UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

AN EXAMINATION OF FORMAL SCHOOL STRUCTURE AND TEACHER EVALUATION

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF EDUCATION

By

ELIZABETH ROSE BRANDS VEREECKE
Norman, Oklahoma
2016

AN EXAMINATION OF FORMAL SCHOOL STRUCTURE AND TEACHER EVALUATION

A DISSERTATION APPROVED FOR THE DEPARTMENT OF EDUCATIONAL LEADERSHIP AND POLICY STUDIES

BY

Dr. Curt Adams, Chair

Dr. Chan M. Hellman

Dr. Patrick B. Forsyth

Dr. William C. Frick

Dr. Beverly J. Edwards

Dedication

I dedicate this dissertation to the first class that I ever taught: the 5th graders at Saint Catherine School in Tulsa Oklahoma. Each of you – unknowing to you – set me on this path by teaching me a great deal about what mattered, inspired, and motivated me. You taught me that no matter how hard the day I had the fortitude and determination inside of me to persevere. You taught me that heroes do not wear capes or their underwear on the outside of their pants. Heroes are the people who do the simple things over and over and over again; they just keep showing up. If I taught any of you even the minutest piece of what I learned the year you came into my life, then I am at a loss for words of gratitude. You changed the trajectory of my life and inspired in me a future career of championing the right that all children deserve – the right to the highest quality education our community can give you. I hope that I gave you this high-quality education. Know that remembering our 5th grade class at Saint Catherine motivates me everyday to be the champion that all children need. In this way, your legacy is greater than you know.

Acknowledgements

I am indebted to the faculty of the Jeannine Rainbolt College of Education at the University of Oklahoma. Of particular note is my dissertation chair, Dr. Curt Adams. I am appreciative of you taking time to encourage in me the academic process of taking responsibility for creating new ideas and understanding from the theory and data. To Drs. Patrick Forsyth, Beverly Edwards, Chan Hellman, and William Frick, this dissertation is better because of your critiques and support.

To my parents, who were my first teachers and my sisters who were my first students in the makeshift classroom in my childhood home. You modeled for me and I practiced on you what would become educational pedagogy that I would put into practice. It is from you that I learned that real learning comes only when you model a joy for learning and when real relationships are formed.

For my husband, who knows that my determination is a close cousin to being stubborn. He serves as my favorite teammate and live-in professional development. Together we embarked on this academic journey motivated by our experiences as classroom teachers, inspired by future career opportunities to serve more students, and in the competitive spirit that keeps us moving forward in life. The heart of this dissertation is due to your patience and encouragement.

Table of Contents

Abstract

Research showing that teachers are the most important school-related factor for student improvement has increased attention on what and how to improve teaching quality in our classrooms. Compounding this is the heightened focus on standards-based outcomes mandated on schools and districts to quantify student growth as the key driver of measuring school effectiveness. The result of this has been an increase in financial and human capital investment to create robust teacher evaluation tools that evaluate teacher performance for the purpose of making personnel decisions and to spur improved teaching quality and effectiveness. The question remains: How can the tools meant to spur increased teacher quality – of primary interest is the teacher evaluation – be used to achieve this intended outcome?

Considering this question through the lens of bureaucracy theory and classical management theory, this research aims to explore how the structural contexts of the school influence perceived usefulness of the teacher evaluation. Narrowing the research focus to organizational structure and structure-related consequences for employee evaluation was chosen because structure is mutable. That is, leaders and other organizational actors have a degree of control over the use of formal structures to guide individual behavior and collective action. Therefore, this research has application for how scholars and practitioners can adjust school structure to enhance the effectiveness of teacher evaluation.

Chapter 1: Introduction

One of the most controversial tools to stimulate teacher and leadership effectiveness is formal performance evaluations (Halliner, Heck, & Murphy, 2014). Performance evaluations are meant to monitor teacher performance and provide feedback so that educators can adapt their instructional methods and pedagogy to the learning needs of students (Sanders & Horn, 1998; Sanders, Wright, & Horn, 1997). Daley and Kim support this intended application of evaluation systems: "Teacher evaluations can fulfill two related purposes of personal growth and accountability" (2010, p. 5). In practice, formal evaluations are more often wrought with negative connotations of bureaucratic oversight, top-down management, the fear of punitive consequences, and decades of failed attempts to stimulate teaching growth and accountability (Daley & Kim, 2010). The advent of higher expectations on schools, tightening budgets, educational reform, and a more complex set of contextual factors facing students of the United States, heightens the stakes for effective teaching (Daley & Kim, 2010). With a considerable amount of resources devoted to implementing new evaluation frameworks, it is imperative to understand how evaluation tools and processes can be used to raise teaching quality.

The increased importance of having a useful teacher evaluation framework has largely emerged from teacher effects studies identifying teachers as the most important school-related factor for student achievement (Daley & Kim, 2010; Darling-Hammond, 1997; Darling-Hammond, 2000; Wright, Horn, & Sanders, 1997). The issue of effective teachers is so important that the National Commission on Teaching and America's Future declared that teaching quality must be at the center of the education reform agenda

(Darling-Hammond, 1997). President Obama's administration, in *A Blueprint for Reform,* places teacher quality at the center of its reform agenda. As President Obama argued, "Our goal must be to have a great teacher in every classroom and a great principal in every school" (U.S. Department of Education, 2010, p. 1). Darling-Hammond echoes the significance of quality teaching in stating, "without a sustained commitment to teachers' learning and school redesign, the goal of dramatically enhancing school performance for all of America's children will remain unfulfilled" (1997, p. 1).

