogue about social change. It may also be important that they view the researcher as on their side or an ally to their cause.

Grbich (2007) noted that several scholars have raised concerns about the implications of the critical paradigm, with particular focus on the question: Who has the power to emancipate whom? (Grbich, 2007). Some viewed the researcher in a positivist or interpretivist role. According to Noblit and Eaker (1999), "the evalutator-critical theorist is in many ways a supreme patron, providing both content and process direction in the service of free uncoerced dialogue" (p. 68). Thus, it is essential that he or she engage in constant reflexivity noting that even as he or she seeks to reveal instances of subjection he or she may be engaging in the same

processes of power (Carspecken, 1996). In addition, it has been noted that critical research may often not reach the very audience it is intended to transform (Rogers et al., 2005). If the knowledge produced is simply passed from one tower of academia to another, then the underlying purpose of transformation may be lost.

Issues of power. Critical research focuses on issues of power, particularly the need to shed light on these issues in order to incur change. Rogers et al. (2005) noted that power relations are often historically-based and that the outcome is the privileging of certain groups over others. This could lead to greater access for some meaning that dominant groups have a greater abundance of social resources, including assets, social rank or station, education, and information (Van Dijk, 1993). According to Van Dijk (1993), power may also be relevant to the connection between action and cognition. For example, "a powerful group may limit the freedom of action of others but also influence their minds" in a manner that has benefits solely for the dominant individual or group (p. 254).

Van Dijk (1993) also discussed the institutionalized nature of power. As a result, power may be vindicated and rationalized by government rules, the media, or literature. Dominant groups often have a hand in the blueprint and execution of power structures. However, it would be unwise to view power in simplistic, black and white forms. Power is rarely found in totality; rather, it

> may be restricted to specific domains, and it may be contested by various modes of challenge, that is, counter power. It may be more or less consciously or explicitly exercised or experienced. Many more or less subtle forms of dominance seem to be so persistent that they seem natural until they begin to be challenged. (Van Dijk, 1993, p. 255)

In this manner, power may be jointly produced by both the dominant and the dominated groups through elaborate social interactions. Therefore, researchers are encouraged to dispose of the standard "good guy" versus "bad guy" dichotomy and acknowledge the more layered and complicated nature of power.

The role of critical research in education. The fact that schooling is both public and institutionalized seems to make it a fertile ground for critical research. In fact, the number of critical studies in education has increased greatly in recent years despite some claims that teachers may see the theory as inconsequential (Johnston, 2004). Everhart (2004) argued that educators are obligated to improve the circumstances of the public institution of education, and he viewed critical theory as a springboard for this improvement. He argued,

> when we accept the premise that the fundamental purpose of education in a free society is to foster a citizenry capable of active and informed participation then it follows that inquiry within and about education should contribute to that objective. (Everhart, 2004, p. 296)

From this stance, for those in the realms of education to engage in the critical is not only advisable but a responsibility that must be taken up.

Examples of critical literacy research studies. In this section I have supplied the reader with a brief introduction to the ontological and epistemological aspects of critical theory. Next, I more closely examine the methodological facets of the theory by examining three relevant literacy studies that take a critical lens.

Rogers (2002) followed one minority family that resided in a section of New York known for high levels of violence and poverty for a period of two years. Through multiple data collection methods, including interview, document collection, photography, and archival

research, she sought to establish through a critical lens how a literacy identity is formed and how the institutionalized and routinized nature of schooling may work to make life-changing decisions about a child's future (Mehan, 1996). Specifically, she traced how one mother consented to special education services for her child despite initial reservations. Rogers (2002) leaned heavily on the paradigm of the critical to make sense of her data. For example, she analyzed critical incidents in the data that are defined as tensions or surprises and relied on critical discourse analysis to make sense of the social and political aspects of the study (Fairclough, 1995). By attending to the local, the institutional, and the societal elements of discourse (Fairclough, 1995), she worked to "illuminate how people make sense of their reality and understand their social positions" (Rogers, 2002, p. 257).

The results of the study demonstrated that June (the mother of the child being evaluated for a special education reading diagnosis) showed a strong literacy identity in her home. For example, she encouraged her children to complete their homework before playing outside, displayed their certificates of achievement proudly on the wall, and kept a bookshelf stocked with books for her children to access. In her kitchen she spoke adamantly of her right as a mother to keep her child out of special education when she stated, "if they argue with me, they argue with me. She will not be in no special education class. I'm the one who has to sign the papers. I'm her mother, not them" (Rogers, 2002, p. 261). However, when attending the special education meeting June vocalized none of these views as the committee reviewed the test results that highlighted her daughter's deficiencies, and she signed the permission form with little comment.

Based on these data, Rogers (2002) concluded that the language of the institution, which is presented as scientific and objective, worked to encourage a submissive stance from June and

played on her history as a student with literacy failure. In this way, the literacy of her family's daily life was discounted in comparison with the social institution of school. Through the paradigm of critical literacy, Rogers (2002) made transparent the power disparities that may exist between schools and the families they serve. She advocated for educators not to negate the critical reading that may occur outside of school as well as the idea that "critical literacy in the classroom offers great potential for reordering the invisibleness of invisible assumptions" (Rogers, 2002, p. 267).

Heath's (2004) study, "The Children of Trackton's Children: Spoken and Written Language in Social Change," provided another example of how one might use a critical lens to view data. She took a political stance and argued that current bureaucratic decisions may affect the welfare of particular cultural groups. For example, she referenced a ruling from the assistance program Aid to Families with Dependent Children (AFDC) that implicated absenteeism on the part of fathers as well as public housing regulations together may undercut the formation of a supportive community.

Heath (2004) followed two individuals, Zinnia Mae and Sissy, both of whom were raised in Trackton, a minority community in the 1960s and 1970s. Years of transcribed audiotapes revealed that for Zinnia Mae the physical move away from Trackton into an urban housing project resulted in an emotional and psychological move away from a culture rich in dialogue, interaction, and community to one of isolation and negativity. Over the course of the study, Zinnia Mae remained on welfare, out of work, and in the same housing project. She most frequently described herself as "tired." While she attended to the external needs of her children, only 14% of talk in the home was between Zinnia Mae and her children, and they experienced very little access to individuals outside of the home. In contrast, in Sissy's home approximately

40% of the audio recorded data revealed conversations amongst three or more individuals regarding work or leisure.

Heath (2004) took a critical stance and argued that "it is essential that we see relationships of knowledge and power that both connect those within communities and disconnect them [especially] in an era in which national and state policies strike at the core of intimate family relations" (p. 207). In this way, Heath's work was both a political statement and a call to action.

The last article I highlight in this section focused on how the language of teachers may influence the manner in which students take up (or do not) their own literacy identity in ways that may (or may not) be transformative. In their theoretical framework the authors, Hall, Johnson, Juzwik, Wortham, and Mosley (2010) relied on the critical as they positioned literacy in terms of power. They noted that, based on Scribner's (1984) work, "language and literacy can be used to help empower communities or groups, or access to literacy can be diminished in order to prevent particular groups from acquiring power to create change" (Hall, Johnson, Juzwik, Wortham, & Mosley, 2010, p. 235). Data collection of teacher and student language in three different classrooms led to the following conclusions: (1) the language teachers use in the classroom may influence which literacy identities students feel they do or do not have access to, (2) this language encourages particular literacy identities while discouraging others both in and out of the classroom, and, (3) the student and teacher literacy identity construction is not arbitrary but deliberate. The researchers held that on the basis of these findings students were pressured to take on the prevailing forms of identity or risk marginalization (Hall et al., 2010). Hall et al. (2010) promoted the notion of critical change by suggesting that educators should

move away from this dominant culture of schooling toward an environment that allows for freedom and acceptance of various literacy identities.

These three literacy studies have demonstrated how one might take up the critical paradigm in an educational research study. While I initially hesitated to declare a critical agenda from the outset of this study, I have come to agree with Fine (1994) that lack of acknowledgement does not change one's purpose. Therefore, I examine my data through the lens of the critical not to impose a Marxist or emancipatory agenda upon my participants but rather in hopes of shedding some light on previously under-considered issues of power structures and the impacts they may have upon teachers, students, and the educational system.

Summary of Chapter

In this chapter, I discussed the long and changing history of teacher evaluation and its impact on political and educational reform. I then traced various studies of top-down policy implementations in school settings and the implications of these enactments. I also closely examined the specific components of Tennessee's TEAM evaluation model that has been newly implemented. Lastly, I highlighted the theoretical underpinnings of the critical research paradigm as well as three examples of literacy studies that employ this paradigm. In the following chapter I describe the methodology for this study.

Chapter 3. Methodology

Chapter Introduction

In this chapter, I provide the methodology of the study. I consider the role of the researcher, the population and setting, as well as data collection and analysis procedures.

The purpose of this study was to describe how the enactment of a particular policy (in this case the implementation of a new teacher evaluation system based on 50% qualitative and 50% quantitative, which takes into account teacher effect and student achievement, data) impacted the experiences of eight first grade teachers in the Tennessee public schools.

Methodological Approach

This study is qualitative, employing ethnographic research procedures (Hatch, 2002). Aligning myself with Hatch's (2002) notion of qualitative research, in this study I work to understand the perspectives of my participants by attempting to answer the following questions: "What is happening here, specifically? What do these happenings mean to the people engaged in them?" (Erickson, 1986, p. 124). In keeping with ethnographic procedure, I attempt to describe the perspective of the individuals of a particular culture, in this case teachers, and to understand their daily experiences through the examination of observations, interviews, and artifact collection (Hatch, 2002).

As detailed in chapter one, I position myself within the constructivist paradigm in that I assume that absolutes are unknowable in our world and multiple realities are always in existence. From this stance, individuals are constantly constructing their own view of the world based on how they internalize it in their own unique ways. Additionally, I must point out that this particular research project will be viewed through the lens of the critical paradigm. While I did not enter into data collection with a specific social agenda, I must acknowledge that throughout

collection and analysis I remained attuned to structures of power and the possibility that raising awareness of such structures might lead to change or transformation (Hatch, 2002).

Participant Selection

My selection of participants was purposive, relying heavily on the work of Merriam (1998), which presumes that the role of the investigator is to "discover, understand, and gain insight and therefore [he or she] must select a sample from which the most can be learned" (p. 61). In this way, my intent was to select individuals who possessed basic similarities. For example, all participants were first grade teachers in the public schools of Tennessee who were at the time experiencing the implementation of the new evaluation model. While these broad criteria were necessary prerequisites for all participating teachers, within those parameters I sought individuals who would likely bring different perspectives or points of contrast to bear on my research. Based on these criteria, I selected three different school districts to study within the state of Tennessee with the idea that the implementation of policy may look different depending on the actions of district leaders. Therefore, an examination of a single district would not allow for productive comparisons. The districts differed in geographic location, size, and student makeup. Amongst the three districts I chose teachers from four different schools. Again, my choices were purposeful. I selected schools that, at least on the surface, appeared to be quite or somewhat different in terms of characteristics or population. For example, within one particular district I chose a Title I school with statistics that demonstrated a population of students with high minority rates and a large number of students receiving free and reduced lunch. Within the same district, I selected another school that was quite statistically dissimilar. It was reported to be in an affluent neighbor and its student population contained very low numbers of minority

students as well as students in poverty. Again, my hope was to provide contrasting environments in order to better glean the different perspectives of teachers in varying situations.

Additionally, I opted to study only first grade teachers. My reasons for doing so were multiple. First, I believed that it would be easier to make connections and comparisons amongst the teachers in the study if they were all instructing the same grade level. Because the TEAM model would be employing quantitative scoring methods, I believed that this might look quite different at the third grade level, for example, than at first grade. Secondly, because of the arbitrary nature of student scores in the primary grades and inclusion of these scores in the evaluation, I felt that it would be timely to look more closely at how this is playing out in primary grade classrooms. Additionally, first grade is arguably the most important grade for teaching children to read and my focus on TEAM has a literacy slant. Lastly, I was a first grade teacher for seven years, so my interest is a personal one. I feel highly connected to that particular student age group.

Access and Entry Procedures

In this section, I identify how participants were recruited for this study. The primary criterion for participation in the study was the implementation of TEAM within each district. Because TEAM was mandated by the state of Tennessee, all public school districts were required to begin the process of implementation. Therefore, any district in Tennessee would meet the primary expectation for participation. However, as outlined in the previous section, I attempted to select three districts that might provide an arena for comparison. After selecting three potential school districts, I perused each district website in an attempt to better understand access procedures in each. Each district had very different access procedures and as a result my contacts were formed in varying manners.

For each district, I first had to secure district permission to conduct research. In Booker County, I began by setting up an appointment with the Director of Instruction to discuss my research. I provided her with a hard copy of my research plan, and we discussed the details, including the purpose and primary features of the study, at length during the meeting. She provided me with the necessary forms to request permission, which would then be signed by participating principals as well as the superintendent. District access required similar forms in both Hampton County and Carson County, although no meeting was required with central office leaders. Rather, I submitted the required research proposal forms to the Office of Research in Carson County and the assistant superintendent in Hampton County. Both Carson and Booker districts also required principal approval before conducting research at a school site. I set up meetings with four principals from these districts. While all principals responded positively to my research, I did not conduct research at one of the schools in Booker County because of lack of teacher interest. Principal signatures were not required in Hampton County. After securing permission at the district and school levels, I then had to locate first grade teachers within these schools willing to participate.

The principal at Whiteside Elementary set up a meeting with his first grade teachers where he explained the purpose of my research and asked if any teachers were willing to participate. After receiving all teachers' verbal agreement, he set up dates and times for me to conduct interviews and classroom observations. The principal at Newton went through a similar process with her teachers: after receiving the teachers' verbal permission she sent me their contact information. From there, I communicated directly with the teachers *via* email to schedule interviews and classroom observations.

The principal at Lindley Elementary set up a meeting for me with his team of first grade teachers. During this meeting I provided them with information about my study and determined their willingness to participate. Two of the four teachers volunteered to take part. From that point I communicated with each of them individually *via* email to arrange interviews and classroom observations.

I communicated directly from the outset with the teachers at Gary Elementary to gauge their level of interest in participation. One teacher from Gary expressed interest; the other explained that she felt her knowledge of the TEAM model was too limited to engage in the interview and observation processes.

After receiving approval at each level (district, school, and classroom), I submitted all appropriate documents to the Institutional Review Board for approval before beginning data collection.

School and Participant Descriptions

This study took place in three school districts in the state of Tennessee. Each district was in the second year of implementation of the TEAM evaluation model. In this section, I will briefly describe each district, school, and participating teacher. All district, school, teacher, and principal names in these descriptions and in the analysis section in chapter four are pseudonyms. The following chart illustrates selected districts and characteristics, selected schools and characteristics, and selected teachers and characteristics.

Table 3: Selected Districts, Schools, and Teachers and Corresponding Characteristics

District	District Characteristics	School	School Characteristics	Teacher	Teacher Characteristics V.E.L. (View of Effective Literacy Instruction)
Carson County	Large Suburban/urban Neighborhood schools 87 schools 3300 teachers 56,000 students 80% white 15% African-American 3% Hispanic	Whiteside Elementary	Arts-integrated 300 students 79% white 74% economically disadvantaged 17% African-American 6% Hispanic TCAP Reading Grade B	Laura	4 years experience in first grade TEAM Lead Teacher V.E.L.: holistic teaching, clear expectations, cross-curricular purposs
				Karen	0 years experience Internships in 2nd and 5th grades V.E.L.: time for students to read, access to a variety of reading materials, modeling
				Allison	4 years experience in 1st grade Master's Degree V.E.L.: small group time, dialogue about books
Hampton County	Small, rural 96% Title 1 16 schools 500 teachers 7500 students 96% white	Gary Elementary	Very rural area 350 students 95% white 53% economically disadvantaged TCAP Reading grade C	Sally	12 years experience in first grade Pursuing reading specialist cert. V.E.L.: differentiated instruction, word work, deeper comprehension

Table 3. Continued.

District	District Characteristics	School	School Characteristics	Teacher	Teacher Characteristics V.E.L. (View of Effective Literacy Instruction)
Booker County	Small, urban 13 schools 500 teachers 7000 students 79% white African 11% Hispanic 7%	Newton Elementary	Urban Focus on science, math, and technology 300 students 90% economically disadvantaged 50% white 35% African-American 11% Hispanic TCAP Reading Grade B	Rebecca	2 years experience public schools, experience in private schools, worked in all elementary grades except 2nd V.E.L.: focus on phonemic awareness, modeling, student response
				Joanne	12 years experience in 3rd, 1st, 4th grades V.E.L.: word study, small groups, literacy centers
		Lindley Elementary	Suburban, affluent 500 students 20% economically disadvantaged 86% White	Kendra	4 years experience in 1st grade V.E.L.: small groups, balanced literacy, phonics
				Jessica	4 years experience in 1st and 3rd grades, taught at military school V.E.L.: differentiated instruction, word work, phonemic awareness

District 1: Carson County. This district is a very large district comprised of both suburban and urban neighborhoods. It comprises a fairly large geographic area that includes affluent neighborhoods as well as low-income housing communities. Most of the students attend schools in their neighborhood, and for this reason schools often reflect the population of the surrounding community. The district has 87 schools total, including 50 elementary schools, 14 middle schools, 13 high schools, and 10 special schools. Carson County employs over 8,000 individuals, half of whom are teachers. Nearly 80 percent of the district's 56,000 students are white, 15 percent are African-American, and 3 percent are Hispanic.

Whiteside Elementary School. Whiteside Elementary is an arts-integrated school located in an urban neighborhood in the Carson County school district. Approximately 300 students attend Whiteside. In comparison to other schools in the district, Whiteside is economically and culturally diverse. 79 percent of the school population is white; African American and Hispanic students comprise most of the minority population. 74 percent of the students are considered economically disadvantaged, and 69 percent receive subsidized lunches. The school received a grade of B in reading based on a three-year average of the TCAP Criterion Referenced Academic Achievement. In terms of value added scoring, Whiteside achieved a mean gain of 1.2 in reading for the 2011-2012 school year.

Teacher 1 from Whiteside Elementary: Laura. Laura was in her fourth year as a first grade teacher at Whiteside Elementary. She interned in kindergarten and second grade classrooms, but her only experience as a full-time teacher has been in first grade. She was serving as a lead teacher for her school, meaning that her principal asked her to assist in conducting classroom evaluations as required by the TEAM model. She described her students as very diverse, with a large influx of them living in government projects at the poverty level. Laura

explained that she viewed effective literacy instruction as: giving students very clear expectations in reading, helping them make holistic connections between literacy and their real life, and making cross-curricular purposes apparent.

Teacher 2 from Whiteside Elementary: Karen. Karen was a first-year first grade teacher at Whiteside. She also participated in internships in second and fifth grade classrooms. She described her students as being much more diverse and coming from a rougher home life than students she instructed during her internships. She noted that many of her students came from divorced and single-parent homes. Karen described effective literacy instruction as giving students ample time to read as well as access to a variety of reading materials. She also noted that modeling different reading and thinking strategies is essential to student success.

Teacher 3 from Whiteside Elementary: Allison. Allison is in her fourth year of teaching first grade at Whiteside. She entered a master's program immediately after completing her undergraduate work in 2008, finished her master's degree in 2009, and began teaching at Whiteside immediately afterwards. She explained that her students comprise a mixed demographic, with many of them being raised in single parent homes or by grandparents. Allison viewed small group time as the most important aspect of effective literacy instruction because it allowed her time to hear her students read and have conversations about books.

District 2: Hampton County. Hampton County is a small school district in Tennessee that encompasses a largely rural area. Very few of the communities in Hampton would be considered affluent and nearly 96 percent of the schools are Title I schools. The district serves approximately 7,500 students in pre-K through twelfth grade, 96 percent of who are white. It is comprised of 11 elementary schools, 1 middle school, and 4 high schools. Hampton employs over 500 certified personnel.

Gary Elementary School. Gary Elementary is one of the 11 elementary schools in Hampton County. It is located in a very rural area, and is surrounded by ample acreage and farms. Approximately 350 students attend Gary. Of these students 95 percent are white. The remaining 5 percent is split evenly between African-American, Asian or Pacific Islander, and Hispanic students. Fifty-three percent of students are considered economically disadvantaged. Based on a three-year average of the TCAP Criterion Referenced Academic Achievement, Gary received a C in reading. In terms of value-added, the mean gain for reading was -2 for 2012.

Teacher 1 from Gary Elementary: Sally. Sally is a native of Hampton County and has been teaching first grade at Gary Elementary for 12 years. It is the only grade level she has taught. Sally described her population as very rural and mostly white. She explained that the population has changed over the last few years as the community has dealt with an influx of drug issues. She noted that many grandparents are now rearing the students she teaches rather than the students' parents. Sally's view of effective literacy instruction included: teaching students at their specific reading level, focusing on building and sounding out words, and accessing deeper levels of comprehension.

