Principal-Led State Evaluation Policy: Teacher Perceptions at Four High Schools

Amanda S. Frasier
Michigan State University, amanda.slaten.frasier@gmail.com

Follow this and additional works at: https://nsuworks.nova.edu/tqr

Part of the Educational Assessment, Evaluation, and Research Commons, Educational Leadership Commons, Elementary and Middle and Secondary Education Administration Commons, and the Social and Behavioral Sciences Commons

Principal-Led State Evaluation Policy: Teacher Perceptions at Four High Schools

Abstract

While policy makers have attempted to standardize teacher evaluation, policy is implemented and enacted by school administrators. This study addresses the following question: Considering the legislative efforts to remove control of evaluation from local figures, do teachers perceive school principals as influencing the implementation of state-level evaluation policy and, if so, in what ways? I examined interviews from 14 teachers across four high schools within a district in North Carolina derived from a larger mixed method case study of teacher perceptions of evaluation policy and classroom practice. The results suggest a state-centralized teacher evaluation policy, such as the one utilized at the time of this study, can look vastly different to teachers at the school-level due to principal enactment of the policy. Furthermore, the data suggest the following themes influenced policy implementation: the capacity of principals to evaluate in a timely manner, what a principal chooses to value in a policy, and the perceived effectiveness of a principal as an evaluator of teaching. By taking a closer look at what is happening "on the ground" between teachers and principals in four schools utilizing the same state-level evaluation policy, the lessons learned in this study can help inform future policies.

Keywords

school principals, policy implementation, teacher evaluation, case study

Creative Commons License

Acknowledgements

Principal-Led State Evaluation Policy:
Teacher Perceptions at Four High Schools

Amanda S. Frasier
Michigan State University, East Lansing, USA

While policy makers have attempted to standardize teacher evaluation, policy is implemented and enacted by school administrators. This study addresses the following question: Considering the legislative efforts to remove control of evaluation from local figures, do teachers perceive school principals as influencing the implementation of state-level evaluation policy and, if so, in what ways? I examined interviews from 14 teachers across four high schools within a district in North Carolina derived from a larger mixed method case study of teacher perceptions of evaluation policy and classroom practice. The results suggest a state-centralized teacher evaluation policy, such as the one utilized at the time of this study, can look vastly different to teachers at the school-level due to principal enactment of the policy. Furthermore, the data suggest the following themes influenced policy implementation: the capacity of principals to evaluate in a timely manner, what a principal chooses to value in a policy, and the perceived effectiveness of a principal as an evaluator of teaching. By taking a closer look at what is happening "on the ground" between teachers and principals in four schools utilizing the same state-level evaluation policy, the lessons learned in this study can help inform future policies.

Keywords: school principals, policy implementation, teacher evaluation, case study

Introduction

Over the last few decades there have been public and political concerns in the United States that locally developed teacher evaluations cannot accurately identify the effectiveness of teachers and that poor teachers are often rated as high performing. For instance, in evaluation systems with only two ratings (satisfactory and unsatisfactory), 99% of teachers earned a satisfactory rating, and in evaluation systems with more than two ratings, 94% of teachers received one of the top two ratings while less than 1% were rated unsatisfactory (Weisberg et al., 2009). The Weisberg et al. study termed this top-heavy sort of assessment as the "Widget Effect," stemming from the belief that all teachers are interchangeable and would be equally effective across all classroom contexts, much like a software widget can be used across multiple technical platforms.

Teachers often earn tenure protections after several years of teaching that have been deemed successful by their evaluations. Other studies demonstrate the impact teachers have on student success, which enhances concerns local evaluation systems, which tend to highly rate the overwhelming majority of their teachers, may make it difficult to remove "bad" teachers (Chetty et al., 2011; Hanushek & Rivkin, 2010). Concerns about the widget effect have gained the attention of politicians and have led to legislative reform, shifting teacher evaluation policies from the

school-level to the state-level. However, despite efforts to center policy at the state-level, evaluation is still enacted at the school-level and centralization may not mitigate the impact of local policy enactors, such as principals, who are often tasked with leading evaluation processes in schools.

The results of other studies and policy analyses have recognized examining teacher perceptions of policy can yield information that is useful for determining the extent to which a policy achieves intended outcomes (Callahan & Sadeghi, 2015; Donahue & Vogel, 2018; Finster & Milanowski, 2018; Firestone et al., 2014; Henry & Guthrie, 2016; Jiang et al., 2015; Sporte & Jiang, 2016). Additionally, the results of recent quantitative studies have identified that teacher perceptions of an evaluation system can widely vary and are impacted by factors such as perceived clarity of the policy, evaluator credibility, fairness of measures, usefulness of evaluation, and feedback quality (Callahan & Sadaghi, 2015; Donahue & Vogel, 2018; Finster & Milanowshki, 2018; Jiang et al., 2015). Moreover, teachers respond to evaluation through the lens of their administration and the way in which principals enact evaluation may influence how teachers perceive the policy (Reinhorn et al., 2017). Therefore, teacher perceptions can provide information on how and in what ways principals influence evaluation policy implementation. A better understanding of how principals grapple with state policy in the context of their individual schools can help improve policies and inform implementation practices.

As with other states at the time of this study, North Carolina's teacher evaluation tool included multiple measures and was intended to serve both formative and summative purposes (North Carolina Department of Public Instruction, 2015). This paper is derived from data from a larger mixed method study (Frasier, 2017). In the larger study, I used state level data (Teacher Effectiveness data and results from the Teaching Working Conditions Survey) to select four high schools in a large district to investigate teacher perceptions of the implementation of the state-level teacher evaluation policy and instrument (scoring rubric). The four schools were selected deliberately to reflect varying combinations of teacher evaluation ratings and teacher attitudes toward teacher evaluations. In the larger study, I conducted a survey of English and Math teachers at four high schools to measure teacher perceptions of evaluation using a Likert-type scale. I also used demographic data to select fourteen teachers who were interviewed in-person twice in the 2016-2017 school year and asked to describe policy implementation in their school. The fourteen teachers' interviews serve as the data source for this paper. The research question I investigate in this paper is: Considering the legislative efforts to remove control of evaluation from local figures, do teachers perceive school principals as influencing the implementation of state-level evaluation policy and instruments, and if so, in what ways?

Background

Traditionally, the evaluation process for teachers in the U.S. has been a local affair consisting of classroom observation and local personnel preferences (e.g., a teacher's ability to coach or teach certain subjects) with limited standardization among the protocols, frequency, or observers (e.g., Tyack & Cuban, 1995). As federal influence has increased via directing state-level policy in schools through both mandates and incentives, some traditionally locally held powers, such as control of teacher evaluation, have shifted and become more centralized, at least in part, at the state-level. Mintrop and Sunderman (2013) describe the federal accountability movement as occurring in three waves, which offer a framework for understanding the progression of school accountability policy. The first wave of school-level testing-based accountability involved

experiments in states and localities, which inspired the second led by President George H.W. Bush and his America 2000 plan and later President William Clinton's Goals 2000: Educate America Act (Vinovskis, 2009). Both plans were precedents for President George W. Bush's No Child Left Behind (NCLB), a renewal of the Elementary and Secondary Education Act of 1965 (ESEA).

The third wave included the Obama administration's 2009 Race to the Top Initiative (RttT) competition and the 2011 ESEA Flexibility Program, which both prompted states to undergo several legislative changes to reform education and encouraged shifting focus onto the accountability of individual teachers rather than schools (Mintrop & Sunderman, 2013). Both programs required states to centralize teacher evaluation policies at the state-level and adopt multiple measures of teacher performance, including a student growth measure where a psychometric value-added model (VAM) that utilized individual student data such as previous standardized testing performance, was used to attempt to determine how much a student improved in their understanding of a content area (U.S. Department of Education, 2009). So, RttT-inspired policies shifted teacher evaluation policies from locally controlled systems to state systems while increasing focus on the effectiveness of individual teachers.

The state of North Carolina during the 2016-2017 school year was an ideal location for examining the convergence of state-level evaluation policy and school-level implementation, because a state-level evaluation instrument that pre-dated RttT was used in North Carolina. By 2009, North Carolina's evaluation system consisted of five observation standards, pre- and post-evaluation conferences between the teacher and their evaluating principal, and a year-end summative conference. Upon the announcement of the RttT competition, a sixth standard for student growth which utilized a value-added model (VAM) of student performance was added to teacher evaluations. So, at the time of this study, all teachers in the state were measured against five observation type standards and one student growth standard (Table 1). Both the observation rubric and the VAM were measures developed by the state, but the observation rubric was administered by local principals.