District 3: Booker County. Booker County is a school district located in a small city. While the district size is similar to Hampton in terms of numbers, Booker County comprises a much smaller geographic area, and thus serves fewer rural students and more suburban or urban students. Within the district there are 8 elementary schools, 2 middle or intermediate schools, 1 high school, and 2 special schools. The district employs approximately 500 teachers. The student population includes over 7,000 ranging from pre-K through twelfth grade. 79 percent of these students are white. African-Americans and Hispanics make up most of the minority population.

48 percent of students are considered economically disadvantaged, and 36 percent of the schools receive Title I funding.

Newton Elementary School. Newton is located in an urban neighborhood of Booker County and is the most diverse of the schools involved in the study. It is a small neighborhood school with fewer than 300 students. Due to outside funding, it has a strong focus on math, science, and technology. The minority population is large; 35 percent of students are African American and 11 percent are Hispanic. 50 percent of students are white. Over 90 percent of the school's student population is considered economically disadvantaged. In 2012, Newton received a grade of B in reading for the 3-year average of TCAP Criterion Referenced Achievement and a mean gain of -1.7 in reading for value added.

Teacher 1 from Newton Elementary: Rebecca. Rebecca began her teaching career in private schools, teaching first in a school in the "deep South" and then in a private school closer to her current home. She then completed her master's degree in education, and as a participant in that program, was required to do a yearlong internship in a nearby school system. After the internship Rebecca was hired as an interim teacher for the district and assisted in all grade levels, from kindergarten through fifth grade, with the exception of second grade. She then applied for and was hired for her current position as a first grade teacher at Newton. It is her first year working there. Rebecca described her students as ranking very high on the needs list and having needs that reach beyond just the educational and academic realms. She noted that parental involvement is very low at Newton. Her view of effective literacy instruction included: a heavy focus on phonemic awareness, plenty of modeling, and frequent student response.

Teacher 2 from Newton Elementary: Joanne. Joanne has been teaching elementary school for twelve years. She initially began her teaching career in a fourth grade classroom in a nearby

state. After two years, she moved to another city in that state and taught first grade for 3 years. She then moved to her current city of residence and began employment at Newton. Since working at Newton, Joanne taught for three years in first grade, three years in third grade, and two years in fourth grade. She is currently in the first grade classroom. Joanne explained that she didn't think much about whether or not her student population was different from surrounding schools. However, she did note that many of her students are non-readers and that they often receive little help at home because their parents are non-readers as well. Joanne's view of effective literacy instruction included: activities breaking apart and visualizing the sounds in words and other components of word study, small group time, and literacy centers.

Lindley Elementary School. Lindley Elementary is located in an affluent suburb of Booker County. In fact, the school is positioned right in the center of the neighborhood and is surrounded by large, upscale homes. Many of the students can walk just a few steps from their yard and be at school. Lindley serves approximately 500 students from kindergarten through fourth grade. Comparatively speaking, the school is not very diverse: 86 percent of students are white, 4 percent African-American, 5 percent Asian or Pacific Islander, and 5 percent Hispanic. 20 percent of the students are considered economically disadvantaged. Lindley received a grade of B for 2012 for TCAP Criterion Referenced Academic Achievement in reading and a mean gain of 3 in reading for value added.

Teacher 1 from Lindley Elementary: Kendra. After completing her undergraduate work in 2008, Kendra spent one year serving in various interim positions that included experiences in third and fourth grades. In between the interim positions she worked as a substitute teacher. Following that year she was hired at Lindley as a first grade teacher. She is currently in her fourth year in that position. Kendra described most of her students as coming from high socio-

economic backgrounds with high parental support. However, she noted that a small handful of students in each classroom are from high-poverty homes, which creates a large gap between the have and have-nots. Kendra's view of effective literacy instruction included: daily small group time in order to work with students at their level, a balanced approach which includes teaching sight words and phonics, and a structure in place which encourages students to engage in real academic work rather than busywork.

Teacher 2 from Lindley Elementary: Jessica. Jessica was in her fourth year as an elementary school teacher. She has taught both first and third grades. She first taught at a military school in a nearby state. Since then she has worked at Lindley as a first grade teacher. Jessica described her student population as generally affluent and primarily white. However, she noted that this year she has a larger number of lower income students than is typical. Jessica's view of effective literacy instruction included: the acknowledgement that different students learn in different ways and inclusion of word work, phonics, and phonemic awareness in daily instruction.

Data Collection Methods

In this section I will address the following: 1) the time line of data collection, 2) types of data that were collected, 3) rationale for data that was collected, and 4) rationale for how data was collected.

Timeline. Data collection began upon receipt of IRB approval in September 2012. I conducted interviews with each participating teacher. I also observed each participant instructing her students in some aspect of literacy. Lastly, during this time I also collected related documents from participants. Data collection ended mid-November 2012. The following chart illustrates data collected and timeframes for each participating teacher.

Table 4: Data Collected and Timeline

Teacher	Data Collected	When	Data Collected	When	Data Collected	When	Data Collected	When
Laura	Interview	October 2012	Observation	October 2012	TEAM composite score	October 2012	TEAM lesson plan, scores, evaluation commentary	October 2012
Karen	Interview	October 2012	Observation	October 2012	TEAM composite score unavailable	N/A	TEAM lesson plan, scores, evaluation commentary	January 2013
Allison	Interview	October 2012	Observation	October 2012	TEAM composite score	October 2012	TEAM lesson plan, scores, evaluation commentary	January 2013
Sally	Interview	September 2012	Observation	October 2012	TEAM composite score	September 2012	TEAM lesson plan, scores, commentary	October 2012
Rebecca	Interview	October 2012	Observation	October 2012	TEAM composite score	October 2012	TEAM lesson plan, scores (still waiting on post-conference)	March 2013
Joanne	Interview	October 2012	Observation	October 2012	TEAM composite score	October 2012	Did not receive TEAM evaluation due to medical issues	N/A
Kendra	Interview	October 2012	Observation	October 2012	TEAM composite score	October 2012	TEAM lesson plan, scores, commentary	October 2012, March 2013
Jessica	Interview	October 2012	Observation	October 2012	TEAM composite score	October 2012	TEAM lesson plan, scores, commentary	March 2013

Data source 1: Teacher interviews. Teacher perspective is a critical piece of this study, and I consider it to be pivotal in forming an understanding of classroom decision-making. Therefore, the primary source of data for this study was based on interviews with the

participating teachers. My rationale in employing participant interviews was to "uncover the meaning structures that participants use to organize their experiences and make sense of their worlds" (Hatch, 2002, p. 91). Hatch (2002) noted that these meaning structures may be not readily visible even to the participant. In this way, qualitative methods can be a means of uncovering perspectives that have previously not been explored. I align myself with the thoughts of Spradley (1979), who articulated who the purpose of ethnographic interviews thus:

> I want to understand the world from your point of view. I want to know what you know in the way you know it. I want to understand the meaning of your experience, to walk in your shoes, to feel things as you feel them, to explain things as you would explain them. (p. 34)

Just as I speak of my own hope to discover "what is on someone else's mind" (Patton, 1990, p. 278), I must also acknowledge that this is a difficult feat to undertake. Participants may be, for a variety of reasons, reluctant to divulge their true feelings, and thus we can never be sure that we are privy to someone's authentic thought processes.

I, as the researcher, conducted and led the interviews, which were scheduled for a particular place and time. Each interview lasted approximately one hour. Before beginning the interviews, I created a list of guiding questions. These questions were informed by Coburn's (2001) work in sense making and Spillane's (1999) thoughts on district mediation. My hope was that my questions would provide the space for my participants to share their individual experiences with the TEAM evaluation model and how they have made sense of it. More specifically, I wanted to touch on how the evaluation may or may have not impacted their literacy instruction and their view of the qualitative and quantitative aspects of the evaluation. The questions were open-ended, and the direction of the interviews was contingent on where the

participants led. When appropriate, I referred to my guiding questions, but ultimately the interviewee set the course of the conversation. In turn, at times I created on the spot questions relating to the nature of the conversation. This stance is in keeping with what Chiseri-Strater and Sunstein (1997) termed "flexible structure" and with my own constructivist stance. As Hatch (2002) noted, "constructivists will be more likely to emphasize flexibility as they seek to create interactive interviews where they share responsibility for questions and answers with informants" (p. 95). In addition, I have to acknowledge that because of my critical leanings the line of questioning at times took on a stance that lent itself to advocating social and political change.

All participants consented to having their interviews recorded. Upon sitting down for the interview, I placed the digital voice recorder between myself and the interviewee. I began recording at the onset of each conversation and stopped recording after we had addressed the final guiding question. Each recording was then saved to my computer as well as to the transcription program Insqribe in files corresponding to each participant. All files were deleted from the digital recorder after being transferred to my computer.

Interviews were transcribed in December 2012 and January 2013. Each was transcribed verbatim in order to provide for a hardier and more accurate analysis of data (Merriam, 1998). Each transcribed document was then saved under in corresponding participant folders on my computer and an external hard drive. I also created individual folders for each participant that contained places for each piece of data. I placed a hard copy of each interview in the corresponding participant folder.

Data source 2: Classroom observations. From the outset of my study, I saw interviews as the focal point for increasing my understanding of teacher perspectives on the TEAM model. I

believed that hearing the opinions and thoughts of my participants was the best way to better understand their views. Nonetheless, I believed that classroom observations would serve as a valuable secondary piece of data that might inform the interview process as well as provide a context for the commentary of the teachers I was studying. As Glesne (2011) noted, understanding the context of the participants' work is tantamount to understanding the perspective of the participants:

"Through being a part of a social setting, you learn firsthand how the actions of research participants correspond to their words; see patterns of behavior; experience the unexpected, as well as the expected" (p. 64).

With this notion as a foundation, I requested to observe each teacher during her literacy block. I simply asked to see some aspect of each teacher's reading instruction. I did not specify a particular area such as shared reading or word work because I wanted the teacher to make the decision about which aspect of reading instruction she felt was most important for me to witness. In some classrooms, I viewed single literacy lessons; in others I observed a literacy block that comprised several reading components in one time period. What I observed was dependent primarily on the manner in which each teacher constructed her literacy instruction.

My place on the participant-observation continuum (Glesne, 2011), which ranges from being primarily an observer in a social situation to primarily a participant, fell closer to that of an observer because I had little or no interaction with the teacher or students during the period of observation (Glesne, 2011). My role was passive (Spradley, 1980), and I hoped to avoid intrusiveness as much as possible. Nonetheless, I believe, as did Labov (1972), that intrusiveness can never be completely avoided. Rather, we as observers always influence the environment we

are observing: “the very presence of a researcher makes any natural context unnatural to some degree” (Glesne, 2011, p. 73).

I arrived at each classroom a few minutes prior to the time the teacher had given me for the start of the lesson in order to get settled before the lesson began. In each classroom I attempted to select a place to sit that would be the least obtrusive yet still allow me the ability to view all classroom happenings and interactions. In most classrooms, I sat at an empty student desk or small group table at the rear of the room. I brought with me a small notepad, pen, and my laptop. I checked with the teacher beforehand to make sure that she would not feel uncomfortable if I took notes on my computer. While I believe that handwritten notes are often less noticeable, I find that I am able include more detailed commentary when I type than when I write notes by hand. During each lesson I used my computer to take field notes that I believed would be essential in the process of analysis. As Sunstein and Chiseri-Strater (2002) pointed out, “the difference between doing fieldwork and just hanging out is the writing. Without writing, the sharp, incisive details about people, places, and cultures are lost to us” (p. 56). I attempted to create notes that were descriptive, and, as Glesne (2011) suggested, I did my best to avoid judgments in my note taking. For each lesson, I attempted to record as many details as possible about the activities and interactions that took place. In many cases I recorded teacher and student commentary verbatim. I drew diagrams and sketches that would help me visualize the layout of the classroom or the nature of particular activities at a later date. Additionally, at times during the lesson I added analytic notes to my field notes (Glesne, 2011). I used brackets to separate these analytic notes from my descriptive notes. In the brackets I noted my own feelings, reactions, or impressions. I did this in the moment in order to avoid forgetting these initial thoughts. I based

my own bracketing on the work of Ely, Anzul, Friedman, Garner, and Steinmetz (1991), which explained that

> bracketing requires that we work to become aware of our own assumptions, feelings, and preconceptions, and then that we strive to put them aside—to bracket them—in order to be open and receptive to what we are attempting to understand. (p. 50)

In keeping with the observer spectrum of the observer-participant continuum, I did not interact directly with students. However, if they approached me I responded or directed them to their teacher. Likewise, at various times some of the teachers approached me to clarify their actions, and I responded to them as well. When the teachers ended direct instruction and students moved into centers or independent work time, I circulated throughout the classroom to get a better grasp of the activities in which they were engaging.

In keeping with the suggestions of Hatch (2002), after completing the observation and leaving each school, I made a point to sit down in a quiet setting and "fill in" my research notes as soon as possible so that I might remember as much of the observation's occurrences as I could. This process involved looking at the raw data (the notes I had taken on my computer) line-by-line and adding any other relevant information. This often included adding more details about the nature of the lesson or interactions. It also involved adding thoughts, reactions, and acknowledgements of bias to the brackets I had created. I made sure that all direct quotes began with an identifier for the person speaking, such as "teacher" or "student," and used quotation marks to indicate transcribed speech.

Once I had completed the process of filling in the raw data, I saved each set of field notes in a file associated with the corresponding participant on both my laptop and external hard drive. I also saved a hard copy of the observation in each participant's folder. The following chart

illustrates the teacher observed, literacy components, grouping types, and the time I spent in each classroom.

Table 5: Observations: Literacy Components, Grouping, Time in Classroom

Teacher	Observed Literacy Components	Types of grouping	Time in classroom
Laura	Calendar Word Wall Practice Lesson on using evidence from the text to describe a character or setting	Whole group Whole group Whole group to small group to whole group	1 ½ hours
Karen	Writing Brainstorm Small group reading Literacy Centers	Whole group Small group Small group	1 ½ hours
Allison	Lesson on using sensory words to visualize using a poem	Whole group to small group to whole group	1 hour
Sally	Lesson on rhyming words, sequencing, and retell	Whole group to partner work to whole group	1 hour
Rebecca	Word work- blending short vowels Word Wall Practice Read Aloud Literacy Centers	Whole group Whole group Whole group Small group	2 hours
Joanne	Lesson on the differences between fiction and non-fiction using picture books Small group reading Literacy centers	Whole group Small group Small group	1 ½ hours
Kendra	Lesson on drawing conclusions Small group reading Daily 5 centers Seatwork to complete Drawing Conclusion Worksheet	Whole group Small group Small group Individual	1 ¾ hours
Jessica	Lesson on drawing conclusions Small group reading Daily 5 centers Seatwork to complete Drawing Conclusion Worksheet	Whole group Small group Small group Individual	1 ½ hours

Data source 3: Unobtrusive data. Hatch (2002) notes that unobtrusive data can provide "insight into the social phenomenon under investigation without interfering with the enactment of the social phenomenon" (p. 116). This is possible because unobtrusive data is not filtered through words or thoughts of those participating in the study (Web et al., 1981). Such documents that can be classified as unobtrusive data can help us better understand the workings of a particular institution (Hatch 2002; Patton, 1990). Based on this premise, I collected artifacts from each participating teacher; I felt these artifacts could give me insights into how her particular institution processed the TEAM evaluation. In most cases, teachers provided me with a copy of their most recent literacy evaluation conducted since implementation of TEAM. However, the nature of the documents varied from teacher to teacher based on their particular experiences. For example, some teachers had been observed multiple times in reading, while others had been observed once in reading and once in math. In most cases, the artifacts included the teacher observation report template and lesson plans with observer score and self-score. As another form of unobtrusive data, I also looked at each district's value added scores in reading as well as the scores of each individual school. I collected these scores from the Tennessee Department of Education Website, a public access website that provides documentation for the Report Card for Tennessee Schools.

After collecting each artifact and public documents, I stored a hard copy in each participant's corresponding folder along with copies of classroom observations and interview transcripts.

Data Analysis Procedures

Timeline. I believe that data analysis often informally occurs in tandem with data collection because even in the first moments of collection we are already analyzing what we see

and hear. We as researchers are already making decisions about what to include or ignore and what questions to ask or avoid (Hatch, 2002). For example, before conducting interviews I was already making decisions about what questions to ask my participants. Likewise, as I took notes during classroom observations I was making decisions about what to include and what to ignore. In this manner, informal analysis was occurring from the outset of my study. Nevertheless, formal analysis began in early January after all interviews had been transcribed. In this section, I detail how I went through the analysis process.

Inductive analysis. My goal as a qualitative researcher is to employ a "systematic search for meaning" through the act of data analysis (Hatch, 2002, p. 148). I employed the strategy of inductive analysis, a qualitative method that begins with the particular and moves to the general. I began by examining fragments of my data, probing for patterns or connections, and then creating meaning. Potter (1996) argued that inductive analysis involves "looking for patterns across individual observations, then arguing for those patterns as having the status of general explanatory statements" (p. 151).

Phase 1: Coding. I began this phase of data analysis by reading carefully through hard copies of my collected data. I did this to get a general sense of my data before beginning the process of formal analysis. I then organized all pieces of data by participant and had on hand a variety of colored pens. I made the decision to do my coding by hand rather than using a computer program. I believed that my data set was small enough to do so, and I did not feel a high level of comfort with existing computer coding and analysis software.

Saldana (2009) defined coding in qualitative inquiry as a "word or short phrase that symbolically assigns a summative, salient, essence-capturing and/or evocative attribute for a portion of language-based or visual data" (p. 226). I chose to engage in the following three

cycles of coding: in vivo coding, descriptive coding, and emotive coding for interview transcripts. I engaged in descriptive coding for observation field notes and collected documents and artifacts. As I coded I tried to keep in mind that my choices are always impacted by my own experiences and positionality and that therefore codes are a reflection of these influences (Sipe and Ghiso, 2004).

Coding cycle 1: In vivo codes. The root of in vivo comes from the phrase "in that which is alive" (Saldana, 2009), and the focus is solely on the words of the participants. From the initial phases of the study, I have been interested in the experiences and perspectives of the teachers who are participating. Therefore, I believed that in vivo coding was an essential mechanism for me to highlight the voices of my participants. I began by going through the transcripts and interviews with a blue pen and underlining actual words and phrases spoken by the participants that I found particularly salient. These are a couple of examples of in vivo codes from Rebecca's interview: "I didn't want to put on a show"; "You can't go into a trial and already have an opinion"; and from Amber's interview: "I feel like I can't ever take a detour." I attempted to note participant language that gave insights into teachers' perceptions and helped me answer my research question.

Coding cycle 2: Descriptive codes. Descriptive codes condense a topic of a particular passage of data into a key word or phrase (Saldana, 2009). After completing the cycle of in vivo coding, I moved into the task of descriptive coding. I applied the method of descriptive coding to all collected documents including interview transcripts, observations, and teacher evaluation data and lessons. I went through each piece of data line-by-line. When I came across passages of data dealing with topics that I believed to be essential to my purpose I summarized them in a key word or phrase. I recorded each descriptive code in red ink in the margin of the corresponding

passage. This is an example of a passage from Sally's interview in which she discussed her feelings about the school-wide average being used as a part of her evaluation score:

> I about fell on the floor. I was very upset with it. I didn't feel like it was fair you know when you look at like what my kids did for the year and I've got a 4 but I've got to sit here with a 2 total because of the school wide.

In the margin beside this passage I wrote the descriptive code "feeling misrepresented by scores." This is an example of a section of data from Laura's interview:

> When the model is part of their teaching, it's not something that's slapped on top of teaching. What I mean when I say that is like let's say I see someone put up their objective. They print it out. They put it at the front. They keep repeating it in very technical language. They haven't made it user-friendly. Obviously it's not part of their daily routine. The kids don't know where to reference it.

In the margin beside this passage I wrote the descriptive code "surface level change." My hope in creating descriptive codes was to answer Saldana's (2009) basic question: "What is going on here?" (p. 1495).

Coding cycle 3: Emotive coding. The third and final cycle of coding in which I engaged was that of emotive, or emotion, coding. I applied this cycle of coding to all interview transcripts. Corbin and Strauss (2008) noted that "one can't separate emotion from action; they are part of the same flow of events, one leading into the other" (p. 7). I posited that emotive coding could be especially helpful for me because of the nature of my research question, namely a focus on teacher perception. Often perception and emotion are closely related. As I went through this cycle of coding, I read through the data line by line. When I came to a passage that I

felt illustrated a particular emotion I wrote a key word or phrase in the margin beside those lines in green ink. The following is a passage from Leslie's interview transcript:

> I was in shock. The first six months I was just in shock. This cannot be. I don't want to be a teacher anymore because of this. I don't really even like the lead teacher position. I think it's unfair to put me in a position where I'm taken away from my students, and I could potentially have my coworkers angry with me.