Table 1
Evaluation Standards

Standards	Type and Ratings
1 Teachers demonstrate leadership.	**Observation (Local)**
2 Teachers establish a respectful environment for a diverse population of students.	Not Demonstrated
3 Teachers know the content they teach.	Developing Proficient
4 Teachers facilitate learning for their students.	Accomplished
5 Teachers reflect on their practice.	Distinguished
6 Teachers contribute to the academic success of their students	**Student Growth (State)** Does Not Meet Meets Exceeds

Literature Review

The research that informs this work includes: (a) literature on how principals serve as policy implementers in schools, and (b) literature on how the capacity to complete evaluation in a timely manner, policy values, and perceived effectiveness impact the implementation of policies by principals.

Principals as Evaluation Policy Implementors

Policies are subject to interpretation and alteration by those who are tasked with enactment in real contexts, resulting in what Lipsky termed "street-level bureaucrats" (2010). Moreover, the results of other studies have demonstrated how school principals influence the enactment of other policies, such as curriculum reform, by influencing teacher sensemaking (the process in which people create meaning of their experiences; Coburn, 2005). For instance, a principal's understanding of a policy can impact the way they participate in and lead social processes of sensemaking, the policy ideas they choose to highlight, and the working conditions they create for teachers (Coburn, 2005). The way in which school principals conduct evaluations and provide feedback can impact the way in which evaluation policy is implemented in schools. Additionally, other studies have shown that teacher perceptions of evaluation are closely linked to perceptions of the evaluating principal (Jiang et al., 2015; Reinhorn et al., 2017) and that the credibility of an evaluator is closely linked to teacher perceptions of the quality of feedback the evaluator provides to a teacher (Finster & Milanowski, 2018). Principals, as the primary evaluator at the school level, become street-level bureaucrats of the policy, and their role in enacting evaluation is important for policy success.

Administrative Capacity

Principals may lack the capacity, in terms of time available, to effectively enact evaluation policy at the school level. Researchers have documented that the roles of principals have shifted over time to include an expanded role as an instructional leader due to changes in both policies and public expectations (Bryk et al., 2010; Louis et al., 2010; Spillane & Kennedy, 2012). The results of other studies suggest the quality of feedback a teacher receives is dependent on principals having the necessary training, time, and resources to devote to provide individualized, actionable guidance (Kraft & Gilmore, 2016; Reinhorn et al., 2017). Due to a deficit of time, administrators may be inadequately trained or unable to provide quality feedback to each teacher. For instance, in New Jersey, requiring more frequent observations under a new policy led to principals trying to enter the results in real time rather than providing actionable feedback after the observation (Callahan & Sadeghi, 2015). So, a principal having the capacity to devote the time to fully implementing the policy influences whether the policy fulfills its intended purpose of providing feedback.

Policy Values

Additionally, principals have varying views on what is most valuable in a policy and may respond to one policy message at the expense of others, leading to varied implementations of the policy (Kraft & Gilmore, 2016; Reinhorn et al., 2017). For instance, when using multiple

measures, an administrator may focus on one aspect of evaluation (such as test scores) at the expense of other measurements (such as observation) or, principals may focus on procedural aspects of evaluating and miss opportunities to provide feedback and inform professional development opportunities (Callahan & Sadeghi, 2015). Additionally, while principals may be willing to label certain teachers as unsatisfactory in informal settings and conversations, they may be less willing to give the same teacher critical evaluations when punitive stakes are attached to such evaluation (Grissom & Loeb, 2017; Kraft & Gilmour, 2016). While one purpose of evaluation may be to rank teachers summatively to inform personnel decisions, Donaldson and Mavrogordato (2018) explained how, despite policymakers' attempts to streamline and standardize teacher evaluations, principals do not use evaluation as a linear process to identify and remove poor teachers. A principal's understanding of evaluation does not occur as a straightforward acting out of policy but is instead complicated by and interacts with factors such as their personal cognitive sensemaking (the process of making personal meaning from experience), relational trust, and organizational capacity. So, implementation is impacted by policy values, or a principal's understanding of an evaluation policy as well as the aspects of the policy on which they focus.

Observer Effectiveness

The evaluating principal's background may also influence teacher perceptions, regardless of whether or not teachers perceive their observing administrator as being effective in executing the policy. Teachers may not feel the feedback from an observation is useful or even valid if they lack confidence in the principal's ability to enact the policy in a way that would yield useful information to improve upon the teacher's classroom practice. However, even a principal who is perceived by teachers as being a fair observer may still be viewed as lacking the effectiveness to effectively evaluate a teacher due a mismatch between the principal's background and the teacher being observed. For instance, it is common for principals to have subject area or grade level experience different from the teacher being observed. A principal's content knowledge influences the way they observe teaching and provide feedback to teachers, with principals being more likely to critique curriculum in a content area that aligns with the principal's background (Coburn, 2005; Nelson & Sassi, 2000). A principal who was formerly a math teacher will focus on different aspects of teaching when observing a math teacher who is teaching content familiar to the principal rather than an English teacher whose content may be less familiar to the observing principal. Additionally, principals report feeling less comfortable providing critical feedback when evaluating teachers outside of their grade level or subject area background, which may result in feedback that is narrowed to certain aspects of teaching, such as classroom management, or other aspects that may be considered less important to the teacher (Kraft & Gilmore, 2016). So, teacher perceptions of principal feedback may be influenced by the principal's teaching background relative to the subject area being taught in the observed classroom.

Researcher Context

I was well-prepared to approach this type of research due to my background as a National Board Certified Teacher with five years of experience in public high schools, including four years of experience in the state of North Carolina. I had some firsthand knowledge of the evaluation policy in this study, as I was teaching in North Carolina when the statewide observation system was adopted and when the student growth standard was added. I went through the same state-based

training on the evaluation tool as the teachers in my study. In fact, my own experiences with this policy roll out is what led me to later pursue this line of research as a graduate student.

When I was in training as a teacher for the roll out of the evaluation policy in this study, members of the North Carolina Department of Instruction came to my school and presented the policy as being highly standardized and fair. The evaluation instrument (rubric) was presented as being not only a reliable tool for rating teachers but also as a useful source of feedback for teachers to improve instruction. Even then, five years before I later defended a dissertation on the larger study for which this paper is derived, I questioned how principals may influence the implementation of this supposedly "standardized" policy and how differences in implementation may ultimately impact teachers. My personal experience led me to suspect that the evaluation instrument could not be completely standardized, because it was being implemented by principals, who are humans with their own perspectives and values that would interact with the enactment of the policy.

I felt if the evaluations were to be used in the high stakes manner North Carolina was proposing at the time, where evaluation results could impact teacher employment, then the way principals implemented the policy could have serious and perhaps negative consequences for teachers. Furthermore, my experiences as a teacher led me to anticipate how the policy might be implemented and how it might impact teachers, which influenced the literature I sought for the study as well as my study design. For instance, as a teacher, I had felt pressure to engage in practices such as drilling students with practice tests when I had a class that would be taking a high-stakes exam, so one literature base I examined was how evaluations could influence changes in teacher practice.

My experiences also granted me greater awareness of the policy atmosphere in which I was investigating and allowed me to effectively engage with interview participants because I already understood what they had gone through in training. As such, my interview participants tended to view me more as a fellow teacher rather than a researcher.

I recognize bias occurs unintentionally, and I constantly acknowledged how my past experiences, particularly as a teacher in the same state as my study, may have impacted data collection. I ensured neutrality in my larger study by writing and reviewing my interview and survey questions beforehand to help ensure I asked non-leading questions and allowed opportunity for clarification from participants. I piloted both the survey and interview questions with other North Carolina teachers in different school systems prior to data collection. Also, the use of multiple types of data in the form of surveys and interviews served as a validity check (I used my initial survey data to triangulate the later interview data; Stake, 2004). My study also features a multiple case design by including various groups of teachers (Math and English teachers, provisionally and professionally licensed, teachers from four different school sites, etc.; Yin, 2009). Additionally, the second set of interviews, which was conducted several months after the first round, served as a type of member check to obtain feedback on the themes and typologies that emerged from the first round of interviews (Deyhle et al., 1992).