In the margins of this passage I recorded the emotive codes "frustration" and "despair." The following is a passage from Allison's interview transcript in which she discussed her lack of understanding of how scoring works within the TEAM model:

> I don't exactly know. There's a lot of words that get thrown around, so I just usually nod my head and go, okay. I know part of it is achievement and part of it is gains. Right? Does that sound right?

In the margin beside this passage I wrote the emotive code "confusion." My hope was that by attending to the emotions of my participants I would be better able to understand their positions.

Phase 2: Analytic memos. Saldana (2009) posited that analytic memoing is a natural result of the coding process and that codes tend to spur the need for written reflection as we ponder each code's deeper insights. Mason (2002) noted that analytic memos can encourage "thinking critically about what you are doing and why, confronting and often challenging your own assumptions, and recognizing the extent to which your thoughts, actions, decisions shape how you research and what you see" (p. 5). I engaged in analytic memoing as a process of reflexivity. As I completed coding cycles for each participant's data set, I then sat down at my computer and wrote my thoughts and ideas. This writing was informal, not always in complete sentences, and often addressed a direct reaction to a code I had created for that participant. I

completed an analytic memo for each participant, and each memo encompassed several typed pages highlighting my thought processes. Within the memos I put codes in all caps and also included segments of participant's quotes. I put my own thoughts in brackets. The following is an example of a segment from one of my analytic memos. This memo is based on Allison's data.

> FORCED INTO DEEPER CHANGE page 1 b/c of announced observations. You had to be teaching that way all the time bc you never know when someone might come in to observe. "It kind of forced you into making your teaching that way all the time so when someone walked in your classroom you weren't trying to do something that wasn't natural to you and the kids weren't going we never do that." [The word forced is interesting here. It seems like Allison has truly changed but what about the word forced? Is that the way change should be?]

The following is an excerpt from one of my analytic memos based on Sally's data:

> CONFUSION Sally addresses the confusion amongst teachers. Says principals didn't have understanding. Lack of training = stress. Makes me wonder—should we implement something in such a quick fashion without ensuring supervisor understanding? How can we ensure that administrators will appropriately implement a system if they don't understand it? Also how much buy in will people have if they have no say/stake in a new implementation?

The act of analytic memoing allowed me to engage simultaneously in the processes of reflection and analysis and helped scaffold the next step in which I began to look for patterns amongst my codes.

Phase 3: Moving from codes to categories. After completing the three cycles of coding with all pieces of data, I began to consider which codes might fit together into categories or

beginning patterns. Grbich (2007) explained that codes can be "segregated, grouped, regrouped, and relinked in order to consolidate meaning and explanation" (p. 21). This process involved considering which codes "look alike" and "feel alike" (Lincoln & Gubba, 1985, p. 347). I began by going back through all my data and writing each code—in vivo, descriptive, and emotive—on index cards. I color coded each index card based on which participant the data was derived from. From example, all codes that came from Karen's data were written in yellow on the index card. Each participant had a different color. Then, in a large space on the floor, I began tentatively grouping index cards together that I felt to be similar. For instance, I began placing index cards together that shared broad basic ideas such as scoring or focus on the rubric.

I went through second and third cycles of categorizing and re-categorizing the index cards. Whereas in the first cycle I concentrated on creating patterns, in the second and third cycles I focused on grouping like patterns together and refining my categories. Abbott (2004) related the process to "decorating a room; you try it, step back, move a few things, step back again, try a serious reorganization, and so on" (p. 215). Similarly, I would group codes together, take a step back and examine them, and, in many cases, move them again. Some codes I removed altogether if I saw that they, while interesting, may not have related to my larger purpose. Some initial patterns were combined into one group if I found them to be similar. For example, I initially had three categories that addressed school level, district level, and state level issues respectively. In the third cycle of coding, I made the decision to group all these issues together, which eventually all fell under the theme of "issues of power."

I began with approximately 80 to 90 codes, which I then grouped into approximately 10 to 15 groups or categories. From there I began the process of creating themes. Richards and Morse (2007) explained the process of moving from coding to creating patterns to creating

themes in this manner: “Categorizing is how we get up from the diversity of the data to the shapes of the data, the sorts of things represented. Concepts are how we get up to more general, higher-level, and more abstract constructs” (p. 157). In this way, I begin to look closely at my patterns and ask the question: What is the big meaning that this pattern coveys? For example, as I noted earlier, I had grouped together issues teachers discussed on the school, district, and state level in reference to the TEAM model. As I studied these codes and patterns, I realized that the big issue, or theme, related to power structures and those in decision-making roles. Therefore, I created the theme “issues of power.” In the diagram below I have demonstrated an example of how I moved from codes to patterns to themes.

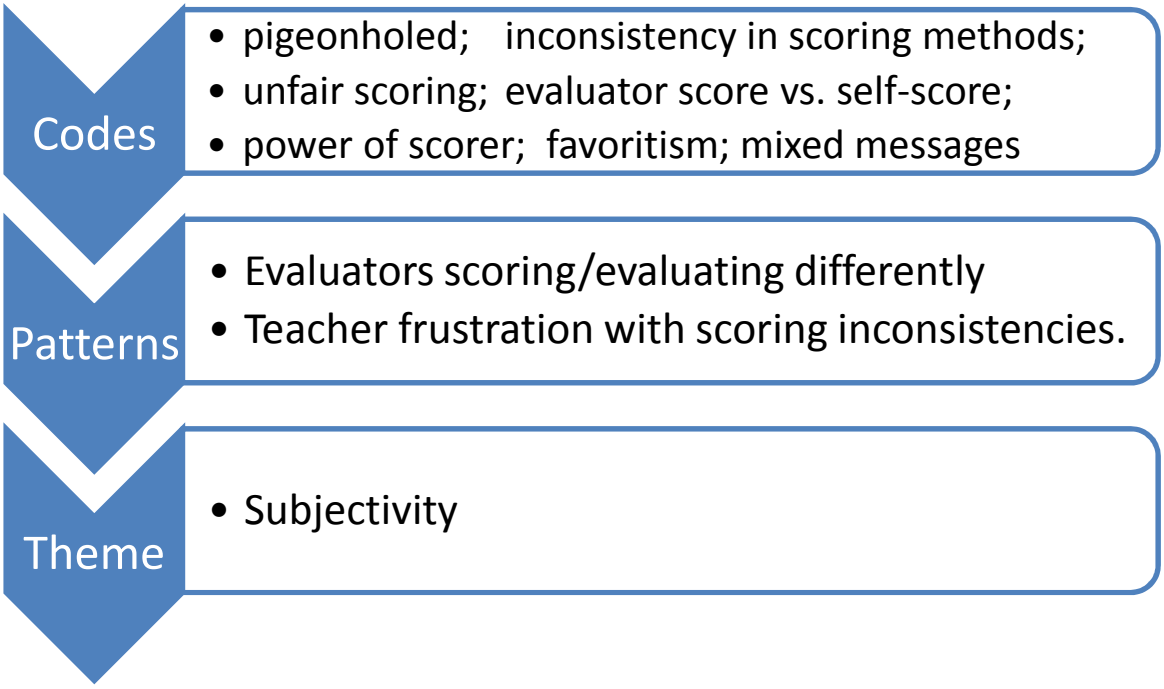


Figure 1: Moving From Codes to Patterns to Themes

After completing this process with all sets of patterns, I created the following themes: teachers want to be held accountable, initial reactions to TEAM, change, administrator role,

subjectivity, issues of power, and possible improvements. I created individual folders with the name of each theme written on the outside. Inside the folder I kept the index cards that had codes and references to corresponding page numbers in the transcripts, field notes, and analytic memos. The cards were paper clipped together and labeled with a sticky note denoting the name of the pattern. I also created an outline for each theme to assist in the writing process that I also stored in individual folders.

Summary of Chapter

In this chapter I have described the methods I employed to collect as well as analyze the data of this study. I also made transparent my own epistemic leanings and discussed how they may impact both the collection and analysis of data. In the next chapter I present my analysis of the collected data.

Chapter 4. Analysis

Chapter Introduction

The purpose of this study was to describe the experiences of eight first grade teachers in the Tennessee public schools in regards to the implementation of TEAM, a new policy that changed the manner in which teachers are evaluated. The study answered the research question, "What are the common elements in teachers' experiences with the TEAM evaluation model?" The following sub-questions were also addressed:

a. What are the common elements in teachers' perceptions of the influence of TEAM on literacy instruction?
b. How does the context of teachers work complicate their perceptions of the influence of TEAM on their work?

In this chapter I provide my analysis of the data collected for this study which works to answer the primary research question and the sub-questions. I have included information relating to the larger themes I created as well as numerous segments of the participants' actual commentary. My hope is that the analysis and data excerpts will help the reader better understand the perceptions of the participants, the context of each participant's particular situation, and how that context may or may not impact individual perception.

The first few sections of this chapter focus on answering the primary research question, "What are the common elements in teachers' experiences with the TEAM evaluation model?" I use the following themes as a framework to address the research question: teachers want to be held accountable, initial reactions to TEAM, change, administrator role, subjectivity, issues of power, and possible improvements.

Teachers Want To Be Held Accountable

Regardless of individual reactions to the implementation of the TEAM model, several teachers interviewed expressed strong feelings about the importance of being held accountable for job performance. For some, this seemed to stem from experiences that led them to believe that a certain percentage of teachers are failing to carry out their jobs in a proficient manner. Both Sally and Rebecca described experiences which led them to value accountability. Sally explained, "I don't have a problem with us being held accountable. I think it needs to happen because there are teachers sitting there doing nothing and you know we have a few at our school." Rebecca, too, cited occasions in which certain peers neglected to carry out their duties to their full capabilities: "I saw teachers, and I know because I was in the room, who never did a lesson plan unless it was their planning evaluation—the whole year. Never a lesson plan at all. Not even jot anything down." Several participants felt that without some measure of accountability teachers such as these would continue engaging in sub-par work.

For some teachers, TEAM has been a method of shedding light on those individuals who are not teaching in an exemplary manner. For example, Joanne described the evaluation system as a method of addressing problematic teaching, or as she puts it, "weeding out the weeds." On a similar note, Laura explained that one positive implication of the evaluation system is that "it really divides the wheat from the chaff. You can see how well a teacher can teach." Sally found the necessity for accountability particularly palpable in a district like hers, where teachers typically were evaluated only every two to three years rather than every year. Depending on the habits of the administrator or the requirements of the district, this could mean that teachers could go many years without being observed at all. Several interviewees viewed this lack of oversight as a means of encouraging unacceptable teaching behaviors.

Some teachers regarded accountability not just as an apparatus for addressing problematic teaching, but also as a way to validate quality instruction. Those who fit into this category found it frustrating to be working hard and engaging in what they viewed as exemplary teaching but to feel that it was never acknowledged. For Rebecca, this validation can only occur through frequent classroom visits. She noted,

> I want you to see what we're doing. I want you to be excited about this too. I'm excited. And you don't want to go woo look at me! So I'm glad they see that. I'm glad they know what you're made of.

From this perspective, the need for accountability is twofold: (1) to work to eliminate ineffective teachers and (2) to validate the efforts of effective teachers.

Initial Reactions to TEAM

While many of the participants expressed positive reactions to accountability in general, they all described negative initial reactions to the implementation of the model. These reactions ranged from feeling inconvenienced to stressed and worried. At the onset of implementation, Laura described teachers as being "very anxious and overwhelmed." She explained that her own reaction was quite negative: "I was just in shock. The first six months I was just in shock. This cannot be. I don't want to be a teacher anymore because of this." Sally, too, explained that the "school system was in uproar. It was a mess. Our second grade teacher cried over this." Initially, many of the teachers described their anxiety and fear as being directly related to job security. They worried that if they could not perform adequately on the new evaluation that their jobs would be in jeopardy. For example, Kendra noted that it was "a lot of pressure all at once. It would keep me up at night almost. It is so, so much pressure like you have to do well or you're going to get fired right away."

Negative reactions seemed to stem from a couple of different areas. Some teachers, such as Kendra, feared that they just wouldn't measure up. Kendra explained, "When I read that rubric my first instinct was I must be a terrible teacher because I don't know how much of this I actually do." These fears seemed to be compounded by the discourse surrounding implementation, which touted the unlikelihood of receiving high scores. Most participants were told that on a scale of one to five, a three would be an achievable goal. Joanne described being "scared to death because they were telling us that three would be the score that you would strive for. Fours and fives would be nearly impossible." This notion did not sit well with many. Jessica explained, "It's like if you were taking a class, you would not be able to make an A. It would be impossible. You feel set up for failure." Lindsey reacted similarly and also compared the evaluation to a traditional grading system. She felt that teachers are particularly sensitive to being limited to what they might view as a mediocre score:

> The model is set up almost for mediocrity because they're telling every teacher who is innately like a perfectionist and a Type A personality, that's really what you have to be to be a teacher, that you're getting a C and really that's good. Don't worry about it. And they're like, this is not okay for me.

Rebecca also explored the idea that teachers on the whole just don't feel comfortable with a mediocre grade: "It doesn't matter what teacher you talk to. Nobody wants a three. It really rubs us raw." While the message seemed to be that teachers should be satisfied with a three, this did not transfer into reality for most. For most of the participants this score just wasn't sufficient in terms of meeting the personal expectations they set for themselves.

All participants found the initial implementation of TEAM to be an unfavorable experience. However, for some these early fears remained prominent as the process was carried

out while for others they were almost completely dissipated. Kendra, for example, found that TEAM wasn't as intimidating as she originally thought because the new system did not mean she was in danger of losing her job. Additionally, Joanne found that the scoring of observations was not as restrictive as she had previously believed: "When we [Joanne and her principal] met, I had fives on some things, some fours and then threes. And I thought, well that wasn't so bad. Fives were a reachable goal." For some participants, on the other hand, initial reactions to TEAM have changed little during the months since implementation. In the next section, I explore what factors may or may not have contributed to changes in opinion of the model as well as change in instructional behaviors.

Change

In this section which focuses on the theme change, I will continue to answer the primary research question by examining the experiences of participants in terms of the differences and commonalities in levels of change. I will also seek to answer sub-question (1) What are the common elements in teachers' perceptions of the influence of TEAM on literacy instruction?

In my interview with Laura, she highlighted the notion that change is hard because often the process of change involves having to acknowledge oneself as flawed in some way. She believed that change cannot occur without this revelation and that our own egos may suffer in some form. As she reflected on the previous months she noted,

> I think we were all in the same place in the beginning. We all had a little too much pride and we all really cared more about ourselves. I don't mean that some of these people don't care about their students but they have to put themselves down to put their students first. I see that I'm not doing something well; I'm going to put my students first, drop my pride and do this.

As I spoke with the interviewees about their personal change as an instructor in relationship to TEAM, two broad groups emerged. One group consisted of individuals whose instruction changed after the implementation of the evaluation model; the other group consisted of teachers whose instruction did not. Within the two broad groups of change versus no change, I found four distinct categories. Under the umbrella of change were: 1) those teachers who connected to TEAM, internalized it, and believed that their instruction was improved as a result and 2) those teachers who did not necessarily believe in the contents of the model but, due to fear of repercussions, felt forced into change and ultimately determined that their instruction was improved as a result. Under the umbrella of no change were: 1) teachers who did not believe in the model and changed their instruction only on a very surface level and 2) teachers who did not believe in the model and refused to change their instruction. Additionally, some teachers could be placed in multiple categories in that pieces of their instruction were subject only to surface level changes while other aspects were altered on a deeper level.

Karen, the newest teacher in the group of participants, is an example of someone who for the most part connected with TEAM, began implementing it in her classroom early, and reported positive reactions to the implementations. She viewed the evaluation system as a catalyst for improvement in instruction. She noted, “I feel like it does really enable us to be able to look at what we’re teaching and really sit down and make sure we are getting this in there, and we are getting this in there.” She believed that she would benefit from classroom observations and the insights of others. Karen explained that “just having different people come in and getting to see you teach” was a positive outcome thus far

> because sometimes what you think you do really well at they have a suggestion for you of how to improve it. And getting you to think about oh I didn't even think I did that or didn't do that. I didn't even realize I didn't put that in the lesson.

Additionally, she argued that the model encourages valuable reflection, which can spur good teaching:

> Last year after all my post conferences I felt like I learned something about myself. They try and give you a positive and something that you can improve on. I felt like that really helped me the best. And it's not all you didn't do this or you need to do this better. It was this is what you're doing good, continue to do this, but this is the piece you can work on. I think it really does help in the reflection process and going home and being like how can I do that better. Now that I know that this isn't really working what can I do next time?

Karen dismissed the notion of TEAM as negative or stressful and pointed out that it is simply a tool to move teachers forward:

> I know a lot of teachers tend to stress out more about it now, but really it should be something that you're doing in your lessons every day. I don't think it is something that's there to scare you or to stress you out. I think it's there to really get you thinking about your lessons and trying to make sure that you're putting into it what you need to be putting into it.

Karen also discussed the necessity of making TEAM a part of a teacher's daily teaching routine. She believed that surface level change is a futile effort:

> If you try and do it for that one lesson that they come in the room, it's not going to work. So it's something that you want to try to do on a daily basis. I think that having that in

> place really makes you think about every day, even though they're only in your room four or six times or whatever it is. I feel like if you try and do it for that one lesson that somebody's coming in there the kids are going to be like, what, this is not our normal routine. So I think it really makes you think what do I need to do on a daily basis and try and get those things in there.

While many of the other teachers may not have implemented TEAM so whole-heartedly at the beginning of the process, a few admitted that they considered it a catalyst for a type of "forced" change. In other words, while they may not have agreed with the premise behind the model, they implemented it anyway because they felt they had no choice, and, as a result, they saw some improvement in their instruction.

Allison is an example of a teacher who felt that she was forced into change, despite her misgivings. As an "apprentice" teacher, someone in the early stages of her career, she felt that she had to follow the requirements laid out by the evaluation model on a daily basis. The requirements became a part of her daily teaching in large part because of the possibility of unannounced observations:

> You had to be teaching that way all the time because you never know when someone might come in to observe. It kind of forced you into making your teaching that way all the time so when someone walked in your classroom you weren't trying to do something that wasn't natural to you and the kids weren't going we never do that. Six year olds would definitely say that.

Although Allison didn't initially seem to want to implement TEAM in her classroom, she reported some positive outcomes in her instruction after doing so. She particularly cited improvements in terms of reflection and intentionality:

> I fine-tuned the way, the facets in which we're supposed to teach, like the way we're supposed to teach our curriculum, and [it] just made us really focus on the why behind why you do certain things and just being really intentional about how you structure a lesson.

The cases of Allison and Karen demonstrate that deeper changes did occur in some classrooms on a daily basis as a result of the new evaluation model. However, in some classrooms change was purely superficial. For the most part, teachers in these classrooms saw a chasm between what they viewed as effective instruction and the requirements of the new model, but they feared the repercussions of non-compliance. Thus, they implemented the requirements on a surface level. Laura described this as putting on "a dog and pony show," in other words, teaching in a manner for the evaluators that you would not on a regular basis. Rebecca explained that "there are complete ways to just almost create this fictitious image of who you are and that not really be who you are." For some that might mean having a prepared lesson on hand "just in case" an evaluator walks in to do an unannounced observation, or it might entail creating a lesson for an announced observation that closely follows the TEAM rubric but looks nothing like the teacher's everyday teaching. While most participants said they did not engage in the process of surface level change, a few found it to be a wise strategy. Joanne relayed an incident in which a peer planned a lesson that she believed would ensure a good score on the TEAM rubric:

> A teacher made the comment that when she knew her announced observation was coming up she sat down and made sure that her lesson had every one of those things [on the rubric]. She normally doesn't do that. She didn't stick with what she normally does. She took out that rubric and actually made sure that she was doing those things. I never did that, but I thought that was pretty smart. And then on the other end you have teachers

> who say I do my lesson plans one way until they want to come look at them and then I make sure it has everything on there that they're looking for, every little compartment. I never did that and it showed in some of my scores. If I had looked at the rubric and paid more attention to it … those [scores] would not have been twos if I had just paid attention to the rubric.

The final group to consider is those who do not believe in the new evaluation model and refused to change their instruction regardless of the consequences. For these individuals, TEAM comes into direct conflict with what they view as effective instruction and best practices for children. Additionally, they may feel that surface level change or "putting on a dog and pony show" is an unethical practice. Jessica explained, "I try to put the children first and the evaluation second." For her the two are separate entities. Rebecca discussed her issue with surface level change as she described teachers who follow the requirements of the model only if they know that an evaluator will be coming to watch them:

> unless it was the TEAM planning and [they were] going to get a grade for it and [they] had to turn it in. And so I don't understand that because that goes against every grain in my logic and thinking, but I've seen them do it.

She also explained her own reaction to this mentality:

> They tell you don't teach to the test but at the same time as a teacher I didn't want to teach to the rubric either. I didn't want to put on a show and do something that I wouldn't normally do every day. So I did pretty much what I normally did and let it kind of be my baseline from that point on and I think that for me was a better perspective.

The teachers who took this stance tended to be willing to receive a lower score on their evaluation in order to continue teaching in the manner they viewed as best practice. Sally

described an environment at her school in which "most of the teachers keep doing what they're doing and carry on whatever happens." Even Karen, who was an example of a teacher who internalized and had positive reactions to the model, cited a scenario related to time management in which she would be willing to sacrifice a higher score for student understanding:

> You've put it [the approximate length of time of the lesson] in the lesson but things don't always go like how you planned it to. I feel like the most important thing is to make sure that the kids are getting it. You don't want to just move on through the lesson even if they're not getting it. I feel like it's more important to make sure that the kids are understanding than to actually get to that closing. So don't rush through just to get all the pieces in I feel like. And you may not get as good of a score because you didn't get that closing in.

In this section, I have highlighted four different positions on change in teaching practice related to implementation of the TEAM evaluation model including refusal to change, surface-level change, forced change, and willingness to change. In the next section, I address specifically how the model impacted literacy instruction in the interviewees' classrooms.

Specific changes in literacy instruction. In the previous section, I focused broadly on participants' levels of change. In this section, I seek to answer sub-question 1: What are the common elements in teachers' perceptions of the influence of TEAM on literacy instruction? Nearly every participant identified both positive and negative changes in their literacy instruction on some level as a result of TEAM. The only exception to this was Sally who reported no change as a result of the model.

Kendra claimed the following as elements of her literacy instruction were improved as a direct result of TEAM: teaching higher-order thinking skills, intentional grouping, providing

students with academic feedback, and lesson planning. She explained that in terms of helping students to engage in the processes of higher order thinking she is "really trying to make sure that [she is] getting them to think beyond just here's a word, let's read it. And think about how we used it and what else does it apply to." She reported that her method of grouping students has become less haphazard. While she has always viewed herself as intentional in relation to small reading groups, she explained that she is now also more intentional when planning other types of groupings such as turning and talking to peers. Additionally, she believed that the amount of preparation going into lesson planning has increased:

> It's pushed me to be more prepared I think. You never know when you're going to be observed with the pop in observations, so I'm more thoroughly prepared than I was before. And I'm making sure that every day, every minute, there's not a minute wasted.

Joanne believed that TEAM has altered the way she thinks about grouping as well as teaching her students to engage in questioning strategies. She explained the manner in which she realized her method of grouping was lacking:

> I realized last year that I was spending more time in whole group than I was spending in small group time and independent practice time. So that kind of looking back at that rubric and them telling me this is what you need to work on. I understood it then.

In terms of questioning, she described the change from using the strategy as a "teaching tool" for herself to using it as a "learning tool" for her students. Previously, she had an essential question written for each lesson, "but it never went past that in any of [her] lessons." Now she has question stems posted around the room, and students work to answer them in teams. She explained,

> So I was able to take what they told me to strengthen or to change with my questioning and once I became comfortable with it and the kids were confident in their questioning abilities, they could then go to that part of the room and answer that question with their team, and then bring it back and teach the [rest of] the class. I have them posted and they're in kid-friendly language. The kids know them. It's not just me utilizing the question. I'm asking them now.

For Jessica, the evaluation model has influenced how she considers the elements of problem solving during literacy instruction. Jessica acknowledged that problem solving is something she previously attributed solely to math instruction and that since using TEAM she has begun to think about how problem solving might be applied to literacy. She noted,

> The big thing that I've really focused on in literacy that I realized I didn't do is, or I'm having a hard time wrapping my mind around, is showing problem solving in literacy. I think it's the way I define problem solving. I immediately think math problems or word problems. So coming up with things that are those problem-solving tasks in literacy. I've been able to kind of fit in a lot with comprehension type questions, responding to literature, [and] classifying and kind of hitting some of those problem solving skills and putting things into groups or justifying their solutions.

Karen felt as if the assessment component of her literacy instruction has been improved. She believed that she now plans assessment in a more intentional manner.

> I feel like I am more aware of what I am putting into my lessons. I feel like I'm actually trying to sit down and make sure that I have that assessment there. I think it's something that everybody tries to do anyway but this really makes you [ask:] am I doing what I need to do?

While most of the interviewees were able to identify positive changes in their literacy instruction, the same teachers also saw some negative implications of TEAM in their teaching of reading. Kendra was particularly bothered that she saw portions of her reading instruction time being consumed by activities and requirements she viewed as unnecessary. For example, she felt that the focus on relaying standards to students was inconsequential in their learning:

> The amount of time I've spent cutting out standards to post on my walls and thinking up ways to tell the students what they're doing when they could get by with an essential question or a learning target. But now having to find a way to post [standards] and go through it and their eyes glaze over when we go over the standard for the day. Little things like that just eat my time up. There's a couple of things I do that I feel like are mainly just in case someone else walks in [rather] than for the benefit of the students.

Kendra also noted that one area that she views as particularly essential when considering effective literacy instruction, small groups, may not align well with the TEAM rubric. Kendra explained that while the evaluation calls for demonstration of different types of groupings, if you were to be observed solely during small group time you most likely would not receive a high score. She cited a recent literacy observation in which she was evaluated during small-group time. She explained that the standards differed in each small group and did not match to the initial standard that she chose for the whole group:

> The main skill that I taught was not the skill that I wanted to hit with them in their small reading groups. Now I haven't had my post conference yet so I don't know if he's going to say that your standard did not match with what you taught in your small groups. I just went ahead and did what was best for them. We talked about main idea because that's what they needed help with that day. I was confused and a little frustrated by that because

> you don't teach the same thing to each group. When you work with small groups you don't hit just one standard. You hit a different standard almost for every group depending on what they need to progress. Even if you are doing one standard, you interpret it differently for each group.

Allison, too, was concerned about how her evaluation score might be affected if she were to be observed during small-group reading. She concluded that the rubric is looking for different groupings, specifically a particular progression from whole-group to small-group to independent work time within each lesson. She feared that small group reading does not fit well into this framework: "I've never had a TEAM during small group time. I'd be a little afraid of what it would look like just because it's such a short amount of time with so many groups."

Jessica saw the following literacy practices as unlikely to receive a high score on the TEAM evaluation rubric: literacy centers, Daily Five, and independent reading. She questioned whether or not an evaluator would be able to assess instruction appropriately if a child was reading independently or working in center:

> I feel like a lot of that does fit into the rubric because they are reading with a partner and then they're working independently and it's a lot of transitioning. But as far as it being higher order or you being able to assess that … you can look at a kid reading to themselves and you can usually tell if they're reading, but you can't really prove that they're reading and comprehending. I think some of those things would kind of be like, well how do you know, how do you assess this?

Although several participants identified sources of both positive and negative change in literacy instruction, some teachers were not convinced that change in their schools was a direct result of the evaluation model and noted that change could be a result of other sources. Allison,

for example, saw the incoming of Common Core Standards as an important catapult for change. Sally pondered whether the changes that were occurring in her school were a result of a new literacy director, the TEAM evaluation, or Common Core Standards. She related that for her personally, change in literacy instruction has been a result of her choice to further her education in a reading specialist program. According to Laura, change in instruction will only occur if individual teachers are willing to accept and internalize new ideas,

> I can see that people if they don't want to change their teaching style, will keep it [the rubric] as a checklist. But if they want to change their teaching style they will make it a part of their life. If you make it a part of yourself, it will really help you be an excellent teacher.

In this section, I have examined participants reported perspectives on change with a particular focus on how their literacy instruction was impacted. Next, I triangulate each teacher's level of change with the following data: 1) classroom observations that I conducted, 2) TEAM observation reports and lesson plans that I collected and analyzed 3) reported beliefs on literacy instruction. The following chart illustrates each of the above criteria. Below I discuss how beliefs, everyday practice, and observed practice may or may not connect for each participant and how these elements may or may not connect to levels of change and acceptance of TEAM.

Table 6: Triangulation of Beliefs, Daily Observed Teaching, and TEAM Evaluation

Teacher	Stated Beliefs about Literacy Instruction	Observed Lesson	TEAM Evaluation	Connection between beliefs, observed lesson, TEAM lesson	Reported Level of Change
Laura	High expectations; real life and cross curricular connections; holistic teaching	Calendar and spelling aloud word wall words- whole group; lesson on using evidence from the text to describe a character or setting; whole group modeling, students move to small group and use a text to find evidence; share	Lesson focus: sorting music into exclamation or statement sentences Introduction: whole group- modeling; Guided/Independent Practice- students create a sentence in small groups and decide whether it is a statement or exclamation Closure- whole group wrap up- listening to a song and deciding whether it fits with statement or exclamation	Highly connected	Forced change; some real change; some surface level change
Karen	Time for students to read; access to variety of reading materials; modeling thinking strategies	Whole group brainstorm on projector- what is fun about Halloween; move to small groups and centers- small group- blending and writing sounds, students fill out character trait sheet using a picture and discuss	Lesson on visualization. Whole group- model using a text to visualize using sensory words. Partner work- students read a text without illustrations and draw what they see. Individual task- students draw another illustration to go with another piece of text.	Some connection	Internalized model; high level of change

Table 6. Continued.

Teacher	Stated Beliefs about Literacy Instruction	Observed Lesson	TEAM Evaluation	Connection between beliefs, observed lesson, TEAM lesson	Reported Level of Change
Allison	Importance of small group time; hearing students read, having conversations about books	Observed TEAM lesson	Focus- using sensory word to help visualize Introduction- students write words associated with 5 senses on corresponding posters; Teacher model reading poem and drawing visualization Guided /Independent Practice: students listen to poem and draw what they visualized; students write why visualization is important Closure- whole group, class discusses how visualization helped them understand/retell the poem	Little Connection	Forced change
Sally	Differentiated instruction; building and sounding out words	Reviews what a cumulative story is and sequencing. Reads aloud. Lists rhyming words from story. Students then draw/write what happened in the story in order on paper using 8 boxes. Teacher also models on smartboard. Students then retell story to a partner.	Small group reading rotations. Each group focused on different skills. The following skills were addressed: spelling sort; beginning, middle, and end; syllables; plot; text-to-self connections; written response to story	Some connection	No Change

Table 6. Continued.

Teacher	Stated Beliefs about Literacy Instruction	Observed Lesson	TEAM Evaluation	Connection between beliefs, observed lesson, TEAM lesson	Reported Level of Change
Rebecca	Heavy phonemic awareness focus; modeling, allowing for student response	Whole group practice blending short vowel sounds using cards; spelling word wall words visually; seatwork- workbook page glue sentence with corresponding picture; read aloud with discussion of beginning, middle, and end; literacy centers	Whole group practice blending sounds; word wall practice; read aloud with focus on Author's purpose; Move to small group and literacy stations; stations included word work, writing, computers.	Highly connected	Some real change; refusal to make surface level change
Joanne	Word study; small group time; literacy centers	Lesson on difference between fiction and non-fiction; use of children's picture books, move to centers and small group; small group focus on blending sounds	Not available	Some connection	Some real change; some surface level change
Kendra	Small group time; differentiated instruction, balanced approach including sight words and phonics; avoidance of busywork	Lesson on drawing conclusions- whole group using smartboard; move to seat work (drawing conclusion worksheet), small groups, Daily 5 centers	Lesson on recognizing irregularly spelled words. Whole group- review of sight words. Guided/ Independent Practice- students write sentences. Conclusion- revisit standard.	Some connection	Some real change; mostly surface level change

Table 6. Continued.

Teacher	Stated Beliefs about Literacy Instruction	Observed Lesson	TEAM Evaluation	Connection between beliefs, observed lesson, TEAM lesson	Reported Level of Change
Jessica	Acknowledgement that students learn in different ways; daily word work, phonics, phonemic awareness	Lesson on drawing conclusions- whole group using smartboard; move to seat work (drawing conclusion worksheet), small groups, Daily 5 centers	Lesson on sorting long and short vowel wounds. Introduction- review standard, brainstorm vowel sounds, complete group sort. Students complete individual sort. Review sort for center work. Move into Daily 5 centers.	Highly connected	Some real change; refusal to make surface level change

Laura (Highly Connected): There was a clear alignment between Laura's observed lesson and TEAM lesson. Both lessons followed a similar format in which she introduced a concept in a whole group setting, moved students into guided practice, and then brought them back to whole group for closure. The lesson I observed could have easily substituted for the TEAM lesson. Laura indicated that after a period of forced change, she internalized the TEAM model and began teaching a manner closely aligned to the expectations of the rubric on a daily basis. My observation of one of her daily lessons seems to verify this. Her stated beliefs were somewhat general, but in both lessons she did seem to employ high expectations and real world connections which she named as primary beliefs.

Karen (Some Connection): There was some connection between Karen's observed lesson and TEAM lesson. For instance, both followed a similar format that moved from whole group work to guided practice to independent work. However, the elements of the observed lesson did not necessarily relate to one another in the same way that the elements of the TEAM lesson did. Additionally, Karen's stated beliefs about literacy instruction which included time for students to

read and access to reading materials were not apparent in these two lessons because students were not reading in either.

Allison (Little Connection): Unexpectedly, Allison's observed lesson also ended up being a TEAM observation. Therefore, I was unable to view an example of her daily teaching. There was little connection between her stated beliefs about literacy instruction and the TEAM lesson. She discussed the importance of small group reading and student discourse surrounding books. Neither of these was present in her TEAM lesson.

Sally (Some Connection): There is a strong connection between Sally's stated beliefs and her TEAM lesson. She explained that she believed strongly in differentiated instruction and her TEAM lesson involved small group rotations that employed different skills depending on the needs of each individual group. Likewise, she explained that she believed in the importance of building and sounding out words and her TEAM lesson involved spelling sorts. The lesson I observed did not demonstrate either of these beliefs. It is interesting to note however that there may be a connection between Sally's lack of change, her beliefs, and her decisions about lesson planning for the TEAM evaluation. Sally's instructional performance aligned closely with her stated beliefs despite the requirements of TEAM.

Rebecca (Highly Connected): There were strong relationships across all categories for Rebecca. She stated that she believed strongly in teaching phonemic awareness, modeling, and student response. Her observed lesson and TEAM lesson demonstrated all three. Additionally, Rebecca's observed lesson was extremely similar to her TEAM lesson which relates back to her refusal to make surface level change. Rebecca noted that she would not make changes she did not believe in and her observations seem to validate this.

Joanne (Some Connection): Due to Joanne's unexpected medical issues, Joanne's TEAM lesson plan was not available. However, there was a relationship between her beliefs and her observed lesson. She said that she believed in word study, small group time, and literacy centers and her observed lesson demonstrated all three.

Kendra (Some Connection): A connection was apparent in Kendra's classroom between her stated beliefs and my observed lesson. She said that she believed in small group, differentiated instruction and a balanced approach which included sight words and phonics. All were apparent in my observed lesson. However, there seemed to be less connection between Kendra's beliefs/ daily teaching and the TEAM lesson. While Kendra did incorporate sight word practice in her lesson she did not incorporate small group differentiated instruction. Instead she employed the more standard format of whole group instruction to guided practice to independent practice to whole group sharing. This decision appeared to align with her level of change which was mostly surface level. She appeared to change her instruction only when being observed by a TEAM evaluator.

Jessica (Highly Connected): There were strong relationships for Jessica amongst all criteria. She stated that she believed in differentiation of instruction for students and both the observed and TEAM lessons included small group time. Additionally, Jessica said she believed in the importance of daily word work, phonics, and phonemic awareness and both lessons incorporated aspects of these instructional practices. These connections also demonstrate Jessica's refusal to make surface level changes as a result of TEAM.

Thus, it seems that change is a salient issue for the participants in this study. In most cases, relationships were apparent between type of change, beliefs, and observed lessons. For many of

the teachers, their level of personal change in terms of instruction also seemed to be influenced by the actions and decisions of their administrator. In the next section I explore how participants viewed the role of their principal in the implementation of TEAM.

Administrator Role

In the next few sections I will continue to answer the question, "What are the common elements in teachers' experiences with the TEAM evaluation model?" Several of the teachers interviewed regarded the administrator role as critical in determining how a particular implementation gets carried out and that, as a result, the evaluation system may look very different from school to school. Additionally, in this section I will begin to answer the sub-question, "How does the context of teachers' work complicate their perceptions of the influence of TEAM on their work?"

Because Rebecca worked in three separate schools during the TEAM implementation process, she believed that she had a clear understanding of these differences. She explained that aspects of TEAM were valued in varied ways from school to school and that this variance can directly impact how a teacher is scored and how that score is viewed. She noted, "I felt like I got skewed results because I worked at three different schools. So I really had three different atmospheres to work in and things were valued at one place that would not be valued at another." She used the manner in which standards may or may not be emphasized as an example:

> You know places would tell me, for example, I'm really looking for your essential questions posted and I want to see the SPI with it. Then I would go to another school, well, they could care less if the SPI was with it.

For Rebecca these differences in how TEAM was interpreted and what was emphasized at a school was a direct result of the values of the administrator:

> It really comes down to how it's implemented, I think with the administrators. What I caught is what they (principals) valued versus what it (the TEAM rubric) actually says. I guess we tend to do that. We have things that we really like and things that we—oh who cares. I don't agree with that so I'm not going to worry about that even though that's part of it.

From Rebecca's viewpoint administrators, whether knowingly or unknowingly, made and will continue to make decisions that impact how an implementation is carried out in a school. In the next section, I detail how three different principals carried out the implementation of TEAM according to the teachers interviewed who are currently working for them. From this group, the following three different administrator approaches were highlighted: 1) the "gung-ho" administrator who implemented TEAM in a very purposeful and active manner, 2) the "we are all in this together" administrator who seemed to not be in favor of the model and sympathized with his teachers, and 3) the "disconnected" administrator who appeared to only engage in the most surface level changes in terms of TEAM.

Principal 1: The "gung-ho" administrator. Mr. Whiteside is an example of a principal who, according to the teachers working for him, made the extra effort early on in the implementation process to boost teacher understanding. He scheduled weekly meetings with his staff and lead teacher trainings to boost understanding. He called the meetings "TEAM conversations" and during these blocks of time teachers in small groups practiced scoring using the TEAM rubric, watched videos of lessons considered to be exemplary by evaluation standards, and studied each indicator. Allison explained that they "took all of August and September to really dive in deep. It was overwhelming but I felt like I knew what was coming I

guess." After two months of studying the model as a staff, they began full implementation in October.

Principal 2: The "we are all in this together" administrator. Mr. Mitchell has been a principal for many years and his standpoint on the TEAM implementation seemed to be one of sympathetic concern for his teachers. He did not implement extra trainings to boost understandings or emphasize the value of scoring but rather appeared to place importance on putting his teachers at ease. Kendra explained that her "principal's attitude in general" was supportive: "He's just very encouraging. He's the most worried about just getting the observations done, not with how well we would perform." Jessica agreed,

> Our principal has been very supportive about it. So we're lucky in that way. And [he] doesn't put too much pressure that we have to make perfect scores because we can't. That's helpful. You don't feel too much pressure from our supervisor at least.

From this stance, Mr. Mitchell may see his role as one to ease the stress of the teachers in regards to the TEAM model rather than to promote it or change teacher behavior as a result of it.

Principal 3: The "disconnected" administrator. Sally described her administrator, Mr. Edgar, as someone who had little understanding of the TEAM model and who did little to further his staff's understanding. According to Sally, teachers at her school received virtually nothing in terms of training or support, which she found to put her at a severe disadvantage in terms of comprehending the model:

> I would have liked to have had more training before we started the whole evaluation process. I feel like we just got it presented to us in a faculty meeting, maybe an hour. It was kind of like here's this section. Here's this section. It was just kind of thrown out there. I would have liked to see more sample lessons, you know, what we're looking for.

Sally explained that she and her peers attempted to understand some of the language of the model, but had difficulty:

> You know we had to go in and analyze some of the vocabulary. I know that's sad to say, you know, what does that mean, what are they asking for? And so, more training I think would have been needed and we didn't get much of that.

She posited that the lack of training increased the stress levels of the staff:

> There's still confusion there, which I feel like the whole evaluation thing has been confusion. Last year when we started it, it was what is this for and what are we talking about here. I feel like it was just thrown out there, and we didn't have a whole lot of training on it, and here it is, and I think everybody's stressed out.