Methodology

School Sites

My study was determined to be exempt by my institution's Institutional Review Board (IRB) under Category 2: Educational Tests, Surveys, Interviews, and Observations because IRB

determined that my adult participants could not be easily identified through my data. Additionally, I had to seek approval from the school district by submitting an application and research proposal. My point of contact for district approval was the Director of Testing and Accountability. Once I had district approval, I sought the approval of school principals at the sites where I wanted to conduct research. This permission process was conducted via e-mail. Next, I contacted the department heads of the English and Math departments at each school, who each agreed to participate and negotiated how I could have access to their teachers. For instance, at the school level, different departments asked me to work with them in different ways. Some departments wanted me to administer my survey instrument, which was used in this paper to select participants and guide interview questions, via e-mail whereas other departments asked me to come to a department meeting to speak with teachers about my study and my interests. Different departments also requested different types of tokens of appreciation to allow access to teachers. Once each department reached a certain threshold of teachers that had completed my survey, I conducted a drawing for a $25 Amazon gift certificate. I had a high rate of completion for the survey (76.27%), so this technique was well-received. Additionally, the teachers who were selected for interviews were monetarily compensated $50 for their time after completing the second interview.

The interview participants in this study were fourteen high school Math and English teachers across four high schools in Broadville County (pseudonyms used throughout). The sample of schools was intended to represent a variety of evaluation conditions and Educator Effectiveness scores. To achieve this, I used publicly available data from evaluation related questions on the 2015-2016 North Carolina Working Conditions Survey and 2015-2016 Educator Effectiveness Database to create Evaluation Condition Scores and Educator Effectiveness Scores for the district. I then created a score for each school in the district by measuring the difference between the school's score and the district average. I selected the four schools intending to represent differing combinations of conditions and scores. Once the schools were selected and had agreed to participate, I worked with the English and Math departments at each school to administer a survey to teachers.

Data Collection and Analysis

The larger study from which this paper is derived was a mixed methods investigation I completed for my dissertation and included the following two research questions:

(1) What, if any, role do report school evaluation conditions and school evaluation status play in shaping teacher motivation, experiences with feedback, and work decisions related to teacher evaluation?
(2) What individual-teacher level factors are associated with differences in teacher motivation, experiences with feedback, and work decisions related to teacher evaluation?

English and Math teachers were selected because research question 2 addressed whether teachers of those subject areas viewed evaluation differently. After graduating, I created separate publications based on my research questions, but while I was writing about the first research question, I saw a new and previously unaddressed topic emerge in the interview data and began forming this article addressing the question: Considering the legislative efforts to remove control

of evaluation from local figures, do teachers perceive school principals as influencing the implementation of state-level evaluation policy and instruments, and if so, in what ways?

For my dissertation, I administered a survey to teachers at the start of 2016-2017 school year. Demographic data, such as the school location, years of experience, and subject area from the survey were also used to select interview participants. I attempted to sample two English and two Math teachers from each school, but I was unable to achieve uniform sampling. For this paper, the survey results were used to help develop questions in the interview phase.

I conducted the first round of interviews in-person and on location at the participants' schools two months into the school year. During the semi-structured interview, I began each by asking teachers to generally explain their experiences with evaluation both in the past and in the current school year. I sometimes used the individual's responses on the survey that was administered for the larger study to move the conversation along. If a conversation stalled, got off topic, or if a teacher was not forthcoming, I sometimes asked prepared clarifying questions specifically about answers individual teachers had provided on the earlier survey. However, I generally let each interviewee steer the conversation, and I would follow-up with questions. While the interview had limited structure and was conversational, I ended each interview by asking every teacher two structured questions: (1) Are evaluations necessary because teachers need to be rated, sanctioned, or rewarded in order to be motivated to do a better job? and (2) Do evaluations yield information that is useful for teachers to improve practice? (A copy of my first interview protocol is available in Appendix A).

I conducted follow-up interviews in-person, on location at the interviewee's school in early March of the 2016-2017 school year. Due to block scheduling, teachers had completed one testing cycle in January and were teaching entirely new courses. Every teacher had been observed by their principals and evaluated on the state created instrument at least once. I began the interview by asking teachers for an update on their observations and the recent testing cycle. I inquired about the teachers' courses in the current semester and asked if they felt any differing pressure versus the prior semester. Finally, I shared with each teacher their school's Evaluation Conditions and Evaluation Effectiveness scores as compared to the district mean and asked teachers to explain if these quantitative scores surprised them or if it was an accurate reflection of their school. Finally, I asked teachers if they thought anything in the current school year would alter those state level scores or if they would expect them to stay about the same after the current year. The final question was an attempt to get teachers to identify any new circumstances at the school that may be contributing to changing attitudes and perceptions about evaluation. A copy of my second interview protocol can be found in Appendix B.

Interviews ran from 20-50 minutes and were transcribed in entirety. The transcripts were then uploaded in Dedoose to be coded. The larger study did not explicitly examine the role of administrators in evaluation and initial coding was guided deductively by the literature reviewed for my dissertation. For the initial study I utilized Dedoose's upcoding feature to first code excerpts based on codes I had deduced from my literature review. I later inductively developed new parent codes and child codes as trends emerged beyond those established by my initial literature review. In Dedoose, parent codes are original level codes and child codes are additional codes added to excerpts within those excerpts that received a parent code. Using a child code will automatically "upcode" an excerpt to the parent code as well.

As I was turning my dissertation into publishable works, I noticed administration mentioned frequently under the following codes from the original study: timeliness, consistency, breadth, and expertise of observer. Despite not being explicitly asked about their principals' role

in evaluation, every teacher spent time in the interview describing the policy's implementation by the principal. So, a new round of coding of the interview data was developed inductively and iteratively for this article (see Appendix C). I first created a new parent code to identify when teachers mentioned administration implementing the policy. All mentions of administration were coded using the code "administrator." I then reviewed the excerpts that had been coded "administrator" and identified three themes in those excerpts: administrative capacity, policy values, and observer effectiveness I created child codes that corresponded to those three themes and coded the excerpts as such. In this manner, all excerpts which received a child code of administrative capacity, policy values, or observer effectiveness had also received the parent code of administration.

The literature review I conducted for this paper supported the creation of these child codes. I conducted the literature review simultaneously while coding my excerpts, so my themes were informed by the trends I noticed in the interviews while also being influenced by literature. This was accomplished by first reviewing all the excerpts I had coded as "administrator" and making memos in the form of a list of topics I thought teachers were saying influenced evaluation at their school. I then compared this list to the literature I was reviewing and was able to form three major "bins" or topics that became my three child codes.

Codes were not mutually exclusive, and excerpts could have many codes. For instance, one excerpt that had the parent code "administrator" could be coded both "administrative capacity" and "policy values." It should be noted that this secondary qualitative analysis that occurred after my dissertation allowed me to identify possible explanations for my data that would otherwise have gone uncaptured in the larger mixed method study. For instance, a hypothesis in my larger study was that by picking schools that had quantitative differences in the publicly available Teacher Effectiveness data and Teaching Working Conditions Survey, I would see significant differences in teacher responses between school sites on the survey responses. Yet, there were no statistically significant differences between the schools. However, the qualitative data in the interviews tells a different story where school principals influence the implementation of the policy at the school level. The role of the school principal and the nuanced differences between schools as expressed in the interviews were missed in the initial quantitative analysis but captured when I revisited the data qualitatively. More detailed explanation of my coding scheme can be found in Appendix C.

School Cases

All four schools met the basic legal and procedural requirements of the state-level teacher evaluation policy; however, the interviews revealed that there were differences in how teachers perceived their principal's policy implementation. Each school will be briefly described.

Riley High School: "Feels Like Entrapment"

Riley was selected for this study because the publicly available working conditions and effectiveness scores were close to the district average. Thus, it was meant to represent a school with an average attitude toward evaluation and teachers who were average in effectiveness. I interviewed two English teachers at this school and no Math teachers were available for interview. During the study year Riley had a new principal, Ms. Jefferson. Ms. Jefferson required the submission of daily lesson plans which she initially frequently reviewed and on which she made

comments. In addition, Ms. Jefferson completed a full observation, complete with formal conferences and as well as additional informal discussions about lesson plan submissions, during the first month of the school year. For instance, Mr. Donaldson, an English teacher, was formally observed three times before our first interview in mid-October. Observing each of the nearly 100 teachers was an impressive accomplishment and required a considerable investment of time.