Sally related her lack of understanding to her principal's own lack of understanding of the model:

> Last year when I asked questions about it, I felt like the principal didn't know either. I don't know if that's just him. I asked a question about a certain area. He didn't really know. So here we are just looking at it like I don't know. Some of the things I was trying to look up online because I didn't know what it meant. I know that sounds awful as a college student, but I don't even understand what that's asking for. If you want to get a five, you have to do what?

Additionally, Mr. Edgar seemed to be disconnected in terms of the observation and reflection processes of TEAM. For example, Sally reported that he did not stay for the entirety of the announced observation lesson, and because of this he had a less favorable opinion of her instruction:

> He was saying you didn't have this group and this group. Well, you know, you left. You missed my high-group readers. He was saying you only had two levels, and I was like there were three but you missed the last group because you left.

Sally also recounted her post conference as lacking in terms of depth and assistance:

> I'm going to be honest with you. He really didn't say much. He would wait for me to evaluate myself because we have to do the reflection and then he would just kind of write the same scores. I handed him the reflection and he just copied the scores I'd given myself down. I don't even know if he did what he was supposed to do. I felt like it was done just to say we'd done it.

From this stance, the administrator role may be pivotal in determining the efficacy of implementation. Additionally, the efforts of the principals seemed to connect back to the level of change in instruction discussed in the previous section. Sally, who works for a principal who often exhibits "disconnected" behaviors in relation to TEAM, reported no change at all in her literacy instruction as a result of the evaluation model. On the other end of the spectrum was Karen, who works for Mr. Whiteside, identified as a "gung-ho" administrator. She reported the most positive reactions to TEAM of the interviewees.

Joanne stated that principal feedback is one of the most important things in bettering her methods of instruction and that that is true regardless of whether the feedback cites the TEAM rubric:

> I don't spend the time to try to understand it [the TEAM rubric]. I just teach the way I know how to teach. I take the feedback that the principals give me and I use it to better the next lesson.

This chart illustrates the relationship between the actions of the administrator and individual teacher change.

Table 7: Type of Administrator and Teacher Change

Teacher	District	School	Type of Administrator	Teacher Change
Laura	Carson County	Whiteside Elementary	Gung-ho	Forced Change; Some real change; some surface level change
Karen	Carson County	Whiteside Elementary	Gung-ho	Internalized model; high level of change
Allison	Carson County	Whiteside Elementary	Gung-ho	Forced Change
Sally	Hampton County	Gary Elementary	Disconnected	No Change
Rebecca	Booker County	Newton Elementary	We are all in this together.	Some real change; Refusal to make surface level change
Joanne	Booker County	Newton Elementary	We are all in this together.	Some change; some surface level change
Kendra	Booker County	Lindley Elementary	We are all in this together.	Some real change; mostly surface level change
Jessica	Booker County	Lindley Elementary	We are all in this together.	Some real change; refusal to make surface level change

Subjectivity

In the next few sections I continue to answer the research question, "What are the common elements in teachers' experiences with the TEAM evaluation model?" In relationship to the administrator role, several participants discussed the notion of subjectivity as problematic in terms of the TEAM evaluation model. While in many cases the school principal was the direct focus of commentary on subjectivity, because teachers could be evaluated by various individuals, including lead teachers and central office staff, the conversation extended past the administrator role. In this section, I highlight some teacher perceptions about the subjectivity of the model, instances they cited that they believed to be examples of it, and, lastly, the relationship they saw between subjectivity and job security.

Laura may view the TEAM model through a slightly different lens than that of her peers because she is the only participant in the study who served in both the role of classroom teacher as well as lead teacher and, thus, evaluator. Laura noted that her position as an evaluator has given her an even clearer picture of the subjectivity of the model: "Even now that I'm a lead teacher, and I'm also observing, I see a lot of the subjectivity as we're evaluating one person as a group and talking together." She cited a particular incident in which she, her principal, a central office employee, another lead teacher, and the curriculum facilitator were all observing one teacher and disagreed on how he should be scored:

> For example, I thought one person's academic feedback was really good. They gave oral feedback. Like, that wasn't good. How can we fix that? Some people saw that that wasn't good academic feedback. I saw it as good because sometimes I do that with my kids, and all his kids love the snot out of him. So obviously they trust him enough for him to say that. So, I'm sitting there thinking, well, hmmm, it could go either way on this.

Laura's personal knowledge of this teacher's abilities and interactions with his students impacted the way she scored him. While other members of the scoring team found his comments to be somewhat inappropriate, she regarded his manner of presenting academic feedback as both necessary and worthwhile. Laura argued that prior knowledge (or lack thereof) of an individual always affects the score he or she receives:

> Let's say Nancy Ross, the head of all this TEAM evaluation stuff comes in and watches me. She would probably give me a lower score than Mr. Whiteside would because he knows my background with these kids. He knows how hard I work with them. He knows that I'm at this point with these kids. They're a lower class. Subjective. That's why no one wants anyone from the outside world to come in and have to evaluate us because of background knowledge and also how you feel about the teacher. Everyone's going to have people they like better than others. You're going to sit in and evaluate someone that you've probably already pigeonholed into a certain type of teacher. You can't go in there without some baggage.

Kendra also believed that scores on the observation component of the evaluation are impacted by subjectivity, by the beliefs and perceptions of the evaluator, as well as by the room for interpretation on the rubric. Kendra stated that self-scores and individual evaluator scores may vary from person to person:

> I did do a self-scoring, which, I read the rubric and, as clear as it is, there's still so much room for interpretation. And I know from talking to other teachers how he (principal) interprets things is different than how the woman who observed me last year interprets things and what they see as more important is different as well. If I hit these three out of

> five, he's going to give me a four, bump me up to a four. But it might be the other two bullets that I miss that she harps on.

While a few interviewees saw the subjectivity of the rubric as an indication of differences of opinion amongst evaluators, openness of the rubric, or background knowledge of the teacher, Rebecca relayed an incident in which a teacher was given a high score, according to Rebecca, because she was nearing retirement:

> I saw where a younger principal didn't want to upset someone about to retire, so they just gave them all fives. And I'm thinking, wait a minute. You told me you wanted to see this and this. And I can walk in the room and know it's not there because those are things you would see posted. If I didn't do that I would have got a one, you know. So there's some—I don't want to say favoritism…

Most of the teachers who touched on the issue of subjectivity sympathized with the evaluators and how difficult it is to be unbiased. Rebecca explained,

> To be unbiased, you know to come to that rubric without an opinion and to sit down and evaluate someone else without an opinion of the person, of their past, maybe of their personality. It would be hard to do that but at the same time when you are in that role that's what you've got to do. It's kind of like being the judge and the jury. You can't go into a trial and already have an opinion.

While some understanding was expressed, several teachers took issue with the notion that job security was tied to what they saw as a subjective scoring method. Both Kendra and Laura said that if the scores are going to have serious implications for job security, then they should not be subjective. Kendra explained,

> I don't know if they were trying to make it more open-ended so we could get our own ideas or what. With that kind of stuff, if you're going to score me on it, I want to know exactly how you would like me to do it.

Laura concurred, noting that she understands the presence of subjectivity but takes serious issue with its impact on job security: "My thing is, alright, that's fine, it can be a little subjective. Some things are. But if it's my job on the line, I don't want it to be subjective at all."

In the next section, I explore teacher's reactions to the issue of scoring more fully particularly in terms of the quantitative component of the model.

Scoring

As detailed in chapter two, each teacher's evaluation score is based fifty percent on qualitative methods, a compilation of classroom observation scores, as well as fifty percent on quantitative methods. The quantitative scores are a combination of school wide data and "other" scores, which will be discussed more specifically here. I address the following: 1) teachers' understanding of scoring methods, 2) how scores are used at differing schools, 3) teachers' reactions to the use of school wide data, and 4) what different scoring methods are currently being employed and possible implications of these methods. In this section I work to answer the primary research question, "What are the common elements in teachers' experiences with the TEAM evaluation model?" by examining the patterns in teachers' reactions and understanding of scoring. I also answer the sub-question, "How does the context of teachers work complicate their perceptions of the influence of TEAM on their work?" in part through discussion of how scores may be used differently in different contexts.

All four schools used school-wide data as thirty-five percent of the quantitative score. The remaining fifteen percent of the quantitative component varied from school to school and

teacher to teacher, with some using school wide data a second time, some using professional development scores, and some using classroom data such as DIBELS, running records, or STAR testing.

When asked how the quantitative portion of their evaluation score was computed, the majority of teachers just weren't sure. In regards to the thirty-five percent of the quantitative score, which is often a measure of school-wide data, Allison said, "I don't exactly know. There's a lot of words that get thrown around so I just usually nod my head go 'okay.' I know it is achievement and part of it is gains. Right? Does that sound right?" She had no more understanding of the remaining fifteen percent of the quantitative score, which may be computed based on various scoring components:

> I don't know. I know that they put all these choices in front of you and then in a nice way you figure out which one you really should choose. I just circled that one. Honestly, I'm not really sure what your choices are.

Kendra's reaction was similar: "I don't know. [My principal] set it all up. Even then I said I don't even understand that other part." While Jessica knew that thirty-five percent of her quantitative score was based on school-wide data, she was unclear as to what comprised the additional fifteen percent: "I can't remember. That sounds bad. I don't remember what it was. I think we had a choice individually. I think I selected to use my professional development score." For Kendra, her confusion related to her belief that she only had ownership over the qualitative portion of the score. "I know my observation [score] because that's what I can control."

While many of the participants expressed confusion over exactly how the quantitative scoring portion was applied, most had strong opinions about the use of school wide data for their score. Because teachers in grades kindergarten through second grade do not currently have a

statewide test on which to base personal scores, at least some percentage of the quantitative component of their evaluation score is based on a valued added school wide average. Amongst the teachers interviewed, there were two distinct reactions to the use of this average: 1) those that saw the validity in its use and the role they played in upper-grade achievement and 2) those who felt that the use of school wide-data was a poor representation of them as a teacher.

Karen, Rebecca, and Laura all agreed that lower-grade teachers are responsible in part for how upper-grade students score and that they should be held accountable for those scores. Karen stated, "I feel like it's a good thing to take on all of the school's scores because you should all be held accountable." Rebecca agreed as well, even though she acknowledged that as a new teacher at the school she will not actually have instructed the upper grade students for a couple more years:

> You know technically speaking the next two years if I stay here you could technically say it's unfair to me because I didn't teach those students that are in the third, fourth, and fifth grades. But, you know in two years this first grade class will be taking the TCAP for the first time so I have an investment in them. I think that's fair. I've laid the foundation. If I've not taught that first grader to read and they go and take the TCAP in third grade, who is that looking bad on? That's my part—reading. And that's huge. So yeah, I feel like that's fair.

Laura also relayed that she has had positive reactions to using the school-wide data; however, she admitted that she might have felt differently if it had negatively affected her evaluation score:

> I was completely okay with that. If I was at a different school, I might not have been. I'll be honest about that. [It's] just an amazing place where everyone works together. If we

> weren't all working together and we weren't all accountable …. I'm accountable for how they read in second grade. I'm accountable for what they know when they get to third grade. You know? First grade is not tested on any science and social studies but we still teach it because it builds and we want to be responsible for that. I know all of them are working really hard and so I truly don't have a problem. If I was at a different school, maybe I would.

For Laura, her belief in her school's sense of community played into her feelings about the use of school-wide data. She appeared to have trust in what the other teachers were doing; therefore, she had positive feelings about the scoring method.

On the other end of the spectrum are those individuals who took issue with the use of school-wide data. Teachers who fell within this category believed for one reason or another that the test was not a fair representation of their abilities as teachers, and because of this it should not be included as part of their evaluation scores.

Sally is an example of a teacher who felt strongly that using school-wide data as a component on her TEAM evaluation score was unfair. She contended that the school wide data created an inaccurate view of her teaching ability. Sally explained that the previous year she had received a score of four out of five on her observation as well as a four out of five on the fifteen percent of the quantitative component that was based on her current student scores on the STAR Literacy test. However, because her score for school-wide data was a one, she ended up with a total TEAM evaluation score of two out of five. She described her reaction:

> I about fell on the floor. I was very upset with it. I didn't feel like it was fair, you know when you look at what my kids did for the year and I've got a four but I've got to sit here with a two total because of the school wide.

Sally especially took issue with the fact that most of the upper-grade teachers received higher scores because their thirty-five percent was based off their students' scores for the current year. Primary grade teachers' thirty-five percent was a compilation of three years of school-wide data. According to Sally, because two years of that data showed very low scores, primary-grade teachers had much lower scores on their evaluations:

> They based this off the whole school for three years. Our school has been low for two of those years so that brought us down. That killed us while our upper grade teachers who have a year … they're sitting there with fours and threes and our kids all got ones when our kids weren't even the ones who took the test.

For Sally, the implications of this formula were unfair to primary-grade teachers.

Kendra also viewed the use of school-wide data as unfair. This view was held even though her TEAM evaluation score was raised by the inclusion of school-wide data. She, like Suzanne, believed that this presents a skewed impression of her abilities as a teacher:

> Because Lindley [her school] tends to have high scores I feel like I'm cheating because really if I looked at my overall scores online, like my observation scores, they were good. It was almost a four. It was like a 3.9 was my average, which … I was okay with that, but then somehow I got a five overall because of just the school, because I had the school's scores twice. And so in my head, I'm not a five. I'm a three … a high three.

Several of the interviewees also focused on the issues of transient populations and pressures on upper-grade teachers when considering whether or not it is appropriate to use school-wide data in evaluation scores. Joanne and Jessica both argued its use was not as valid at schools with transient populations because lower-grade teachers may or may not have taught

many of the students in upper-grade classrooms; as a result they may not feel that they have a stake in those students' achievement. Joanne noted,

> We have such a transient population, there may have been you know fifteen kids out of fifty that they actually taught in kindergarten or first grade or second grade. The others they never had their hands on them. They never had the chance to teach them and get them prepared for that.

Several of the participants also addressed the amount of pressure that has been placed upon upper-grade teachers since school wide data became of part of each teacher's evaluation score. Joanne had sympathy for upper-grade teachers because of her experience teaching third grade the previous year: "I know how it feels to be an upper grade teacher and everybody else looking at you saying, you made my evaluation scores go down because your kids didn't pass the TCAP test." Kendra and Jessica reported that the pressure is particularly high for upper-grade teachers at their school for two reasons. First, their school is comprised of kindergarten through fourth grade so scores are based on growth from third to fourth grade; thus, school wide data is based on the test scores of the three fourth grade teachers. Kendra argued, "I think it's just unfair on the fourth grade teachers especially since we've changed our schools and now it's all on fourth grade teachers. Three teachers for the whole entire school. I just think it's a little unfair." Jessica also commented that at a high achieving school such as hers, it may be difficult to show growth:

> I had actually taught third grade the year before so my kids' scores were what was being looked at as compared to fourth grade and I had several kids make perfect scores. They'll be lucky to make a perfect score again. And if they do, it's still nothing even though they made a perfect score, which is amazing. I understand wanting to see growth, but when

> you have really high children, which we do at this school, a lot of our kids are in the upper ninetieth percentile, they can't show but so much growth because they're starting at such as high point. So that's going to be a struggle for our school as far as the scores.

Lastly, some of the teachers who saw using school wide data in conjunction with teacher evaluation as problematic believed that its use was giving too much power to one test on one day of the year. Joanne concluded that the use of school wide data may spawn an extreme overemphasis on the test, particularly in upper-grade classrooms:

> I mean the first month of school I heard "TCAP" no less than twenty times from upper-grade teachers the first month of school. And so there's this push to get them to know what they need to instead of in the lower grades we're not worrying about a test. All we're trying to get them to do is read, get them to know strategies to read a word. And that whole focus is forgotten in the upper grades. It's on test, test, test. Can you answer this type of question? What strategies do you need to do to eliminate four answer choices? It becomes about test-taking strategies instead of reading strategies, instead of what do I need to do to become a better reader.

Perhaps as a result of some of the very issues touched on by the participants, several districts have begun to reconsider how primary grade teachers should be scored on the quantitative portion of the TEAM rubric. Most districts are considering the use of or a combination of two possibilities: 1) adding a cumulative test for primary grade students and/or 2) using classroom data.

Two of the schools in the study began implementing the SAT-10 test in first and second grades, and they planned to use these scores as a portion of the TEAM score. Some of the teachers believed that using such a measure could, in some ways, be more indicative of their

teaching performance. For example, Allison liked that it would mean a focus on the progress of her students: “I’m glad it’s off of gains and not based off of achievement. That’ll be helpful.” However, the teachers also noted various issues they saw arising as a result of implementing a test in primary classrooms.

One issue that Laura and Allison found particularly disconcerting was whether or not their students had the stamina to adequately demonstrate their abilities on such a test. Allison described how the SAT-10 pre-test process went for her students in the fall:

> Just the testing process in September for a first grader is awful. You go thirty minutes into it, and they’re just going to start putting a C for every answer. They don’t care. It’s a little bit frustrating for that to be what it [the evaluation score] is to be based off of.

Laura agreed, noting that while she liked the content of the test, it was too long for a first grader:

> I like the test. It was a good accurate assessment except it gave about sixty questions when it could have given literally twenty. They needed to take into account where a first grader is at the beginning of the year.

Additionally, both teachers felt that they were blindsided by the implementation of the SAT-10 and therefore did not know how to adequately prepare their students. Laura explained,

> I don’t like that I’d never seen it, heard of it. I had no idea what was about to be put in our hands. I had no idea what the layout would look like, how long it would take, what pacing I needed to keep. I didn’t like that.

Allison also believed they should have been better prepared for the implementation of the SAT-10:

> I just wish they would have given us a year to give this test. And then maybe this year would have let us have it be school-wide, and then next year let it be our kids once we

> have a bearing on what we're dealing with. Because they [upper grade teachers] do TCAP prep for weeks and weeks to get them ready. We don't know what to prepare our kids for. It feels a little bit unfair to me.

Some teachers explained that their district or school was considering using classroom data as a primary component of the quantitative portion of their TEAM evaluation score. Some teachers would be given the choice of which data they wanted to apply amongst a certain selection of data. This choice usually included DIBELs data, STAR Literacy data, and, in a few cases, running records data. Opinions on what was most appropriate varied from teacher to teacher. Sally believed that the STAR literacy data was an accurate reflection of the progress in reading she had made with her students over the course of a year. She said, "That's geared more towards our kids and what we did." She also considered the use of running records but questioned whether or not they would appropriately fit into the equation required by TEAM. Both Rebecca and Joanne argued that DIBELS is the best indicator of their performance as literacy teachers and therefore is a good choice for the quantitative component of their evaluation score. Rebecca described DIBELS as covering "the whole gamut. We're covering the whole reading spectrum there." Joanne agreed and saw DIBELS as measuring her students' "ability to read." Kendra, on the other hand, questioned the accuracy of DIBELS as an assessment tool, which, she noted, will be used as part of her evaluation score in the upcoming year:

> Some kids have trouble with the isolated phonics that DIBELS tends to do. They have that nonsense word test and some kids are really great with that and others do what good readers do and they're trying to listen to themselves say the word and it doesn't make any sense to them so they don't know what they're supposed to be saying. That's what I tell my kids to do when they're sounding out a word is to listen to themselves and to put

> together that with what makes sense in the sentence. That particular test doesn't allow them to do that whereas a running record does as they sound out words because there's a context for the words.

In fact, Kendra posited that running records would be the best measurement and the assessment most closely aligned with her performance:

> I would prefer to use running record scores because to me that is more clearly defined. They start at a level 3, and they got to a level 18. I did my job. They've made progress. I do my reading groups on the running records. I talk to parents about the running record level. I do my report cards based on running record level.

According to the interviewees, most districts and schools had not decided how primary-grade teachers will be scored. Many district officials are discussing the idea of moving away from the heavy use of school wide data, but final decisions had not yet been made. As Laura noted, "K and 1 are still kind of in this floating place."

In the next section I build on the role of the district and state by examining how the participants view these entities.

Issues of Power

In keeping with a critical framework, in this section I continue to consider the patterns amongst teachers' perceptions of TEAM by touching on the following issues relating to power structures that were noted by the teachers interviewed: 1) teachers' understanding (or lack thereof) in terms of who makes the decisions that they are asked to carry out, 2) disconnect between classroom teachers and those in power, 3) variations in central office participation at the school level, and 4) comfort level with voicing opinions about implementation.

Several teachers expressed a lack of understanding of exactly who were the architects of the TEAM model. Throughout the interviews the creators of TEAM were referred to with a generic "they." When questioned as to whom "they" refers, most teachers were non-committal, often vaguely saying, "the state?" Sally's reply was as follows: "The state? I guess. Whoever designed this thing. I don't even know who designed this thing. You know I don't even know who made this and I wonder if they were ever educators either." Kendra also noted that TEAM was a production of the state, but she too did not give specifics as to who comprised the state. Additionally, Kendra discussed why she believed that the state felt that such a measure was needed and why it may have been viewed negatively by practicing teachers:

> In their [the state's] effort to set clear boundaries, clear consequences, it just came across too harsh maybe. Part of it I think probably coupled with the education association losing their collective bargaining rights, it just seemed like a lot of negative things, spotlights being put on teachers. It kind of feels like the better you do the more you're penalized in a way. I don't know if it's the general school culture and teachers just feeling put upon all the time or what. I feel at least it came down from the state.

From Kendra's standpoint, TEAM was implemented by the state in an effort to set clearer expectations for teachers. However, she believed that many teachers viewed it more as an attack than as guidance.

Additionally, several of the participants saw a disconnect between "they" and "me." In other words, they expressed a tension between their beliefs based on their roles as classroom teachers with the beliefs of those in power to make decisions. For Rebecca and Joanne, individuals in these positions had little understanding of day-to-day classroom life and instruction; rather, they focused on outcomes and results. Rebecca explained,

> They want to see your results but the only time they care is at the end of the year and they want to see if you've shown growth. Show me your AYP. They don't really care that you taught that child half the alphabet because when they came to you they had no print concepts. They think we read from right to left. They don't care. They don't care about that …. There's no room for all that. The personal part is gone out of it, which I guess it should be. I guess that's what a rubric is for.

Joanne also saw a tension between teacher goals and state expectations. She concluded that she puts more weight on her own opinion of her performance than on the opinion of the state. Joanne, like Rebecca, connected this to the state's lack of understanding of her classroom and her students' progress: "They may not perform the way the state thinks they should, but we know we've done the best we can. We've exhausted all possibilities."

For Jessica, mandates for classroom teachers were not helpful or even appropriate, but mandates were a reality to be accepted and dealt with. She believed that it was just the way of public education: "There's always some type of evaluation. It's not necessarily kid-centered or student-centered, but it's just the way things are."

While many of the participants viewed the state in a negative or disconnected manner, opinions tended to be more positive when teachers considered leaders in their district or central office. Nonetheless, the workings of and manner of support from district leaders varied greatly from district to district.

Kendra spoke well of her district and actually referred to her administrators as a mediating force between classroom teachers and the state: "I feel like our district itself was very encouraging, not just encouraging, they calmed us [teachers] down a lot." She believed that district leaders assisted in helping teachers process the new evaluation model and supported

them, as a result she viewed TEAM in a more positive manner than many of the other teachers. She also viewed her district leaders as protecting her from state mandates: "I feel like my school district will still try to take care of me. And to know that I am a good teacher."

Rebecca expressed how support and contact from those in central office may vary greatly from district to district. She described how infrequently she saw district leaders in the previous school system in which she was employed and how much more positive her experience at Newton Elementary had been:

> The only time I saw people from the round building was if they were coming to give me a score. Here that's a totally different story. They led my new teacher training. They taught me. It was great. It was a personal relationship.

Joanne maintained that there is a possibility for compromise between practicing teachers and those in positions of power: "I think they can come up with some sort of common ground so there's not … the teachers aren't feeling like they didn't get a fair shot." In the meantime, however, some teachers appeared to feel hesitant about speaking out against current mandates. Near the end of her interview, Rebecca looked towards the door nervously, as if she was afraid that someone may have heard her comments: "I'm telling you, I'm going to end up …. I'm going to have to hitch a ride out with you. Sorry. I'm playing. We're at the end of the hall. We're good."

Final Issues with the TEAM Model and How the Model Can Be Improved

Problems with time and the checklist that isn't a checklist. Several of the participants mentioned that TEAM trainers were explicit when they argued that the elements of the TEAM rubric should not be viewed as a simple checklist. However, some teachers believed that because

evaluators checked off whether or not a teacher demonstrated a particular behavior, it was very much a checklist. Karen noted,

> You know we've been told don't look at it like it's a checklist, but trying to make sure that you're thinking about all of those pieces that need to go in there … I feel like, well, all the meetings we go to—it's not a checklist. But try to make sure you have the differentiation in there and that you have the assessment in there. I kind of see it as a checklist.

Laura agreed, "They always use this phrase that it's not a checklist, but it very much was when it was presented."

Many of the participants saw the checklist as particularly problematic in relationship to time constraints. According to the teachers, an instructor is required to demonstrate all of the required rubric elements but the elements must all be demonstrated within a particular time frame. If all elements are not demonstrated, then a teacher's score is lowered. If teachers go over the allotted time for an observed lesson, they will also be penalized in terms of scoring. Jessica explained,

> We counted every bullet. They're looking for twelve main things in the lesson, but each part has about eight or nine subparts to it. That's a lot to show in one lesson, as far as the time constraints, too. If you go over the time you are penalized.

Many argued that it is unrealistic to address all the elements on the rubric in every lesson. Jessica maintained,

> It's so much at one time. You may do that in your lesson the next day to coincide. You may hit all those components in a week, but they're only seeing one little portion of the

> unit. They can't possibly see everything. You can't possibly show everything in that one little snapshot.

Allison was in agreement, explaining that she does engage in all the teaching practices required by TEAM in a single lesson but that they may occur over the course of the day, week, or unit:

> We have five days to cover it so we spend little blips of time doing different things that over the course of the week are going to hit everything TEAM asks for. Over the course of a week I think everything TEAM asks for we cover just naturally, but to fit it all in one lesson isn't really how I teach.

For some of the teachers time restraints can also cause tensions between scoring well on the evaluation rubric and attending to students' learning needs. Jessica, Kendra, and Karen all described feeling pulled between addressing all the required components while simultaneously engaging in best practices for student learning. Jessica said she felt tempted to end her lesson earlier than she might otherwise despite productive learning on the part of her students:

> If you go over, they penalize you on the pacing portion of the rubric. Sometimes students are really talking and learning very well and you don't want to stop that, but at the same time you're watching the clock the whole time, too.

Kendra also discussed how the number of requirements on the rubric was in opposition to what she viewed as effective teaching:

> In an effort to make sure I was hitting everything, I tried to jam pack way too much in and I feel it just kind of flooded their heads a little bit instead of getting them time to really digest what we were talking about. I had to get my two problem solvings in. I had to get this kind of thinking in. And I had all these different kinds of activities. Sometimes

> I just feel like it's just too much. And that first grade twenty minute lesson that you have their attention for. Then I feel rushed. Then it doesn't help them at all.

For several of the teachers, these tensions between the expectations set forth on the rubric and their teaching may create a misrepresentation of them as an instructional leader. They argued that evaluators could see them engaging in all the practices required by TEAM but not necessarily during a single lesson. Kendra said, "This is supposed to be a true reflection of my teaching. I don't feel like it's doing a good job of that."

The TEAM rubric should be more level-specific. Several of the participants argued that because TEAM is created for kindergarten through twelfth grade students, it is too generic and would be more beneficial if it was customized for particular populations. For Rebecca this would mean customizing it for individual school populations:

> One [change I would suggest is that there] be room for the principal to customize it to our population and our personality as a community because there is none of that and that's where I think it is so generic and so almost it's too black and white, too vague, too generic, general. You know, the teacher does this all of the time. The teacher does this some of the time. What does that mean? The teacher does this three times. I don't know. What? Who cares? Either she does or she doesn't.

Others believed that the rubric should be customized for specific age levels of students. According to this view, what a teacher should be doing at the first grade level is very different than what a high school teacher should be doing, and thus these differences should be reflected in the rubric. Kendra viewed the single rubric for all levels as problematic: "It's frustrating with the rubric because it's one rubric for the entire span of grades. It's just not a one-size-fits-all kind

of thing." Both Sally and Jessica had similar feelings and advocated for tailoring a rubric based on grade level. Sally explained,

> I think I would understand it better if they were gearing it more to, okay, this is what K-1 and 2 teachers should be doing, and this is what it should look like, and this is 3, 4, 5. Instead of all of us at the same level.

Jessica, too, expressed her belief about the need for differentiation on the rubric:

> It's a rubric for K-12. I feel like it should be adapted to at least primary and secondary and high school. Some of the things, too, are geared more toward elementary. I'm thinking of my friends who teach high school. How do they do that with high school students? It's not really appropriate to do. At least tweak it a little to fit in more with what's developmentally appropriate for those children.

How TEAM implementation could have gone smoother. As addressed previously, most of the teachers initially had negative reactions to the TEAM implementation process. Laura reasoned that this was a direct result of the speed with which the evaluation was implemented, arguing that the state did not have a handle on TEAM before they required schools to start implementation:

> At first it was very difficult because I don't think that they knew exactly what they wanted. So they're kind of putting the cart before the horse really. They put it out there before it was ready, and they slapped it on us without enough administrators. They just had no idea of the capacity it would take.

Laura postulated that this contributed to a breakdown in communication:

> Anytime you just slap something on there that no one really truly knows what's going on and the communication is down, no one's going to be okay with it. I think they got really lucky that teachers are still trying to hang in there.

Laura, Sally, and Jessica all noted that more training and support would have been beneficial during the implementation process. Laura rationalized that the lead teacher training would have been beneficial for all teachers to better understand TEAM:

> I think the lead teacher training that I took should have been given to everyone. Everyone would have had some basic knowledge of what they're going in to and not felt like someone was coming to just steal their job out from under them.

Jessica believed that more examples on how to perform expected instruction would have been helpful as well:

> That's what I've struggled with because I don't really know what that looks like in language arts. I've just collaborated with my fellow teachers and asked them what they think. I haven't been given from central office or anyone doing the observations, any kind of model of what that looks like in reading.

Laura also argued that a slower implementation process might have made the transition smoother. She suggested implementing just one of the three rubrics (instructional, planning, environment): "I think what would have been better is for them to maybe instead of adding three rubrics, there's instructional, planning, environment, maybe just start with planning and let us look at instruction and environment." Generally, most participants seemed to feel that the state jumped in to the process too quickly before individuals at all levels (state, district, and school) had an appropriate level of understanding of the new model.

Giving teachers a say. According to the teachers I interviewed, the model could be greatly improved by incorporating teacher input, which several participants argued was lacking in the creation of the TEAM model. Sally believed that the job of evaluation creation would be best suited for "exemplary teachers, your teachers that are doing it right." She believed that giving teachers a say would result in an improved model, one that more fully takes into account individual students, individualized instruction, differentiated instruction, and student-teacher interaction.

Considering Context: It's Complicated

In the previous sections of this chapter, I have primarily focused on answering the primary research question, "What are the common elements in teachers' experiences with the TEAM evaluation model?" I discussed the following commonalities or themes that I created as a result of participants' commentary: teachers want to be held accountable, initial reactions to TEAM, change, administrator role, subjectivity, scoring, issues of power, general issues with TEAM, and how participants believe the model can be improved. In earlier sections of this chapter I began to answer the sub-question, "How does the context of teachers work complicate their perceptions of the influence of TEAM on literacy instruction?" In this final section, I will continue to focus on the question of context by considering how context differs for each participating teacher and how these differing elements of context may or may not be impactful on outcomes of change.

In determining specific contexts for each individual teacher I considered the following factors: 1) years of experience, 2) percentages of economically disadvantaged students in the school, 3) level of diversity in the school, 4) TEAM composite score, 5) measures used for 35% of quantitative score, 6) measures used for 15% of quantitative score, 6) school value-added

score, and 7) district policy on student evaluation measures used. Primarily, I related these criteria of context to the level of teacher change. However, I also considered how these factors might or might now influence teacher reaction to the use of school wide data and preferred scoring methods. The following chart illustrates the criteria stated above for each teacher and the corresponding school and district.

Table 8: Issues of Context; Part 1

District	School	Teacher	Experience in years	Eco Dis.	Diversity	TEAM Score	35%	15%	Value Added Score
Carson County	Whiteside	Laura	4	74%	Diverse	5	School wide avg.	School wide avg.	1.2
		Karen	0	74%	Diverse	N/A	School wide avg.	School wide avg.	1.2
		Allison	4	74%	Diverse	5	School wide avg.	School wide avg.	1.2
Hampton	Gary	Sally	12	53%	Little Diversity	2	School wide avg.	STAR Literacy	-2
Booker	Newton	Rebecca	2 (public)	48%	Highly Diverse	3	School wide avg.	DIBELS	-1.7
		Joanne	12	48%	Highly Diverse	3	School wide avg.	DIBELS	-1.7
	Lindley	Kendra	4	20%	Little Diversity	5	School wide avg.	School wide avg.	3
		Jessica	4	20%	Little Diversity	5	School wide avg.	School wide avg.	3

Table 8: Issues of Context; Part 2

District	School	Teacher	District Policy	Reaction to use of School Wide Data/ preferred scoring method	Connection between classroom observation, TEAM observations, beliefs	Level of Change
Carson County	Whiteside	Laura	Undecided; considering SAT-10	Approved	Highly connected	Forced change; some real change, some surface level change
		Karen	Undecided; considering SAT-10	Approved	Some connection	Internalized model; high level of change
		Allison	Undecided; considering SAT-10	Mixed/Bench mark Tests	Little connection	Forced change
Hampton County	Gary	Sally	District allows schools to choose 15% measurement.	Disapproved/ STAR Literacy Tests	Some connection	No change
Booker County	Newton	Rebecca	District allows schools to choose 15% measurement.	Approved/ DIBELS	Highly connected	Some real change; refusal to make surface level change
		Joanne	District allows schools to choose 15% measurement.	Mostly disapproved/ DIBELS	Some connection	Some real change; some surface level change
	Lindley	Kendra	District allows schools to choose 15% measurement.	Disapproved/ Running Records	Some connection	Some real change; mostly surface level change
		Jessica	District allows schools to choose 15% measurement.	Mixed/ Measurement from beginning to end of year	Highly connected	Some real change; refusal to make surface level change

In this section, I discuss how the above criteria relating to context may or may not have resulted in patterns which impacted change in the instructional practices of the participants.

Years of Experience: The participants in the study had levels of experience which ranged from 0 to 12 years. While it would be difficult to generalize years of experience as a prevalent factor in level of change, it is interesting to note levels of change on the two ends of the spectrum. Karen, who had the least amount of experience reported the highest level of acceptance of the model. Sally, in contrast, was one of the two teachers with the most experience and reported no change at all in her instruction. The rest of the participants fell somewhere in the middle of the spectrum with years of experience ranging from 2 to 12 years and varying levels of surface level and real change reported.

Percentages of Economically Disadvantaged and Diverse Students in the School: In this study the percentages of economically disadvantaged and diverse students in the school seemed to have little impact on teachers' acceptance of the model and corresponding levels of instructional change. For example, teachers at Whiteside Elementary had a high percentage of economically disadvantaged students and the most positive reactions to the model of all teachers studied. Over half of Gary Elementary's student population is economically disadvantaged, but teacher change and acceptance was lowest here. Newton Elementary had the most diverse student population and Lindley Elementary had the least diverse, but the reported levels of change were similar for both groups.

Scoring Relationships: There was some relationship between individual TEAM composite scores and instructional change. For example, two of the teachers at Whiteside Elementary, Laura and Allison, who reported internalization of TEAM strategies after a period of forced change also

received high composite scores. In contrast, Sally, who reported no instructional change at all received a 2, the lowest composite score of all the teachers. This relationship does not necessarily hold true across the board, however. For example, Kendra and Jessica described only small to moderate changes to their instruction, but both received high TEAM composite scores.

There are also apparent relationships between the following: TEAM composite scores, school value added scores, choice of measurement for the 15% component of the qualitative data, and teacher reaction to the use of school wide data and their preferred scoring methods. Because 35% of each teacher's TEAM composite score was based on school value added scores, teachers who worked at school with high value added scores had higher composite scores. For example, Lindley Elementary received a value added score of +3 and Kendra and Jessica received composite scores of 5. The four teachers in the study with the highest composite scores also teach at schools with the highest value added scores. The opposite also holds true. Teachers with the lowest composite scores also worked at schools with the lowest value added scores. For example, Sally received a 2 for her composite score and her school, Hampton Elementary, received a -2 value added score.

The schools in the study that received high value added scores opted to use that score twice, once for the 35% component of the quantitative portion of the evaluation score and once for the 15% component of the quantitative portion of the evaluation score. Schools such as Hampton and Booker that received lower value added scores opted to only use the school wide data for the 35% component and replace the 15% component with some other form of measurement. For example, Hampton used STAR Literacy Tests and Newton used DIBELS assessment scores.

For the most part, there was also a relationship between the value added score and teachers reaction to the use of school wide data. Laura and Karen both work at a school who received high value added scores and both approved of the use of school wide data. Sally, on the other hand, works at a school that received very low scores and strongly disapproved of the use of school wide data. Exceptions to this were Kendra and Rebecca. Kendra's school, Lindley Elementary, received a high score and yet she still disagreed with the use of school wide data. Rebecca's school, Newton Elementary, received low scores and yet she still agreed with the use of school wide data.

Summary of Chapter

In this chapter I have presented my analysis of eight first grade teachers' perceptions of the new Tennessee Evaluation Acceleration Model. In my analysis, I have attempted to answer the following research questions:

1) What are the common elements in teachers' experiences with the TEAM evaluation model?
    a. What are the common elements in teachers' perceptions of the influence of TEAM on literacy instruction?
    b. How does the context of teachers work complicate their perceptions of the influence of TEAM on their work?

Questions 1 and a. were answered through discussion of the following themes: teachers want accountability, initial reaction to TEAM, change, subjectivity, administrator role, issues of power, scoring, and possible improvements to the model. Question b. was addressed through the themes of administrator role and scoring as well as comparisons of contexts amongst teachers. In the final chapter I will provide discussion and implications for the previous analysis.

Chapter 5. Discussion and Implications

Chapter Introduction

In this study, I examined the experiences of eight first grade teachers as a new teacher evaluation model was implemented in their state. Through the means of interview and classroom observation, I worked to understand how each teacher perceived the implementation of this model. The findings of this study support several notions that have been discussed in previous literature. Additionally, it provides possible insights for future policy directives as well as specific implications for Tennessee's current TEAM evaluation system.

Discussion

In this section I highlight the points of agreement between previous research studies and my own. I focus on the following commonalities: the need for an evaluation system, the problematic nature of top-down policy and relationships to teacher beliefs and context, the influence of those in decision-making roles, and the messiness of applying scoring data.

The need for an evaluation system. The need for teacher accountability has been a rallying cry particularly in the public realm in recent years. To varying degrees, many researchers have agreed that evaluation is a necessary means to further the professionalization of teaching and increase the rigor of our schools (Shulman, 1987). While some in the public arena believe that teachers seek to avoid the critical eye, this study demonstrated that many teachers were in favor of evaluations. This indicated that teachers often welcomed critique when they could use it as a springboard to improve their instructional methods. Teachers who were in agreement with this need also appeared to want to raise the quality standards of their profession by addressing the issues of problematic teachers. Lastly, they wanted to demonstrate their own expertise because of their pride in their work.

The problematic nature of top-down policy and the relationship to teacher beliefs and context. This study also aligned with much of the research in regards to the messy nature of policy mandates in the schools. McGill-Franzen (2000) noted that the complex nature of teaching often results in unpredictable policy outcomes. Likewise, Spillane (1999) found that the amount of actual change in classroom practice as a result of policy mandates was variable and that instructional changes ranged from nonexistent to extreme. This study indicated similar results as some participants demonstrated deep-level changes in their teaching, others showed only surface-level change, and still others demonstrated no change at all in their classroom practice.

The reasons for these varying levels of change could be linked to Coburn's (2001) theory of sensemaking. She argued that sensemaking is a natural process in which teachers process the messages from policy makers and then attend to or disregard them in their instructional practice. Likewise, the teachers in this study employed their own processes of sensemaking, and in doing so elements of the TEAM model were put into practice fully or partially, or these elements were disregarded completely as teachers attempted to translate them into workable classroom resources (Coburn, 2001). Just as in Coburn's (2001) study, variability in implementation of TEAM methods indicated that policy makers cannot expect a basic input-output equation when it comes to mandates. In other words, although those in decision making roles expected the TEAM model to be carried out in a detailed and specific manner, this expectation does not mean that the implementation played out as planned at the classroom level. Spillane (1999) referred to teachers as the ultimate policy brokers; their belief systems are complex and varied, and beliefs always impact how teachers make sense of any mandate (Richards & Lockhart, 1994). Buehl and Fives (2009) posited that a teacher's willingness to embrace policy reform is almost always reliant on an individual's world-view and belief system. Likewise, in this study, teachers' beliefs about the

appropriate nature of classroom practice often impacted which measures of TEAM were implemented and which were ignored. The triangulation of daily classroom observations, TEAM observations, and stated beliefs demonstrated that despite policy mandates many teachers will continue making classroom decisions that they view as most appropriate. For example, Sally, Rebecca, and Jessica continued to teach in the manner they believed to be most appropriate even if it meant a lower score on the TEAM evaluation. Other teachers such as Joanne and Kendra made instructional changes on a very surface level, but ultimately continued teaching in a manner closely tied to their beliefs about appropriate practice. Kagan (1992) noted that in order to begin the difficult process of changing beliefs, individuals must be given "extended opportunities to examine, elaborate, and integrate new information into their existing belief systems" (p. 77). This study also validated this argument. Allison, Karen, and Laura all received on-going professional development opportunities that allowed them to dialogue about the TEAM policy implementation. Not surprisingly, this group of teachers was more willing to embrace the policy than were the other teachers.