The teachers were critical of the new administrator's approach to observation. The teachers perceived that the new administrator equated the completion of paperwork with good teaching. Initially, both teachers stated that they received increased feedback. As Mr. Donaldson surmised:

> This year, I've been observed three times already, which is fine, it's great to get feedback. The frequency, yeah, it has skyrocketed. Yeah. I mean, well, everybody's super kind in their evaluation, it really is... Being able to actually get some feedback is a nice thing.

Mr. Donaldson also explained how the increase in observations had been coupled with the submission of formal lesson plans which were also being commented on: "She's implemented also a weekly lesson plan submission that she'll comment on and... Yeah, to actually have the feedback to where it's not just seeming like it's another box to check. Yeah, that's been good. I've appreciated it, for sure." Mrs. Macdonald initially said of the principal:

> The woman who's evaluating me right now is very strict. For instance, she took 12 of [the standards] and emailed me and said, "I want you to respond how you reflect [on] each of these standards on a daily and on a yearly-basis." One of the standards was like "Differentiate your lessons." And to answer how I do that on a daily and then on a yearly basis was pretty bogus. So, I probably had to write an additional hour of paperwork to receive my evaluation… and then she took my responses and she put together probably the most detailed evaluation I've ever had in 12 years of teaching.

However, by the second interview, both teachers had started taking advantage of Ms. Jefferson's tendency to focus on paperwork completion over content and stopped devoting as much time to submitting lesson plans. For instance, after the first few lesson plan submissions, Mrs. Macdonald stopped receiving feedback and had since invested less time in the process. Mrs. Macdonald was also skeptical that the very high marks she received on her initial observations were an accurate reflection of her teaching. She stated that the feedback she had received lacked substance and suggested Ms. Jefferson was measuring her ability to meet deadlines rather than "actual teaching ability."

Meanwhile, Mr. Donaldson was not observed by Ms. Jefferson again between the interviews, and he stated he had started "recycling" lesson plans on his daily submissions, which he said went unnoticed by the principal. Mr. Donaldson stated he observed "pushback from some quarters" of the teachers regarding the increased frequency and intensity of observations, and some teachers felt the approach was meant to catch them doing something wrong. He stated he had overheard others talk "of dissatisfaction, or of people feeling like there's an entrapment factor" to the principal's observation procedures. Mrs. Macdonald said of her principal's approach to evaluating teachers:

I don't feel like I've grown at all as teacher. And I've done it for how many months, eight months? It's an hour every Sunday night. I'm not getting any feedback on it. We're just, the teachers who aren't doing it are getting reprimanded.

I asked Mrs. Macdonald if she felt like she was being rated fairly and she said her evaluations were very positive, "I feel like she is rating my positivity on whether or not I've done my paperwork. That is all."

Phoenix Alternative High School: "A Culture of Learning and Growing"

Phoenix was an alternative school that specialized in teaching students who were failing out of or otherwise unable to perform in traditional high schools. Phoenix had the most positive view towards principal evaluation, and its teachers ranked as average on effectiveness as measured with publicly available data. Two English teachers and one Math teacher were interviewed at Phoenix.

During our interviews, Mr. Brown, a second-year English teacher who was new to Phoenix, often compared his first-year teaching at another school to his current year at Phoenix and said the difference was related to the school administration. The day of our first interview, Mr. Brown found out he was scheduled for his first observation of the year in the next week. He stated that at his previous school, he had felt anxiety about his evaluations and perceived extreme pressure on teachers to achieve high test scores. He stated everything in his observations seemed to be linked back to state testing and described his first year as being "micromanaged by policies meant to increase student achievement on tests." He expressed concern over how student achievement on tests would factor into his observation at this new school, as students attending alternative school do not usually produce high test scores.

I observed a noticeable difference in Mr. Brown's anxiety about evaluation between his first and second interview. In the first interview, Mr. Brown expressed nervousness about how evaluation would play out in the unique new classroom context he now taught in as he previously had faced retribution if his students did not perform well. He stated:

So, I feel like that kinda pressure on, you know, the kids, you gotta get them up two percentage points, and it's just like that kinda pressure is off which is one reason it kinda drew me to this school. I know there's some challenging situations with students, but at least it's actually focused on students, not tests.

By the second interview, Mr. Brown said he had accepted testing did not matter in the same way at Phoenix as it did at his previous school. He surmised:

[There] are bigger fish to fry here [than academics], especially on the social and emotional level. Some of these kids are dealing with a lot [which] matters more in the grand scheme of things. Some of these kids need to have social skills as opposed to knowing how to take a test.

Mr. Brown stated the administration at Phoenix did not have low expectations, but instead, "They understand, in fact, they really understand what the teachers are dealing with, like how a classroom looks. And I feel like their perceptions, they align maybe pretty well with the teachers."

Mrs. Street was a veteran English teacher who previously served as a curriculum coach at Phoenix. She returned to the classroom to finish her teaching career with a few years left before retirement. She talked extensively about the autonomy afforded to teachers at Phoenix, which was something she felt did not occur at other schools. She explained:

> We have guiding principles for our school, but how I wanna meet those students' needs; my principal, Mr. Conard really leaves that up to me as a professional. And I work with other teachers, with Special Ed., with special education teachers, the curriculum coach, other people in the county as well.

Mrs. Street stated the principal of Phoenix allowed and encouraged teachers to try new things to reach the unique population of students they served. Mrs. Street described how she had felt in other schools, particularly during observations:

> I would be very nervous to try new or out of the ordinary [methods]. I would stick to something more scripted, something tried and true. Here we have the freedom. We are not going to be marked down for trying a strategy or trying something with students and it fails.

The teachers interviewed for the study also brought up the culture of Phoenix, which Mrs. Street described as being one of "learning and growing," where every teacher is willing to accept feedback from the others. Mrs. Street explained, "[F]eedback is necessary… it is not a bad thing or just a good thing; it is how can we all learn from each other." Mr. Forest, a Math teacher, described conversations he had with colleagues, "I feel like I can walk into [my administrator's] office at any time, and I wonder if my colleagues at other schools feel the same way."

Central High School: "Just a Joke Here"

Central had a very negative view of evaluation when compared to the district average, but the teachers measured above average on their effectiveness scores. The interviews revealed there were differing observation experiences between the Math and English departments. Math teachers were observed for several consecutive years by the main principal, Mr. Nichols, who had several years of experience as a Math teacher prior to becoming an administrator. In contrast, the English department was observed for several consecutive years by an assistant principal, Mr. Reward, who allowed teachers to complete their own evaluations. There were three English teachers and two Math teachers interviewed at Central.

The three English teachers felt their department was strong and did not require a lot of oversight. Mrs. Hoard stated, "we have really, really strong PLCs (Professional Learning Community meetings), and that has developed over the past three years or so, so I think that definitely could be a big part of that, and I think that also gives us confidence." One teacher, Mrs. Williams, attributed Mr. Reward's assignment to the department as a testament to teacher skill and explained she felt the principal did such a good job at hiring that the teachers at Central did not need to be evaluated:

> We're a pretty strong department… I think that our principal is really, really good at hiring. And he hires people that are so good that he doesn't have to chase them

around. And he also knows that, even though he wouldn't admit this, I know more about this subject than he does, and he hires people; he wants me to be the expert in this room. He doesn't wanna be the expert.

All three English teachers described evaluation under Mr. Reward in a similar way: the assistant principal would sit in for part of a class and then largely leave the assigning of scores and comments to the observed teacher. Mrs. Williams described taking over the computer from Mr. Reward and typing up the evaluation for him, "And I think I'm fair," she added.

One English teacher, Mrs. Hoard, described observations similarly, "So the evaluator I have right now is very, very laid back and pretty much I go in there and we talk, and he writes down whatever I tell him to write down." Mrs. Hoard also described an observation that did not occur with students. Instead, Mr. Reward completed the evaluation after he observed her conducting a department meeting. She explains:

> I was one of the teacher leaders for professional development this year within our school…and it was a really, really quick lesson. They were basically breakout sessions, and teachers were coming just to get different strategies, kind of in little 20-minute chunks. So, he saw a 20-minute little snippet… And he sat here in this classroom and filled out his form on the computer, and that was pretty much it. And then we signed off on it later.

As a result, Mrs. Hoard was not formally observed teaching students at all during the study year. Mrs. Hoard said she did not like that the observations were conducted so "haphazardly," but felt secure that the results of her evaluation would have no effect on her professionally. One of the Math teachers indicated in his interview that he had been observed once under Mr. Reward and described similar experiences, so the approach may have transcended subject areas.