The professional development opportunities at Whiteside Elementary are just one example of how context impacted levels of change and acceptance in terms of teacher belief systems and instructional decision-making. While policy makers may believe that TEAM can be implemented in a blanket fashion across the state of Tennessee, this study demonstrated the difficulty of employing any policy in a rigidly consistent manner across numerous districts, schools, and even classrooms. Each teacher in this study experienced the implementation of TEAM in a different way. These varying experiences were often a result of differences in context at the district, school, and classroom level. The contrasting ways in which district

leaders and administrators translated policy messages as well as school demographics and teacher experiences always impact the manner in which a policy is carried out.

The influence of those in decision-making roles. Prior research indicates that individuals in positions of power or leadership roles influence not only the initial creation of policy but also the manner in which policy is carried out in the classroom (Spillane, 1996; Wise, 1987). The results of this study indicated similar findings and pointed to people at varying levels of power as influences towards change. Some researchers have argued that the distance of policy makers from the classroom may lead to disconnect and make policy fulfillment a difficult task (Eisenhart, Cuthbert, Shrum, & Harding, 2001; Toll, 2001; Wise, 1987). In this study, teachers often demonstrated feelings of disconnect and even disgust in terms of policy makers, which was indicated by their inability to identify exactly who "they" (those in power) were as well as their skeptical feelings about "their" knowledge of daily classroom life. This response indicated that when teachers lack trust in those creating classroom mandates, they may be much less likely to implement said reforms.

Additionally, this study validated the research that notes the impact of those in roles of power at the district and school levels as well. Spillane's (1996) and Coburn's (2006) work reminds us that district leaders and administrators often influenced the way a problem is framed in a particular community. Because messages from policy creators are filtered through them, district leaders and administrators must always be considered in terms of the effects of classroom practice. Likewise, in this study the actions of district leaders seemed to be quite significant. On the one hand, teachers who received greater support and training from their district appeared to be more confident in their understanding of the TEAM model and, in turn, more supportive of its

implementation. On the other hand, those who received little support and training were less likely to embrace the model.

Perhaps the most influential finding in this study was the role of the individual administrator at each school site. Coburn (2008) imparted the notion that problem framing is a key element in the success of school reform. This study validated this idea in that the way in which each administrator framed the policy mandate had important implications for the way it was carried out in classrooms. For example, the principal who framed TEAM as important and worthwhile had teachers who showed deeper instructional changes than did the principal who framed TEAM as just one more thing to check off the "to-do" list. This finding indicates that the support of those in leadership roles, and particularly support of the administrator, is vital to the success of policy change.

The messiness of applying scoring data. The TEAM evaluation model focuses heavily on the use of value-added data when calculating teacher evaluation scores. While most educational researchers concurred that the value-added model greatly improved upon past attainment models, many worried that these measures weren't reliable enough to be used for high-stakes decisions such as teacher evaluation, hiring, firing, and performance-based pay (Papay, 2011; Rothstein, 2007; Rubin et al., 2004). The results of this study indicated that the use of value-added data may be problematic, particularly in primary grades, because of the variance in the way it is used across schools, districts, and states.

Additionally, the variances in the way scores are applied make it difficult to make broad comparisons of TEAM scores. Again, context comes into play as a mitigating factor. For example, one teacher's score might reflect DIBELS assessment scores while another reflects SAT-10 scores and yet another reflects running records scores. These variations in context make

it very difficult to make definitive comparisons across teachers, schools, and districts based solely on TEAM composite scores.

Implications

In this section, I highlight possible implications of this study both in regards to specific ideas for the TEAM model as well as larger implications for teacher evaluation in general.

Specific implications for the TEAM model. In this section I address possible ways that the TEAM model could be improved. These include: differentiation of the rubric and consideration of the issues of time and pacing. These changes are based on issues that several participants noted as problematic. Both of these ideas seem to involve minor variations that could potentially improve the model.

Differentiation of the TEAM rubric. The teachers in the study indicated that the TEAM rubric in was a "one-size-fits-all" model. Several of them believed that it would be beneficial to modify the rubric to fit more appropriately with specific age groups or grade levels. For example, there could be a rubric for primary grades, intermediate grades, middle grades, and high school. The teachers believed that this would make the rubric more appropriate for evaluating teaching a specific demographic of students.

Consideration of issues of time and pacing. For many of the teachers interviewed, the expectations concerning time and pacing restricted what they believed to be appropriate classroom instruction. Teachers felt that they must demonstrate all required elements of the rubric within a particular time period. This was problematic for two reasons: 1) teachers felt that they must often avoid "teachable moments" and thus stop student learning at times when they viewed it as necessary and productive and 2) teachers argued that although it is impossible to touch on every requirement during a single lesson they are demonstrating all requirements over

the course of a day, week, or unit. This indicated that a closer look at the requirements for pacing and its connection to the rubric may be necessary. While time and pacing are arguably notable issues in the planning of instruction, it is worrisome to think that they might take precedence over student learning.

Implications for the literacy classroom. It seems that TEAM does encourage some behaviors that may result in improved literacy instruction. For example, some teachers in the study moved from basic questioning and lower-level thinking strategies to more appropriate use of critical thinking in the classroom. TEAM also seemed to assist others in moving from whole-group instruction to including different types of groupings as well as moving toward more detailed lesson planning and strategizing.

We must be cognizant that when encouraging these important teaching behaviors we do not want to disregard those tried and true methods that we know to be indicative of good literacy instruction. The teachers in this study believed that the TEAM model discouraged the practices of small group instruction, independent reading, and literacy centers. It is unclear whether this message was intended or unintended on the part of policy makers. Regardless, it was the way teachers made sense of the policy implementation. Numerous researchers have written about the multitude of benefits that result from the implementation of these three common literacy practices (Allington, 2002; Cunningham & Stanovich, 1998; Franzese, 2002; McGill-Franzen, 2006; Morrow, 2002). It is concerning that teachers felt the need to discontinue literacy practices that they believed to be valuable and that research has shown to be valuable to tailor instruction to the TEAM rubric.

Broader implications for teacher evaluation. Previously, I touched on possible specific implications for the Tennessee TEAM model as well as implications for literacy instruction in

terms of TEAM evaluation. In this final section, I address broader implications for teacher evaluation policy including the following: speedy implementations, scoring, consideration of differentiation for teachers, and teacher autonomy.

Speedy implementations/"the race." It seems that our education system is currently heavily laden with the discourse of speed. In terms of education we want things to be done well, but we also seem to want to do them quickly. Hence the current phrases: "Race to the Top" and "Tennessee First to the Top." In keeping with this notion of speed, it seems that Tennessee was anxious to be one of the first states to put in place a model of teacher evaluation that tied teacher evaluation scores directly to their students' achievement. While Tennessee can now claim the title of one of the first to implement this model of evaluation, the speed of the TEAM implementation appears to be the cause of many problems and concerns for teachers. As a result, teachers did not understand the model, and even those in leadership roles appeared to be improperly trained in the language of the model. This finding indicated that policy makers might be wise to slow down and gauge levels of competence before beginning to implement large-scale policy mandates. Field testing and more training could have curbed many of those initial problems of the TEAM implementation.

Scoring and making comparisons. This study demonstrated that comparing TEAM evaluation scores across schools, districts, and the state was a messier process than previously imagined. As the data has indicated, evaluation scores may be affected by a plethora of factors including: evaluator subjectivity, administrator expectations, and the variety of scoring methods. While it seems appropriate to give schools choice in testing methods and scoring application, I would also argue that this very choice makes it impossible to make fair comparisons and judgments across multiple districts, schools, and teachers. These scores could be helpful as

another layer of information, another way to understand our students' progress, but they are inappropriate when used as the basis for making high-stakes decisions.

Differentiation for teachers. Another important consideration may be that a one-size-fits-all teacher evaluation model isn't appropriate for every teacher. Burden (1986) posited three stages of teacher career development in which teachers are located at different points on the career development spectrum, with some needing more direct supervision and guidance and others needing far less. This study's findings support this notion in that TEAM seemed to further the teaching of some while stagnating the progress of others. For example, one teacher admitted that she was spending almost all of her instructional time in whole-group teaching, and TEAM encouraged her to think about alternative methods of grouping. Another teacher, however, was already employing a variety of creative grouping strategies throughout the day, and TEAM forced her to take a step backwards and move to a more "traditional" sequence of grouping where the class moves from whole-group to small-group to independent work time. Thus, I would argue that different levels of training and development, depending on the needs of the teacher, may be appropriate (Stallings & Krasavage, 1986).

Teacher autonomy and giving teachers a say. Lastly and perhaps most importantly, this study encourages consideration of the issue of teacher autonomy. Various researchers have lamented the essentiality of teacher autonomy in terms of furthering professionalism, stake in student learning, and success of policy implementation (Allington, 2002; Carlone et al., 2010; Weathers, 2011). This study validates those arguments. Teachers expressed their feelings of helplessness in terms of policy implementation. In many cases they found themselves engaging in practices they considered to be in direct conflict with what they believed to be effective instruction. When teachers are given appropriate autonomy they are happier, more devoted to

their profession, and have a greater stake in student outcomes (Allington, 2002; Malmberg & Hagger, 2009). From this stance, I believe policymakers would be wise to avoid rigid mandates that discourage teacher choice and teacher decision making.

Additionally, I would argue that it is essential that teachers be given a voice in the creation of evaluation systems. We have a long history of excluding teachers from the educational conversation (examples: Bush's Education Summit and The National Reading Panel). When teachers are given a say, policy has a greater chance of succeeding (Hall, 2005), and a bridge might begin to be built between "they" (the policy makers) and "those in the trenches" (the teachers).

Epilogue

In this epilogue I provide information about future directions for the teachers and final thoughts about TEAM evaluation. This information is based on data collected from follow-up interviews that I conducted in April of 2013. The questions I asked were based on the following: implications of TEAM, job placement for the upcoming year, and final thoughts.

Laura: Laura will continue in her role of TEAM lead teacher for the 2013-2014 school and therefore will continue evaluating and scoring her peers based on the TEAM rubric. She noted that at times this role has created divisions and tensions between her and the other teachers at her school. Nevertheless, she has agreed to continue in the role because of her respect and support for her principal. Because her composite score was a 5, there have been no negative implications for her as a result of TEAM. She believes this will be the case again in terms of next year. Laura noted that the most positive aspect of TEAM this year was collaboration between teachers that resulted from the TEAM professional development trainings implemented by her principal. Lastly, she explained that the most negative aspect of TEAM this year for her was watching how stressed her peers appeared when she walked into their classrooms to evaluate them.

Karen: Karen is certain that she will remain at Whiteside Elementary for the 2013-2014 school year, but she is uncertain as to whether or not she will remain in first grade. Karen has not yet received a composite score for this year. However, based on her TEAM evaluation scores thus far she believes that there will be no negative implications for her as a result of TEAM. Karen viewed the feedback from evaluators as the most positive aspect of TEAM this year. She noted that some of her peers believe the feedback is too harsh, but she saw it as a path to improvement.

Karen explained that if she could change anything about TEAM, she would make all the observations unannounced because she finds preparing the announced lessons overly stressful.

Allison: Allison will remain in first grade at Whiteside Elementary for the 2013-2014 school year. She will also be moving from apprentice to professional status which means she will have two TEAM observations next year instead of four. There were no implications for Allison based on her previous composite score and she believed the same will be true for the upcoming year. Allison noted that the most positive aspect of TEAM this year was the TEAM professional developments implemented by her principal. She explained that these trainings pushed her to be more intentional in her teaching and also tied Common Core to TEAM. Allison reported that if she could drop anything from the TEAM requirements for next year it would be to alter the rigid nature of the model. She explained that she feels that she must fit into a particular mold that isn't always conducive to good teaching.

Sally: Sally will remain in first grade at Gary Elementary for the 2013-2014 school year. The following year she hopes to move into a reading specialist position. She has nearly completed the requirements for the reading specialist license. Sally had to be observed more frequently this year as a result of her last TEAM composite score. She noted that repercussions for next year are still unknown and will depend on how well her school performs on the yearly TCAP achievement test. She now has a new principal and sees improvements in how TEAM is handled as well as improvements in the general workings of the school. Sally explained that she has thought more about incorporating different types of groupings as a result of TEAM. She stated

that if she could get rid of one element of TEAM it would be the focus on one achievement test and its connection to her evaluation score.

Rebecca: Rebecca is pretty positive that she will continue in her current position as a first grade teacher at Newton Elementary for the 2013-2014, however, decisions have not been finalized. She does not believe that there will be any negative implications for her as a result of TEAM. For Rebecca the most positive aspect of TEAM this year has been that her principal and the other TEAM evaluator from central office both have experience as teachers. She noted that because of their experience they use the TEAM rubric as a form of positive motivation. Their actions have led to an increase in positive feelings about TEAM for Rebecca. Rebecca explained that if she could change one thing, it would be to make scoring a 5 more attainable.

Joanne: In January Joanne had to take a medical leave of absence because she began losing her eyesight. She will not be returning to Newton Elementary for the 2013-2014.

Kendra: Kendra will return to Lindley Elementary as a first grade teacher for the 2013-2014 school year. She expects that this year's composite score will be similar to last year's score of 5 and as a result doesn't foresee any negative implications as a result of TEAM. Kendra expressed concern that her last TEAM evaluation involved the application of a model she was unfamiliar with, the Learning Focus Model. She wasn't sure if the use of this model indicated a different direction for her district. She also noted that there was conflict in her district in terms of how to apply DIBELS testing data to the TEAM quantitative score. Kendra stated that the most positive aspect for her this year as a result of TEAM was that she became more proficient in

differentiation. She does wish that they could disregard the focus on displaying and repeating the standards because she doesn't view this focus as beneficial for her first graders.

Jessica: Jessica will not return to Lindley Elementary for the 2013-2014 school year. Her husband is in the military, and they will be relocating to another state. Jessica will spend the summer looking for a job in her new home town. She noted that she is now more proficient in instructing her students in higher order thinking strategies as a result of TEAM, and she will take these skills with her to her new place of employment. Jessica hopes that in her new state she will be able to say good-bye to what she views as an overemphasis on posting and relaying standards and doing things for others that she views as unbeneficial for students.

List of References

Abbott, A. (2004). *Methods of discovery: Heuristics for the social sciences.* New York: W.W. Norton.

Adamczyk, D. (2011, April 5). The value of value-added assessment. *ASHA Leader,* p. 27.

Allington, R. (2002). What I've learned about effective reading instruction from a decade of studying exemplary classroom teachers. *Phi Delta Kappan*, *83*(10), 740-747.

Allington, R. (2002). *Big brother and the national reading curriculum: How ideology trumped evidence.* Portsmouth, N.H.: Heinmann.

Allington, R., & Cunningham, P. (1996). *Schools that work: Where all children read and write.* New York: Harper Collins.

Atkins, A. O. (1996). Teachers' opinions of the teacher evaluation process. Retrieved from ERIC document Reproduction Service. (ED398628)

Bell, T. H. (1988). *The thirteenth man: A Reagan cabinet memoir.* New York: Free Press.

Begum, S. (2012). A secondary science teacher's beliefs about environmental education and its relationship with the classroom practices. *International Journal of Social Sciences and Education*, *2*(1), 10-29.

Berends, M., Chun, J., Schuyler, G., Stockly, S., & Briggs, R. J. (2002). *Challenges of conflicting reforms: Effects of New American Schools in a high poverty district.* Santa Monica, CA: RAND.

Bond, G. L., & Dykstra, R. (1967). The cooperative research program in first-grade reading instruction. *Reading Research Quarterly*, *2*(4), 5-142.

Bracey, G. (2004). Serious questions about the Tennessee Value-Added Assessment System. *Phi Delta Kappan*, *85*(5), 716-718.

Bryant, D., Clifford, R., & Feinberg, E. (1991). Best practices for beginners: Developmental appropriateness in kindergarten. *American Educational Research Journal*, *28*(4), 783-803.

Buehl, M. M., & Fives, H. (2009). Exploring teachers' beliefs about teaching knowledge: Where does it come from? Does it change? *The Journal of Experimental Education*, *77*(4), 367-407.

Burden, P. R. (1986). Teacher development: Implications for teacher education. In L. Katz & J. Ruth (Eds.), *Advances in teacher education* (Vol. 2, pp. 185-219). Norwood, NJ: Ablex.

Burris, C. C., & Welner, K. G. (2011). Conversations with Arne Duncan: Offering advice on educator evaluations. *Phi Delta Kappan*, *93*(2), 38-41.

Burris, C. C., & Welner, K. G. (2011). Letter to secretary of education Arne Duncan concerning evaluation of teachers and principals. *National Education Policy Center,* 1-8.

Carlone, H. B., Haun-Frank, J., & Kimmel, S. C. (2010). Tempered radicals: Elementary teachers' narratives of teaching science within and against prevailing meanings of schooling. *Cultural Studies of Science Education*, *5*, 941-965.

Carspecken, P. (1996). *Critical ethnography in educational research.* New York: Routledge.

Cavanagh, S. (2011, September 14). Race to top winners feel heat on evaluations. *Education Week*, 20-21.

Chiseri-Strater, E., & Sunstein, B. S. (1997). *Fieldworking: Reading and writing research.* Upper Saddle River, NJ: Prentice Hall.

Ciardiello, A. V. (2004). Democracy's young heroes: An instructional model of critical literacy practices. *The Reading Teacher*, *58*(2), 138-147.

Clark, C. M., & Peterson, P. L. (1986). Teachers' thought processes. In M.C. Wittrock (Ed.), *Handbook of Research on Teaching, 3* (p. 255-296). New York: Macmillan.

Clark, D. (1993). Teacher evaluation: A review of the literature with implications for educators. Retrieved from EDRS Digital Theses. (ED359174)

Coburn, C. E. (2001). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis*, *23*(2), 145-170.

Coburn, C. E. (2006). Framing the problem of reading instruction: Using frame analysis to uncover the microprocesses of policy implementation. *American Educational Research Journal*, *43*(3), 343-379.

Colby, S., Bradshaw, L., & Joyner, R. (2002, April). *Teacher evaluation: A review of the literature*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Thousand Oaks, CA: Sage.

Costello, R., Elson, P., & Schacter, J. (2008). An introduction to value-added analysis. *Catholic Education: A Journal of Inquiry and Practice*, *12*(2), 193-2003.

Creighton, D. (1997). Critical literacy in the classroom. *Language Arts*, *6*(6), 438-445.

Croninger, R. G., & Valli, L. (2009). Where is the action? Challenges to studying the teaching of reading in elementary classrooms. *Educational Researcher*, *38*(2), 100-108.

Cross, C. T. (2004). *Political education.* New York: Teachers College Press.

Cuban, L. (1986). *Teachers and machines.* New York: Teachers College Press.

Cuban, L. (1995). The hidden variable: How organizations influence teacher responses to secondary science curriculum reform. *Theory Into Practice*, *34*(1), 4-11.

Cuthbert, A. M. (1984). A case study of policy-induced staff development in a local school district. *Dissertation Abstract international, 45,* 2025A. University Microfilms No. DA8421858.

Daley, G., & Kim, L. (2010). A teacher evaluation system that works. Working paper, National Institute for Excellence in Teaching.

Danielson, C. (2010). Evaluations that help teachers learn. *The Effective Educator*, *68*(4), 35-39.

Darling-Hammond, L. (1990). Teacher evaluation in transition: Emerging roles and evolving methods. In J. Millman and L. Darling-Hammond (Eds.), *The new handbook of teacher evaluation.* Newbury Park, CA: Sage.

Decker, L. E., & Rimm-Kaufman, S. E. (2008). Personality characteristics and teacher beliefs among pre-service teachers. *Teacher Education Quarterly*, *35*(2), 45-64.

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). New York: State University of New York Press.

Eisenhart, M. A., Cuthbert, A. M., Shrum, J. L., & Harding, J. R. (2001). Teacher beliefs about their work activities: Policy implications. *Theory into Practice*, *27*(2), 137-144.

Elmore, R., Peterson, P., & McCarthy, S. (1996). *Restructuring in the classroom.* San Francisco: Jossey-Bass.