Conversely, the Math department described slightly more positive experiences with observation due to the quality of feedback they received from Mr. Nichols, who had been a Math teacher. The Math teachers described how Mr. Nichols would identify things in the observed lesson that may have otherwise gone unnoticed by the teacher or make suggestions that were practical and could be used to improve instruction. Both Math teachers expressed gratitude that they had an administrator who, as Mr. Williams stated, "knows the Math" and could effectively identify if something went wrong. Overall, the Math teachers seemed to feel the feedback received from observations was valid due to their administrator's background as a Math teacher. Mr. Proffit explained, "That's always been so much more helpful than someone who has never been a Math teacher… because we can discuss things."

I asked the teachers to help explain the negative impression of evaluation at their school. Aside from the discrepancies with observations, teachers took issue with the student growth standard. Teachers stated the consensus in their school, including from administration, was testing was, as Mr. Proffit phrased it, a "fact of life," and they wanted students to do well on tests; however, the teachers overwhelmingly did not see value in the tests as sources of feedback or as valid measures of student gains. The teachers universally claimed they felt their principals felt similarly about testing. As Mr. Augustus surmised, "Evaluation is kind of just a joke here. We go through the motions, but no one, not even our administrators, expect it to be more than just checking another box."

Charles High School: "Cover Your Behind"

Charles High School measured slightly below average in terms of teacher's favorability toward evaluations and below average teacher effectiveness scores. Charles was the smallest of the three traditional high schools and featured an initiative to improve Math scores where students took an Introduction to Math course prior to taking Math I. Math I test scores contributed to the schoolwide growth score while the introductory course counted as an elective for students. Of the four high schools in this study, teachers at Charles described a focus on testing results that was much more intense than the other three high schools. There were two Math and two English teachers interviewed from Charles.

Mrs. Ranier, an English teacher, stated that due to the testing focus, she felt the faculty at Charles was "analytical and cynical," but said there was a disconnect between what the principal and teachers thought needed to be done. She explained:

> This is an interesting faculty that even in what I think of as the arts people who I don't think of as logic in their stream, like English and visual arts, even members of those parts of the faculty are very analytical… I think it's because our faculty, we don't have a lot of non-analytical people. Even the teachers who are new to the school are not new to teaching. They've been teaching at least four years, most of them more, and when you've been teaching that long, you know you don't get a lot of information [from evaluations].

Mrs. Ranier attributed the issue to the principal's lack of experience, as Mrs. Warner had limited teaching experience before becoming an administrator. Mrs. Ranier felt she had to engage in a lot of required "cover your behind" activities that the principal expected of teachers at her school. She stated, "I really liked it better… when I was trusted to do my job and do it well. And I think I did a better job because I was less anxious, there was less stress."

Mr. Eagle, an English teacher who had previously taught out-of-state but was in his first year at Charles, described a general feeling of "being watched" four times in the first interview, "I think people feel like there is this mentality that you're being watched, especially if you've been put on whatever kind of list of like, 'This isn't a good performing teacher,' you're watched incredibly closely and evaluated harder." While Mr. Eagle felt the principal seemed very enthusiastic about his performance as a teacher, he acknowledged there seemed to be an "invisible list of bad teachers" who were watched more frequently. He also noted an intense focus on student growth on tests. He had good assessment scores his first semester with his seniors, which he attributed to pure luck, and he had been approached about teaching an End of Course (EOC) class the following year because the principal had identified him as someone who had the potential to produce high test score gains from students.

Some of the views of the two English teachers were echoed by the two Math teachers who talked about the testing results-driven atmosphere of the school; however, the Math teachers did not express the same pressures and fear of oversight that the English teachers expressed. There are three possible explanations for this. First, the closeness of the departments differed in professional relations, personal relations, and physical location. The Math department consisted of veteran teachers and was very tight knit, even meeting on the weekends to play board games. In contrast, the English department consisted of many newer teachers, including both English teachers interviewed for this study. The English department was also spread across the school instead of

housed in adjacent classrooms in one hallway like the Math department. Secondly, the Math teachers were not observed by the principal, but were observed by an assistant principal who had a Math background, and the teachers expressed confidence in his ability to evaluate. Finally, aside from students taking an Intro to Math course, the Math teachers also received a lot of outside support in the form of a curriculum coach, whereas English teachers were primarily receiving feedback from the administrator, who reportedly did not have robust teaching experience.

Findings

Analysis of the data revealed that despite policy attempts to mitigate the impact of local actors, teachers perceived the implementation of evaluation policy differently depending on their principal. There were three themes which illustrate ways in which teachers perceived principals impacting the success of the policy: administrative capacity, policy values, and observer effectiveness. These themes emerged from the coding of the interview transcripts and are supported by the results of other studies as reported in the literature review. As mentioned in the methods section, because the research question in this article emerged after my dissertation, I conducted the literature review while simultaneously reviewing excerpts of interviews that were related to principals. Therefore, codes were created iteratively but further influenced by the new literature I was examining as I was looking for trends but also checking what I was seeing against the work of others.

Administrative Capacity

The overall success of evaluation policy to both rate teachers and provide feedback hinges on a principal's ability to enact and support the policy and the theme of administrative capacity surfaced in all four schools. Teachers specifically mentioned principals not devoting the time needed to conduct evaluations in a way that could accurately assess teaching ability and provide useful feedback. Teachers in two schools, Riley and Phoenix, mentioned frequent administrative presence in classrooms. These schools are the largest and smallest schools in the study, respectively, and while both administrators devoted time to observing and evaluating their teachers, the evaluation experience as described by teachers was vastly different.

Ms. Jefferson devoted much of her time as a new principal at Riley to observing her staff and providing oversight on lesson plan submissions; however, her feedback was limited to processes and not to instruction. According to the teachers from Riley, Ms. Jefferson began narrowing her focus to teachers who struggled to meet deadlines. According to teachers, Ms. Jefferson had a formalities-driven approach where an adherence to procedure was valued rather than quality of work. All the evaluations she conducted followed formal process, and she did not enter classrooms informally. Ms. Jefferson was able to demonstrate the capacity to evaluate her teachers, but the teachers I interviewed felt her approach failed due to the values she emphasized, which will be discussed subsequently.

In contrast, teachers at Phoenix felt supported by administration and worked in an atmosphere of learning and improvement guided by their administrator. While teachers did not necessarily view evaluation, and particularly the testing component, in a positive manner, they were satisfied with the way evaluations were conducted at their school, with the feedback they received, and with the administration's approach to the policy. Teachers felt administrators were familiar with the work of the teachers and the evaluation was already engrained in current practices

because the principal was often informally in the classroom. Aside from Riley, Phoenix was the only school that required lesson plans. It was also the only school where teachers mentioned administrator presence in the classroom aside from evaluation. Overall, while the staff at Phoenix followed the evaluation policy, they were free to work together to create definitions of success in meaningful and supported ways.

The interviewees at Charles perceived administration did not devote enough time observing teaching to accurately gauge ability, but there was a heightened focus on the testing component of evaluation. Also, teachers at Charles perceived that their test scores would influence their observations, which will be discussed more in the next section.

At Central, the teachers perceived the evaluation was perfunctory and administrators did not care about the policy, only performing the tasks of evaluation to, as one teacher described, "check a box." The case of Mr. Reward, the assistant principal who allowed teachers to complete their own evaluations or who did not observe teaching at all, is illustrative of that perspective.

Policy Values

The aspect or aspects of the policy on which principals chooses to value (or devalue) also influences teacher perspectives on the policy. Specifically, what an administrator valued in the policy, how those values were interpreted by the principal, and whether those values seemed credible to teachers, mattered to the teachers across all four schools.

For instance, what an administrator chooses to value in teachers and teaching may matter greatly in how the observation protocol gets interpreted, as was the case with Riley's new principal, Ms. Jefferson. According to teachers, Ms. Jefferson valued coherence to a process in her evaluation and her feedback was limited to the process. The teachers reported Ms. Jefferson tended to equate "good teaching" with the submission of lesson plans and teachers expressed concern that this impacted how observation ratings were assigned. According to the teachers, Ms. Jefferson placed emphasis on the completion of tasks rather than on what she saw happen in classrooms. Teachers stated they felt like this was a "gotcha" situation and other aspects of their teaching were not taken into consideration in their evaluations. Further, the teachers, Mrs. MacDonald and Mr. Donaldson, stated they began to "game" the system by recycling lesson plans or changing the plans to fit their own purposes. While the principal seemed to have a lot of time to devote to evaluating her teachers, the teachers did not feel the way the principal was interpreting the criteria of the evaluation was credible. The situation at Riley demonstrates principals can choose to prioritize certain actions of teachers or interpret the observation instrument in a way that allows for such prioritization.

Similarly, the teachers at Charles perceived that administration valued testing data over what was observed in classrooms and expressed that they could receive better observation scores if they produced high test scores in students. Furthermore, this might have impacted the English and Math departments differently. Mrs. Ranier, an English teacher, perceived a challenging, test-focused workload and lack of support from administration that expected teachers to cover their own "behinds" and claimed those conditions negatively impacted the motivation of the English teachers. The English Department was less supported than the Math Department. English teachers were spread throughout the school, did not know each other well (the teachers I interviewed did not know each other's names at the first interview), consisted mostly of newer staff members, did not utilize a curriculum coach, and did not have an introductory course to help students prepare for testing. English teachers felt like administration was valuing the "wrong aspects" of their teaching and that they were unsupported in trying to produce good test scores. The English teachers

expressed a fear of retribution from administration if they demonstrated poor performance and both English teachers at Charles were the only teachers in the study who talked about leaving the profession.

In contrast, the Math teachers at Charles also felt that test scores were the "wrong thing" to look at for evaluating teachers, but they were given supports to help boost achievement. For instance, the work of Math teachers was aided by the Introductory Math courses, which were meant to increase student success on the Math I test, and by a curriculum coach who was able to aide teachers with instructional decisions. Additionally, the Math Department all had classrooms in the same hallway and the teachers were personally close. During an interview with a teacher, another Math teacher came in to borrow a board game which he was going to set-up for a "game night" with the other Math teachers that evening. Neither Math teacher expressed the professional concerns or distress the English teachers described, so while Math teachers may have felt testing was the "wrong" focus, they did not perceive the focus as an obstacle to their work.

What a principal chooses to devalue in policy also matters. The teachers at Phoenix agreed the testing component of evaluation held little value in the context of their school, and the principal's actions supported that perception. Teachers at Phoenix were autonomous and did not perceive that administration equated test scores as effective teaching. While state policy allowed alternative schools, which have highly transient student populations and would have difficulty obtaining test score growth with students, to be exempt from the traditional calculation of the growth standard for school effectiveness scores, it did not for teacher effectiveness scores. In other words, alternative schools were not penalized if their students did not achieve satisfactory growth on standardized tests, but the evaluation scores would still be attached to individual teacher's evaluations as feedback on performance. The teachers felt the administration at Phoenix made feedback an ongoing process that was not driven by the values of the evaluation tool. Overall, teachers at Phoenix did not value the evaluation tool, but they did value the observation process and valued feedback that would allow for improved practice.

Observer Effectiveness

The perceived effectiveness of administrators mattered to teachers, and this was most obvious in the case of Central. The English teachers at Central described their observer, Mr. Reward, as an administrator who did not possess the skills needed to complete the observation instrument properly. Mr. Reward would often ask teachers to complete their own evaluation ratings and in one case, conducted an observation at an inappropriate time. A teacher described how an assistant principal had made serious mistakes with overseeing testing which teachers felt had impacted student test scores and may have further contributed to negative feelings toward evaluation. Additionally, teachers perceived administration as dismissive of evaluation and teachers viewed their principals as being unable to provide accurate or useful feedback, which is a stated policy goal of evaluation.

There is also some evidence that suggests the effectiveness of a principal in the subject area being observed may also matter, particularly regarding the quality of the feedback received. For instance, the Math teachers at Central and Charles described situations under which they had been observed by an administrator with a Math background and they expressed gratitude that their principal "knew the math" and could validate their ability as a math teacher. Similarly, despite a positive culture of feedback, the Math teacher at Phoenix mentioned he wished his principal could fully recognize the innovative things he was doing with Math in his classroom and was writing a

small grant with other teachers in the district to provide training for principals to better understand the subject area.

Additionally, teachers who had positive experiences with curriculum coaches described receiving useful feedback that was directly relevant to their work in the classroom. Along with having an observing principal who had been a Math teacher, the Math teachers at Charles also had frequent observations from a curriculum coach. In contrast, the English department was observed by an administrator with limited teaching experience and had no curriculum coach. It is interesting that there was such a discrepancy between the two subject areas in the same school, and it is possible that the principal's background as a Math teacher helped drive school policy changes that supported Math teachers without similar consideration to the English department. While it may not be feasible to always have an administrator with the same subject area background as the teacher being observed, having supports for the subject area in place, which include a source of feedback from an informal observer, like a curriculum coach, may help improve the implementation of evaluation policy.

Success of Policy Implementation

The evaluation policy at issue in this study had two purposes: to serve as a summative assessment of teacher effectiveness and to provide ongoing feedback for improvement. The goals seemed best met when teachers perceived their administration was successful in each of the three themes that emerged in the data. Based on teacher perceptions, Phoenix's principal had the most success in implementing evaluation policy. The principal stressed values of the policy that were sensitive to the school context and in agreement with those of the teaching staff. Additionally, the principal took time to observe teachers frequently and provide feedback both formally and informally. While the Math teacher did mention that the principal did not "know the math," the principal was deemed effective as an evaluator due to his understanding of the school context and ability to facilitate improvement.

According to teacher perspectives at the other three schools, the other principals had varying levels of success with the policy. At Central, the purposes of the policy were largely ignored and there was reportedly a shared, negative view of the policy overall. The requirements were completed but at the bare minimum. The principals did not demonstrate the capacity to implement the policy and effectiveness was certainly questioned, not just in subject area experience but also in a basic ability to fulfill the policy requirements. So, teachers may have felt their ratings were accurate, sometimes because they were self-assigned, but they did not perceive the feedback being useful.

The principal at Riley demonstrated the capacity to implement the policy; however, what she valued in the policy was at odds with what the teachers valued, and her understanding of teachers and teaching, and thus her effectiveness, was questioned. Teachers at this school felt the ratings were inaccurate and the feedback was not useful due to the principal targeting the wrong values in the policy.

Finally, at Charles there appeared to be uneven success for the policy. The principals at Charles seemed to demonstrate some capacity to implement the policy, as demonstrated by the policy supports given to the Math teachers. The role of the principal's background (as a Math teacher) in developing those supports is unclear but may be impacted by his effectiveness in the subject area. However, the English department was not granted similar supports and had an administrator with very limited teaching experience. It is possible that the principal did not

understand how to support the English department and thus, the policy was unevenly applied across subjects and teachers in the two subject area reported different levels of success with the policy.

Discussion

The findings of this study raise several implications for both policy and practice and raise the need for additional lines of inquiry. Additionally, my findings support prior work, such as that done by Coburn (2005) and Finster and Milanowski (2018), demonstrating the impact principals have on policy implementation and extend this work by showing this impact remains in a context where state policy attempted to intentionally curtail local influence. The state-level evaluation policy in North Carolina was meant to be implemented universally across schools and to correct the issues raised by critiques of locally developed policies (e.g., Weisberg et al., 2009). State legislation instituted strict rules for schools and principals to comply with the policy and required the process and scores to be recorded at the state level, yet the teachers interviewed in this study did not perceive evaluation being implemented in the same way across schools. In other words, principals do act as street-level bureaucrats and implement the evaluation policy in differing ways (Lipsky, 2010). In addition, my findings shed light on the ways in which capacity, policy values, and effectiveness contribute to whether evaluation policy implemented by principals successfully rates teachers and provides useful feedback. There are three implications from these findings.

First, if policy is to be successfully implemented in schools, principals must devote ample time to enacting the policy. Teachers in my study reported more success with the policy when principals were familiar with the work of their teachers outside of formal observation and felt this knowledge resulted in ratings that were more accurate and feedback that was more useful. However, principals do not have infinite amounts of time to devote to implementing evaluation policy and must make choices about which aspects of their job to emphasize and to what extent. For instance, the principal at Riley attempted to devote large amounts of time to observing her teachers and providing feedback, but by the second interview the teachers felt feedback was no longer useful, and the principal was no longer able to fully provide feedback. Other principals, like Mr. Reward at Central, devoted minimum effort to evaluation. Mrs. Williams lauded how great Mr. Reward was with buses, athletics, and handling student disciplinary issues, but the same devotion was not applied to teacher evaluation. It is possible that Mr. Reward was allocating little effort and time to evaluations because he had so many other responsibilities.