Ely, M., Anzul, M. Friedman, T., Garner, D., & Steinmetz, A. M. (1991). *Doing qualitative research: Circles within circles.* London: Falmer.

Erickson, F. (1986). Qualitative methods in research on teaching. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.) (pp. 119-61). New York: Macmillan.

Ertmer, P. A. (2005). Teacher pedagogical beliefs: The final frontier in our quest for technology. *Educational Technology Research and Development*, *53*(4), 25-39.

Everhart, R. B. (2004). Toward a critical social narrative of education. In G. W. Noblit, S. Y. Flores, & E. G. Murrillo (Eds.), *Postcritical ethnography* (pp. 59-76). New Jersey: Hampton.

Fairclough, N. (1995). *Critical discourse analysis. The critical study of language.* New York: Longman.

Fenstermacher, G. D. (1978). A philosophical consideration of recent research on teacher effectiveness. In L. S. Shumlan (Ed.), *Review of Educational Research*, *52*(1), 7-30.

Fenstermacher, G. (1986). Philosophy of research on teaching: Three aspects. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 37-49). New York: Macmillan.

Fine, M. (1994). Dis-stance and other stances: Negotiations of power inside feminist research, In A. Gitlin (Ed.), *Power and method: Political activism and educational research* (pp. 13-35). London & New York: Routledge.

Firestone, W. (1993). Alternative arguments for generalizing from data as applied to qualitative research. *Educational Researcher*, *22*(16), 16-23.

Franzese, R. (2008). *Reading and writing in kindergarten: A Practical Guide.* New York: Scholastic.

Friere, P. (2000). *Pedagogy of the oppressed* (30th anniversary ed.). New York: Continuum. (Original work published 1970).

Fuller, B., Wood, K., Rapoport, T., & Dornbusch, S. M. (1982). The organizational context of individual efficacy. *Review of Educational Research*, *52*(1), 7-30.

Giroux, H. A. (1993). Literacy and the politics of difference. In C. Lankshear & P. L. McLaren (Eds.), *Critical literacy: Politics, praxis, and the postmodern* (pp. 367-377). Albany: State University of New York Press.

Glazerman, S., & Seifullah, A. (2010). An evaluation of the teacher advancement program in Chicago: Year two impact report. *Mathematica Policy Research*, 1-35.

Glesne, C. (2011). *Becoming qualitative researchers: An introduction.* Boston: Pearson.

Grbich, C.(2007). *Qualitative data analysis: An introduction.* London: Sage.

Green, T. E. (1971). *The activities of teaching.* New York: McGraw-Hill.

Guerra, P. L., & Nelson, S. W. (2009). Changing professional practice requires changing beliefs. *Phi Delta Kappan*, *90*(5), 354-359.

Hall, L. (2005). Teachers and content area reading: Attitudes, beliefs, and change. *Teaching and Teacher Education*, *21*(4), 403-414.

Hall, L, Johnson, A. S., Juzwik, M. M., Wortham, S. E. F., & Mosley, M. (2010). Teacher identity in the context of literacy teaching: Three explorations of classroom positioning and interaction in secondary schools. *Teaching and Teacher Education*, *26*, 234-243.

Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first grade classroom make a difference for children at risk of school failure? *Child Development*, *76*(5), 949-967.

Haney, J. J., & McArthur, J. (2002). Four case studies of prospective science teachers' beliefs concerning constructivist teaching practices. *Science Education*, *86*(6), 783-802.

Harris, D. N. (2010). Clear away the smoke and mirrors of value-added. *Phi Delta Kappan*, *91*(8), 66-69.

Harste, J. C. (2000). Supporting critical conversations in classrooms. In K.M. Pierce (Ed.), *Adventuring with books; A booklist for pre-K-grade 6* (12th ed., pp. 507-554). Urbana, IL: National Council of Teachers of English.

Hatch, J. A. (2002). *Doing qualitative research in education settings.* Albany: State University of New York Press.

Heath, S. B. (1983). *Ways with words: Language, life, and work in communities and classrooms.* Cambridge, England: Cambridge University Press.

Heath, S. B. (2004). The children of Trackton's children: Spoken and written language in social change. In R. B. Ruddell & N. J. Unrau (Eds.), *Theoretical models and processes of reading* (5th ed., pp. 187-209). Newark, DE: International Reading Association.

Heitin, L. (2011, October 19). Evaluation system weighing down Tennessee teachers. *Education Week*, *31*(8), 1-14.

Helle, L., & Murtonen, M. (2001). Computer science and teacher students' beliefs concerning the nature of expertise in their field of study. *European conference on learning and instruction* (vol. 28, pp.1-17). Fribourg, Germany: Publishers.

Jennings, N. E. (1996). *Interpreting policy in real classrooms: Case studies of state reform and teacher practice*. New York: Teachers College Press.

Johnston, B. J. (2004). Critical theory, critical ethnography, critical conditions: Considerations of postcritical ethnography. In G. W. Noblit, S. Y. Flores, & E. G. Murrillo (Eds.), *Postcritical ethnography* (pp. 59-76). New Jersey: Hampton.

Juel, C., & Minden-Cupp, C. (2000). Learning to read words: Linguistic units and strategies. *Reading Research Quarterly*, *35*, 458-492.

Kagan, D. M. (1990). Ways of evaluating teacher cognition: Inferences concerning the goldilocks principle. *Review of Educational Research*, *60*(3), 419-469.

Kern, D. (2007). Flying with critical literacy wings. *The NERA Journal*, *43*(2), 87-89.

Knoblauch, C. H., & Brannon, L. (1993). *Critical teaching and the idea of literacy.* Portsmouth, NH: Boynton/Cook.

Kroeger, D. (2008). Thoughts on narrowing the policy-practice gap: Literacy teachers as critical bureaucrats. *Journal of Reading Education*, *34*(1), 46-50.

Labov, W. (1972). *Sociolinguistic patterns.* Philadelphia: University of Pennsylvania Press.

Lee, C., & Runyan, B. (2011). From apprehension to critical literacy. *Journal of Educational Thought*, *45*(1), 87-105.

Lewison, M., Flint, A., & Van Sluys, K. (2002). Taking on critical literacy: The journey of newcomers and novices. *Language Arts*, *79*(5), 382-392.

Lincoln, Y.S., & Gubba, E.G. (1985). *Naturalistic inquiry.* Beverly Hill, CA: SAGE Publications.

Lortie, D. (1975). *School teacher.* Chicago: University of Chicago Press.

Loup, K. S., Garland, J. S., Ellett, C. D., & Rugutt, J. K. (1996). Ten years later: Findings from a replication of a study of teacher evaluation practices in our 100 largest school districts. *Journal of Personnel Evaluation in Education*, *10*, 203-226.

Luna, C., Botelho, M. J., Fontaine, D., French, K., Iverson, K., & Matos, N. (2004). Making the road by walking and talking: Critical literacy and/as professional development in a teacher inquiry group. *Teacher Education Quarterly*, *31*(1), 67-80.

Malmberg, L., & Hagger, H. (2009). Changes in student teachers' agency beliefs during a teacher education year, and relationships with observed classroom quality, and day-to-day experiences. *British Journal of Educational Psychology*, *79*, 677-694.

Marshall, J., & Klein, A. M. (2009). Lessons in social action: Equipping and inspiring students to improve their world. *The Social Studies*, *100*(5), 218-221.

McDaniel, C. (2004). Critical literacy: A questioning stance and the possibility for change. *The Reading Teacher*, *57*(5), 472-481.

McGill-Franzen, A. (2000). Policy and instruction: What is the relationship? In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research volume III* [Kindle version]. New Jersey: Lawrence Erlbaum Associates.

McGill-Franzen, A. (2006). *Kindergarten literacy.* New York: Scholastic.

McGill-Franzen, A., & Allington, R. (2006). Contamination of current accountability systems. *Phi Delta Kappan*, *87*(10), 762-766.

McGill-Franzen, A., Lanford, C., & Adams, E. (2002). Learning to be literate: A comparison of five urban early childhood programs. *Journal of Educational Psychology*, *94*(3), 443-464.

McNeil, J. D. (1981). Politics of teacher evaluation. In J. Millman (Ed.), *Handbook of teacher evaluation.* Beverly Hills, CA: Sage.

Mehan, H. (1996). The construction of an LD student: A case study of the politics of representation. In M. Silverstein & G. Urban (Eds.), *Natural histories of discourse* (pp. 253-276). Chicago: University of Chicago Press.

Merriam, S.B. (1998). *Qualitative research and case study applications in education.* San Fransico: Jossey-Bass.

Meyer, L. A., Waldrop, J. L., Hastings, C. N., & Linn, R. L. (1993). Effects of ability and settings on kindergarteners' reading performance. *Journal of Educational Research*, *86*(3), 142-160.

Mitchell, D. E., & Kerchner, C. T. (1983). Collective bargaining and teacher policy. In L. S. Shulman & G. Sykes (Eds.), *Handbook of teaching and policy* (pp. 214-234). New York: Longman.

Morgan, W. (1997). *Critical literacy in the classroom: The art of the possible*. London & New York: Routledge.

Morrell, E. (2010). Critical research and the future of literacy education. *Journal of Adolescent & Adult Literacy*, *53*(2), 96-104.

Morrison, F. J., & Connor, C. M. (2002). Understanding schooling effects on early literacy: A working research strategy. *Journal of School Psychology*, *40*, 493-500.

Morrow, L.M. (2002). *The literacy center: Contexts for reading and writing* (2nd ed.). York, ME: Stenhouse.

National Institute for Excellence in Teaching. (2012, May). Best Practices Center Portal. Web. http://nietbestpractices.org/

Nespor, J. (1987). The role of beliefs in the practice of teaching. *Journal of Curriculum Studies*, *19*(4), 317-328.

Noblit, G. W., & Eaker, D. J. (1999). Evaluation designs as political strategies. In G. W. Noblit (Ed.), *Particularities: Collected essays on ethnography and education* (pp. 61-79). New York: Peter Lang.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects. *Educational Evaluation and Policy Analysis*, *26*(3), 237-257.

Ogan-Bekiroglue, F., & Akkoc, H. (2009). Preservice teachers' instructional beliefs and examination of consistency between beliefs and practices. *International Journal of Science and Mathematics Education*, *7*, 1173-1199.

O'Quinn, E. J. (2005). Critical literacy in democratic education: Responding to sociopolitical tensions in U.S. schools. *Journal of Adolescent & Adult Literacy*, *49*(4), 260-267.

Otterman, S. (2011, May 14). Teacher reviews will put more focus on state tests. *New York Times,* p. 17.

Papay, J. P. (2011). Different tests, different answers: The stability of teacher value-added estimates across outcome measures. *American Educational Research Journal*, *48*(1), 163-193.

Patton, M. Q. (1990). *Qualitative evaluation methods.* Newbury Park, CA: Sage.

Pembroke, E., & Goedert, E. (1982). What is the key to developing an effective teacher evaluation system. *NASSP,* 29-37.

Pianta, R. C., Belsky, J., Houts, R., L. F., Morrison, The National Institute of Child Health and Human Development (NICHD) Early Child Care Research Network. (2007). Opportunities to learn in America's elementary classrooms. *Science*, *315*, 1795-1796.

Pianta, R. C., & La Paro, K. M. (2003). Improving early school success. *Educational Leadership*, *60*(7), 22-24.

Potter, W. J. (1996). *An analysis of thinking and research about qualitative methods.* Mahwah, NJ: Erlbaum.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, *66*, 221-227.

Powers, S. W., Zippay, C., & Butler, B. (2006). Investigating connections between teacher beliefs and instructional practices with struggling readers. *Reading Horizons Journal*, *47*(2), 122-157.

Pressley, M., Allington, R. L., Wharton-McDonald, R., Block, C. C., & Morrow, L. M. (2001). *Learning to read: Lessons from exemplary first-grade classrooms.* New York: The Guilford Press.

Pressley, M., Rankin, J., & Yokoi, L. (1996). A survey of instructional practices of primary teachers nominated as effective in promoting literacy. *The Elementary School Journal*, *96*(4), 363-384.

Ravitch, D. (2010). *The death and life of the great American school system: How testing and choice are undermining education.* New York: Basic Books.

Redinger, L. (1988). Evaluation: A means of improving teacher performance. *Research Report.*

Rettig, P. R. (2004). Beyond organizational tinkering: A new view of school reform. *Educational Horizons*, 82(4), 260-265.

Richards, J. C., & Lockhart, C. (1994). *Reflective teaching in second language classroom.* Cambridge, England: Cambridge University Press.

Richards, L., & Morse, J. M. (2007). *Readme first for a user's guide to qualitative methods* (2nd ed.). Thousand Oaks, CA: Sage.

Richardson, V. (1996). The role of attitudes and beliefs in learning to teach. In J. Sikula (Ed.), *Handbook of research on teacher education* (pp. 102-119). New York: Simon & Schuster/Macmillan.

Robelen, E. W. (2011, June). Frustration at heart of Washington rally. *Education Week, 30*(35), 14-15.

Rogers, R. (2002). Between contexts: A critical discourse analysis of family literacy, discursive practices, and literate subjectivities. *Reading Research Quarterly*, *37*(3), 248-277.

Rogers, R., & Mosley, M. (2006). Racial literacy in a second-grade classroom: Critical race theory, whiteness studies, and literacy research. *Reading Research Quarterly*, *41*(1), 462-495.

Rogers, R., Malancharuvil-Berkes, E., Mosley, M., Hui, D., & Joseph, O. J. (2005). Critical discourse analysis in education: A review of the literature. *Review of Educational Research*, *75*(3), 365-416.

Rothstein, J. (2007). Does competition among public schools benefit students and taxpayers? *American Economic Review*, 97(5), 2026-2037.

Ruddell, R. B. (1997). Researching the influential literacy teacher: Characteristics, beliefs, strategies, and new research directions. In C. K. Kinzer, K. A. Hinchman, & D. J. Leu (Eds.), *Inquiries in literacy theory and practice: Forty-sixth yearbook of the National Reading Conference* (pp. 37-53). Chicago: National Reading Conference.

Saldana, J. (2009). *The coding manual for qualitative researchers.* Thousand Oaks, CA: Sage.

Sawchuck, S. (2011, May 18). NEA proposes making a shift on evaluation. *Education Week*, *30*(31), 1-19.

Schacter, J., & Thum, Y. M. (2005). Tapping into high quality teachers: Preliminary results from the teacher advancement program comprehensive school reform. *School Effectiveness and School Improvement*, *16*(3), 327-353.

Schmidt, R., Armstrong, L., & Everett, T. (2007). Teacher resistance to critical conversation: Exploring why teachers avoid difficult topics in their classrooms. *The NERA Journal*, *43*(2), 49-55.

Schochet, P., & Chiang, H. (2010). *Error rates in measuring teacher and school performance based on student test score gains.* Washington, D.C.: Institute of Education.

Schunk, D. H. (2000). *Learning theories: An educational perspective* (3rd ed.). Upper Saddle River, NJ: Merrill/Prentice Hall.

Scribner, S. (1984), Literacy in three metaphors. *American Journal of Education*, *93*, 6-21.

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*(1), 1-22.

Shulman, L. S. (1983). Autonomy and obligation: The remote control of teaching. In L. S. Shulman & G. Sykes (Eds.), *Handbook of Teaching and Policy* (pp. 484-504). New York: Longman.

Sipe, L. R., & Ghiso, M. P. (2004). Developing conceptual categories in classroom descriptive research: Some problems and possibilities. *Anthropology and Education Quarterly*, *35*(4), 472-85.

Siu-Runyan, Y. (2007). Critical literacies: Stepping beyond the NCLB act. *The NERA Journal*, *43*(2), 63-71.

Slavin, R. E. (1986). The napa evaluation of Madeline Hunter's ITIP: Lessons learned. *The Elementary School Journal*, *87*(2), 165-171.

Smarick, A. (2011). Diplomatic mission. *Education Next*, *11*(1), 56-63.

Smith, F. (1998). *The book of learning and forgetting.* New York: Teachers College Press.

Spillane, J. P. (1996). School districts matter: Local educational authorities and state instructional policy. *Educational Policy*, *10*(1), 63-87.

Spillane, J. P. (1999). External reform initiatives and teachers' efforts to reconstruct their practice: The mediating role of teachers' zones of enactment. *Journal of Curriculum Studies*, *31*(2), 143-175.

Spradley, J. P. (1979). *The ethnographic interview*. New York: Holt, Rinehart, & Winston.

Spradley, J. P. (1980). *Participant observation.* Fort Worth, TX: Harcourt.

Stallings, J., & Krasavage, E. M. (1986). Program implementation and student achievement in a four year Madeline Hunter follow-through project. *The Elementary School Journal*, *87*(2), 117-138.

Stribling, S. M. (2008). Using critical literacy practices in the classroom. *The NERA Journal*, *44*(1), 34-38.

Stuhlman, M. W., & Pianta, R. C. (2009). Profiles of educational quality in first grade. *The Elementary School Journal*, *109*(4), 323-342.

Sundquist, J. L. (1968). *Politics and policy: The Eisenhower, Kennedy, & Johnson years.* Washington, D.C.: Brookings Institution.

Sunstein, B. S., & Chiseri-Strater, E. (2002). *Field working: Reading and writing research* (2nd ed.). New York: Bedford/St. Martin's.

TAP. (2012, May). *The System for Teacher and Student Advancement*. National Institute for Excellence in Teaching. Web. http://www.tapsystem.org/about/about.taf

Tennessee First to the Top. (2012, May). *Tennessee Educator Acceleration Model (TEAM) Evaluation System Handbook*. National Institute for Excellence in Teaching. Web. http://nietbestpractices.org/

Tennessee First to the Top. (2011, May). Tennessee Department of Education. Web. http://www.tn.gov/firsttothetop/

Theirot, S., & Tice, K. C. (2009). Teacher's knowledge development and change: Untangling beliefs and practices. *Literacy Research and Instruction*, *48*(1), 65-75.

Toch, T. (2008). Fixing teacher evaluation. *Educational Leadership,* 32-37.

Toll, C. A. (2001). Can teachers and policy makers learn to talk to one another. *The Reading Teacher*, *55*(4), 318-325.

United States. National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform : a report to the Nation and the Secretary of Education, United States Department of Education*. Washington, D.C.: The Commission.

Vagle, M. D. (2011). Lessons in contingent, recursive humility. *Journal of Adolescent & Adult Literacy*, *54*(5), 362-370.

Valli, L., Croninger, R. G., & Walters, K. (2007). Who (else) is the teacher? Cautionary notes on teacher accountability systems. *American Journal of Education*, *113*(4), 635-662.

Van Dijk, T. (1993). Principles of critical discourse analysis. *Discourse and Society*, *4*(2), 249-283.

Weathers, J. M. (2011). Teacher community in urban elementary schools: The role of leadership and bureaucratic accountability. *Education Policy Analysis Archives*, *19*(3), 1-33.

Wedman, J., & Robinson, R. (1988). Effects of extended in-service on secondary teachers' use of content reading instructional strategies. *Journal of Research and Development in Education*, *21*(3), 65-70.

Weiss, M. J. (2008). *Examining the measures used in the federal growth model pilot program.* Paper presented at the annual meeting of the Society for Research on Educational Effectiveness, Washington, D.C.

Wharton-McDonald, R., Pressley, M., & Hampston, J. M. (1998). Literacy instruction in nine first-grade classrooms: Teacher characteristics and student achievement. *Elementary School Journal*, *99*, 101-128.

Williams, S. D. (2001). Part 1: Thinking out of the pro-verbal box. *Computers and Composition*, *18*(1), 21-32.

Winerip, M. (2011, March 7). Evaluating New York teachers, perhaps the numbers do lie. *New York Times,* p. 15.

Wise, A. E. (1987). *Legislated learning: The bureaucratization of the American classroom*. Berkeley: University of California Press.

Wise, A. E., Darling-Hammond, L., McLaughlin, M. W., & Bernstein, H. T. (1985). Teacher evaluation: A study of effective practices. *The Elementary School Journal*, *86*(1), 60-121.

(2011). Grades for teachers. *New York Times, PaperClip Communications,* p. 6.

VITA

Renee Moran was born and grew up in Marshall, NC. She earned a Bachelor of Science in English and Elementary Education in 2001 from the University of North Carolina at Greensboro. She then graduated with a Masters of Arts in Curriculum and Instruction from San Diego State University in 2004. She worked as an elementary school teacher in grades kindergarten, first, and fifth from 2003-2010 in the public schools of North Carolina and California. Renee is currently employed at East Tennessee State University as an assistant professor in reading education. Upon acceptance of this dissertation, Renee will have graduated with a Ph.D. in Teacher Education with a concentration in Literacy Studies from the University of Tennessee at Knoxville in 2013.