Secondly, there are several aspects of policy that principals can focus on, and while it may be difficult to devote equal focus to all aspects of the policy, principals should consider how their foci may impact the work of teaching. If the goals of the policy are to provide accurate ratings of teacher effectiveness and to provide feedback to improve practice, then administrators should consider focusing on aspects of the evaluation that will facilitate work towards those goals. If the value of what the principal chooses to focus is unclear or unimportant to teachers, then teachers may find the ratings to be inaccurate and the feedback to not be useful due to a focus on what some teachers in this study termed, the "wrong things." While the evaluation policy at issue in this paper was uniformly implemented at the state and district level, the policy was differently implemented by administrators across schools, and in the case of Central, across departments within a school. These attempts at standardization are not inherently a bad thing, but it should be clear to all stakeholders that evaluation is going to be conducted differently and that is why a clear expression of the expectation and the purpose of evaluation by the principal who is tasked with implementing

the policy is important. Clearly delineating expectations and goals could make the process more valuable to teachers.

Finally, the effectiveness of an evaluator is important. Neither the ratings nor feedback provided by evaluation will seem valid to teachers if they perceive their evaluator as ineffective. My results suggest that principals need the necessary training to successfully implement evaluation policy. Furthermore, while it may be impossible to have an observing principal with background in every subject area and grade level, principals should work with other personnel, including subject area specialists, to better understand what effective teaching may look like in different contexts and to better understand how the work of teachers can be best supported.

There are some limitations to this study. First, as a qualitative study of 14 teachers in four schools navigating one policy, the results are not generalizable. The results may be different had other individuals or locations been chosen. The results may also be policy dependent and could look different if repeated in another state using a different evaluation system. Also, this data is derived from a larger study, and the scope of the larger project did not intend to capture perceptions of principal implementation (Frasier, 2017). Additional data including observation of the policy in action and interviews with principals would shed better light on the nuances of each school. However, given the political context of the evaluation policy (other states have adopted similar schemes) and the themes which emerged across all schools despite their differences, it appears that the role of principals in the implementation of a centralized evaluation policy should be further investigated.

Overall, the standardization of evaluation at the state level is not inherently a bad thing as previous research has demonstrated the limitations of locally developed systems (e.g., Weisberg et al., 2009). However, this study raises some limitations of the evaluation system studied here. If evaluation is meant to move beyond serving as a tool for ranking teachers and is meant to also provide useful feedback for improving teacher performance, then more research should be conducted to determine the best ways to ensure that principals are able to do so and move beyond simple compliance with policy mandates. Additionally, the trade-offs of moving from a local system to a state-centered one should be examined and considered by policymakers. At this time, it is unclear whether moving from a local to state system provides more accurate rankings of teacher performance and/or more useful feedback for improving teacher performance. The costs and benefits of moving from local systems to state centralized needs further investigation to determine the utility of both systems as levers to improve teacher performance.

Further research that extends beyond the perceptions of teachers is needed to determine the extent to which local level actors, such as principals, impact the success of evaluation policy. Examples of future research in that area may include the following research questions: (1) How do principals perceive the effectiveness of teacher evaluation policy? and (2) How do principals perceive their role in influencing teacher evaluation policy implementation? Additionally, further studies across different school contexts, including different grade levels, and across other evaluation policies with similarly stated goals would also further elucidate the role principals play in the successful implementation of evaluation policy.

References

Bryk, A., Sebring, P. B., Allensworth, E., Luppescu, S., & Easton, J. (2010). *Organizing schools for improvement: Lessons from Chicago.* University of Chicago Press.

Callahan, K., & Sadeghi, L. (2015). Teacher perceptions of the value of teacher evaluations: New Jersey's ACHIEVE NJ. *International Journal of Educational Leadership Preparation*, *10*(1), 46-59.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2011). *The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood.* National Bureau of Economic Research. http://obs.rc.fas.harvard.edu/chetty/value_added.pdf

Coburn, C. E. (2005). Shaping teacher sensemaking: School leaders and the enactment of reading policy. *Educational Policy*, *19*(3), 476-509.

Deyhle, D. L., Hess, G. A., & LeCompte, M. D. (1992). Approaching ethical issues for qualitative researchers in education. In M. LeCompte, W. Millroy, & J. Preissle (Eds.), *Handbook of qualitative research in education* (pp. 598-641). Academic Press.

Donahue, E., & Vogel, L. R. (2018). Teacher perceptions of the impact of an evaluation system on classroom instructional practices. *Journal of School Leadership*, *28*(1), 31-55.

Donaldson, M., & Mavrogordato, M. (2018). Principals and teacher evaluation: The cognitive, relational, and organizational dimensions of working with low-performing teachers. *Journal of Educational Administration*, *56*(6), 586-601.

Finster, M., & Milanowski, A. (2018). Teacher perceptions of a new performance evaluation system and their influence on practice: A within-and between-school level analysis. *Education Policy Analysis Archives*, *26*(41), 1-48.

Firestone, W. A., Nordin, T. L., Shcherbakov, A., Kirova, D., & Blitz, C. L. (2014). *New Jersey's pilot teacher evaluation program: Year 2 final report*. Rutgers Graduate School of Education.

Frasier, A. M. S. (2017). *Examining Teacher Perceptions of the Relationship Between Evaluation Policy and Teacher Practice in a North Carolina School System*. Michigan State University.

Grissom, J. A., & Loeb, S. (2017). Assessing principals' assessments: Subjective evaluations of teacher effectiveness in low- and high-stakes environments. *Education Finance and Policy, 12*(3), 369-395.

Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review, 100*(2), 267-271.

Henry, G. T., & Guthrie, J. E. (2016). Using multiple measures for developmental teacher evaluation. In J. A. Grissom & P. Youngs (Eds.), *Improving teacher evaluation systems: Making the most of multiple measures* (pp. 143-155). Teachers College Press.

Jiang, J. Y., Sporte, S. E., & Luppescu, S. (2015). Teacher perspectives on evaluation reform: Chicago's REACH students. *Educational Researcher*, *44*(2), 105-116.

Kraft, M. A., & Gilmour, A. (2016). Can principals promote teacher development as evaluators? A case study of principals' views and experiences. *Educational Administration Quarterly, 52*(5), 711-753.

Lipsky, M. (2010). *Street-level bureaucracy: Dilemmas of the individual in public service.* Russell Sage Foundation.

Louis, K., Dretzke, B., & Wahlstrom, K. (2010). How does leadership affect student achievement? Results from a national US survey. *School Effectiveness and School Improvement, 21*(3),

315-336.

Mintrop, H., & Sunderman, G. L. (2013). The paradoxes of data-driven school reform: Learning from two generations of centralized accountability systems in the United States. In D. Anagnostopoulos, S. R. Rutledge, & R. Jacobsen (Eds.), *The infrastructure of accountability* (pp. 23-40). Harvard University Press.

Nelson, B. S., & Sassi, A. (2000). Shifting approaches to supervision: The case of mathematics supervision. *Educational Administration Quarterly*, *36*(4), 553-584.

North Carolina Department of Public Instruction. (2015). *NC Educator Evaluation System.* http://www.dpi.state.nc.us/effectiveness-model/ncees/

Reinhorn, S. K., Johnson, S. M., & Simon, N. S. (2017). Investing in development: Six high-performing, high-poverty schools implement the Massachusetts Teacher Evaluation Policy. *Educational Evaluation and Policy Analysis, 39*(3), 383-406.

Spillane, J. P., & Kenney, A. W. (2012). School administration in a changing education sector: The US experience. *Journal of Educational Administration, 50*(5), 541-561.

Sporte, S. E., & Jiang, J. Y. (2016). *Teacher evaluation in practice: Year 3 teacher and administrator perceptions of REACH* [Research brief]. University of Chicago Consortium on Chicago School Research.

Stake, R. (2004). Qualitative case study. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed., pp. 443–466). Sage.

Tyack, D. B., & Cuban, L. (1995). *Tinkering toward utopia.* Harvard University Press.

U.S. Department of Education. (2009). *Race to the top program executive summary.* www2.ed.gov/programs/racetothetop/executive-summary.pdf

Vinovskis, M. (2009). *From a nation at risk to no child left behind: National education goals and the creation of federal education policy.* Teachers College Press, Columbia University.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness.* The New Teacher Project. http://widgeteffect.org/downloads/TheWidgetEffect.pdf 2009

Yin, R. K. (2009). *Doing case study research* (4th ed.). Sage.

Appendix A

Interview 1 Protocol

Teacher ID #   _____

Location _____

Date of Interview _____

Length of Interview _____

Start time _____ Stop time _____

Interviewer _____

Before recording:

Thank you very much for allowing me to interview you. Your responses will help me better understand the relationship between teacher practice and teacher evaluation. As with any part of this study, you can withdraw your consent to participate at any time and you do not have to answer any questions that you do not want to answer. Please be aware that I want you to be honest in these interviews, even if it means saying things you think I might not want to hear. With your permission I would like to record the interview so that I can concentrate on what you are saying rather than note-taking. Is that okay?

Structured questions for all interviewees:

- Would you please explain your experiences with evaluation both in the past as well as in the current school year? (Begin)
- Are evaluations necessary because teachers need to be rated, sanctioned, or rewarded in order to be motivated to do a better job? (End)
- Do evaluations yield information that is useful for teachers to improve practice? (End)

Potential general prompts for all interviewees:

- Can you describe for me how evaluation occurs in your school?
- Do you use the scores from your evaluation to change your instruction? Explain.
- How do you prepare for an evaluation?
- Do you think evaluations impact the way you teach? Explain.
- Do you teach/plan to teach differently when you know you are going to be observed?
- Do you teach/plan to teach differently when your students are being tested?
- Do you feel like the evaluation system accurately captures your ability as a teacher? Explain.
- Are you responding to evaluation differently this year as opposed to last year?
- Is there anything else you think I should know about teacher evaluation?

Potential prompts based on survey responses:

- Respondents in your school seem to feel _____. Do you agree/disagree? OR can you tell me more about that?
- I noticed that you put that you _____. Can you explain what you mean by that?
- In the survey you say that you _____, can you give me an example of that?
- Some teachers in your school put _____, do you feel that is an accurate view? Why or why not?

Appendix B

Interview 2 Protocol

Teacher ID # _____

Location _____

Date of Interview _____

Length of Interview _____

Start time _____ Stop time _____

Interviewer _____

Before recording:

Thank you very much for allowing me to interview you again. Your responses will help me better understand the relationship between teacher practice and teacher evaluation. As with any part of this study, you can withdraw your consent to participate at any time and you do not have to answer any questions that you do not want to answer. Please be aware that I want you to be honest in these interviews, even if it means saying things you think I might not want to hear. With your permission I would like to record the interview so that I can concentrate on what you are saying rather than note-taking. Is that okay?

Potential Questions:

- How have your observations been this year?
- How did testing go last semester?
- What courses are you teaching this semester?
- Do you feel any difference in pressure on you this semester? Why?

Asking for contextual opinion on their school based on quantitative scores.

- I would like to share some scores with you some data on your school. I have two numbers I would be interested in hearing your opinion about. The first is an Evaluation Condition Score. This number was calculated using evaluation-based questions from the Teacher Working Condition Survey administered biannually. The number is a measure of your school's distance from the district average. A positive number means the condition is higher than the district average and a negative means it is lower. Your school's number is ___. Do you feel like that is an accurate representation of how teachers at your school feel about evaluation?
- The second score is an effectiveness score. This number was calculated using data on the average scores of teachers on their evaluation. The number is a measure of your school's distance from the district average. A positive number means that the overall average score of teachers at your school is higher than the district average and a negative means it is lower. Your school's number is ___. Do you feel like that is an accurate representation of how teachers at your school score on evaluation?

Appendix C

Transcript Coding Scheme: Interview Level Coding Descriptors: Teacher, School Site, Subject Era, Experience Level

Code	Description	Example
Timeliness*	The teacher makes a reference about the availability or timing of the evaluation or evaluation results.	Yes, it's (not) very helpful, and it's not even broken down. I can look at the overall goals, but if I were given a classroom assessment, I could look at oh, these are the questions they missed and these questions are all aligned and I could go oh, I could actually look at the questions and go this is what I think they were missing, or I could talk about it with the kids. It's too little, too late. So, I'm just like "Yeah, I met my standard. I'm still green. I'm not gonna get fired today."
Consistency*	The teacher makes a reference about the consistent application of the evaluation across or between individuals, schools, counties, etc.	It's all so subjectional, yeah. And, man, it's even based on how [the principal] felt that day. It's either overly rosy, or it's just stupid, and so yeah, it's pretty useless to me. And then I've told you about the administrator feedback, I think that the most positive feedback I've gained has been from my peers, I think, my teaching peers that know the content better than anyone else, know what I'm going for more than anyone else, already know me too. There's also this level of intimacy and friendship that we have with one another, where I am able to listen to them... I've listened to are better than anyone else, I mean they know what they're talking about. So, I care about that.
Breadth*	The teacher makes a reference to the breadth or coverage of the evaluation.	And I say that because now when I sit in, and I get evaluated, it's more than that. And do I help to hire new teachers? No. Am I part of a professional organization? No, because I have to pay my mortgage and it's incredibly expensive to belong to a professional organization. And also, do I think my job is so big at this point that if an administrator wanted to fire me, they could follow me around and nickel and dime me for things about lunch duty and bathroom duty, and did I leave my class alone when I went to the bathroom, and did I have the number of copies I needed, did I record my grades in a timely manner? Do I think I could be fired based on that? Yes. So, in that way, evaluations are exhausting.

Expertise of Observer*	The teacher makes a reference to expertise of the evaluators. Could be on subject area or otherwise.	The person that does our evaluations knows little to nothing about what we do and isn't particularly interested. We literally write our own evaluations. He can't spell, for one thing, and he will make you look so, like such a... He looks like such an idiot. So, we finally... It takes forever but we finally just say, "Give me the computer," and we start typing, because if we don't do that, we'll be misrepresented, simply because he's incompetent when it comes to that. He's really good at some aspects of the job, don't get me wrong. This one, he's incompetent. And so, I write my own evaluations. And I think I'm fair. What I see as another compliment to us as a department is that the principal puts this guy in charge of us because he knows that we don't really need a lot. He puts the people who are better at this sort of stuff in charge of departments that really need a lot of assistance.
Administrator (Parent Code)	The teacher references their administrator enacting the evaluation policy.	Last year, I didn't get observed until after exams in the spring, and it was an administrator that was not the one that was assigned me, and then very quickly, they were in again after another snapshot… the day after. And then this year, I've been observed three times already, which is fine, it's great to get feedback.
Administrative Capacity** (Child Code of Administrator)	The teacher references their administrator's ability to enact and support the policy and the theme of administrative capacity surfaced in all four schools.	I always wondered too, I guess I'm empathetic toward the administrators, they, like the teachers, are given so many tasks to complete daily, too many possibly to keep up with that type of paperwork. So, I feel empathy for them too.
Policy Values** (Child Code of Administrator)	The teacher references what they perceive their administrator values in evaluation.	Yet they want us to be reflective, and as someone who reflects you have to identify your weaknesses. Right now, they're pushing so much on data, and everything we're doing has to be backed by data, and again I just feel like this is taking teaching and trying to make us more robotic and trying to make us not value relationships and building relationships with students. That isn't an area that's ever evaluated… teacher-student relationship rapport

Observer Effectiveness** (Child Code of Administrator)	The teacher references their perceptions of an administrator's effectiveness to complete the evaluation. This could refer to their effectiveness as an evaluator in general or be related to their content or pedagogical knowledge.	So, I do remember in the beginning, and this was a long time ago, but the big push was on using technology in the classrooms. So, we had graphing calculators. And the evaluators, who were not math people, just assumed that that was an excellent use of technology, which maybe at the time it was, but it was actually just another tool that we were already using. So, the math teachers kinda rode along for a while.
*Codes from dissertation where I noticed a trend where teachers spoke about administrators influencing the policy. ** These codes are the themes addressed in this article as discussed in my findings and supported by my literature review.		

Author Note

Amanda Slaten Frasier is a National Board Certified Teacher and holds a Ph.D. in Educational Policy from Michigan State University. She currently holds a faculty position at University School at East Tennessee State University. Please direct correspondence to amanda.slaten.frasier@gmail.com.

Article Citation

Frasier, A. S. (2021). Principal-led state evaluation policy: Teacher perceptions at four high schools. *The Qualitative Report, 26*(9), 2769-2796. https://doi.org/10.46743/2160-3715/2021.4696