In theory, teacher evaluation is an essential formal structure for improved teaching quality in schools (Sanders, Wright, & Horn, 1997). Teacher evaluations are meant to ensure an effective teacher in every classroom by providing targeted feedback, coaching, professional development, and when necessary exiting ineffective teachers from schools (Wise, Darling-Hammond, McLaughlin, & Bernstein, 1984). However, teacher evaluations do not always work in practice. Research on performance-based evaluations paints a mixed picture of effectiveness. Hallinger, Heck, and Murphy (2014) argue that the policy logic supporting performance-based teacher evaluations remains considerably stronger than the empirical evidence on its implementation and effectiveness.

Although there are some positive findings on the effectiveness of new performance-based teacher evaluation systems (Coggshall, Max, & Bassett, 2008; Daley & Kim, 2010; Hallinger, Heck, & Murphy, 2014), it is the case that the policy objectives of improved achievement and closed achievement gaps set forth by No Child Left Behind and Race to the Top have not been realized (Hallinger, Heck, & Murphy, 2014). In attempting to understand the lack of progress toward better student outcomes, some have

argued that good teaching is more complex than the frameworks acknowledge, that value added estimates associated with many models are unreliable, and that evaluation is the wrong driver for improvement (Fullan, 2011; Hallinger, Heck, & Murphy, 2014). A factor that has not received much attention has to do with the effects of organizational structure on teachers' experience in the evaluation process.

This study draws on bureaucracy theory and classical management theory to explain why in some schools teacher evaluation processes may be experienced favorably whereas in other schools the same framework and processes may be perceived as ineffective. Narrowing the research focus to organizational structure and structure-related consequences for employees was chosen because structure is mutable. That is, leaders and other organizational actors have a degree of control over the use of formal structures to guide individual behavior and collective action (Hoy & Sweetland, 2001).

Formal structures become increasingly important in larger, more complex organizations where low-productivity, high-employee turnover, and low-employee satisfaction leave organizations vulnerable to failure (Dalton et al., 1980; Ford, 1973; Pierce & Dunham, 1976; Weed, 1971). Organizations need structure; however, formal structures can often be applied in ways that hinder processes and outcomes (Hoy & Sweetland, 2001). The degree to which structural conditions enable or constrain how teachers experience formal evaluation has implications for how school leaders organize and coordinate evaluation processes.

With data from schools in a mid-sized urban school district, this study used a non-experimental correlational design to measure the relationship between school structure and teacher perceived usefulness of the evaluation framework. The study was based on

the assumption that teacher perceptions of the effectiveness of evaluation stems from the formal structure within which the evaluation tool is used. Such a belief was derived from two sources of evidence. First, decades of research on teacher evaluation show inconsistent findings regarding the usefulness of evaluation in measuring teacher effectiveness and promoting professional growth (Borman & Kimball, 2005; Kimball et al., 2004; Milanowski, 2004; Webster & Mendro, 1997; White, 2004). Second, evidence within bureaucracy theory and classical management theory show that the same action within different structures can have dissimilar meanings based on how employees perceive their environment (Goe & Croft, 2009; Hackman, Oldhman, Janson, & Purdy, 1975; Hallinger, Heck, & Murphy, 2014; Organ & Greene, 1981; Shepard, 1970; Stone & Porter, 1975; Wolf et al. 1997).

Extending research using bureaucracy theory and classical management theory to teacher evaluation suggests that an enabling structure, where teachers perceive that the rules and procedures are adaptable to their unique needs, may provide the context for evaluation to be experienced as supporting teacher growth. In a highly bureaucratic structure, where those in power determine procedures unilaterally, teachers may perceive the same evaluation tool as punitive. The degree to which teachers perceive the school structure may alter their perceptions of performance evaluation. Thus, it is important to determine what organizational effects, if any, are related to teacher perceptions of the evaluation process.

This study attempts to shed light on the plausible reason for inconsistent findings in the evaluation literature by examining how teacher perceptions of formal school structure may contribute to a useful evaluation experience. The study does not argue for

or against a particular evaluation design; rather, it examines whether or not formal school structures can explain differential teacher experiences with performance evaluation. The evidence has application for how scholars and practitioners can adjust school structure to enhance the effectiveness of teacher evaluation.

Statement of the Problem

Teaching conditions have rarely been more challenging then they are today (National Center for Education Statistics, 2015). Federal and state dollars are more limited than ever and contextual factors challenging communities affect student learning and learning opportunities (Hal & Briar-Lawson, 1997; National Center for Education Statistics, 2015). At the same time, the stakes for higher and equitable learning outcomes have never been greater (Bush, 2001). Teacher evaluation systems are intended to be a strategy for improving teacher quality and in turn raising student outcomes (Gates Foundation, 2013; Hallinger, Heck, & Murphy, 2014; Milanowski et al., 2005; Odden & Wallace, 2008). Research asserts that improving teacher performance can improve student achievement (Danielson, 2007; Milanowski, 2004; Sanders & Horn, 1994).

Milanowski, Kimball, and White (2004) studied a large urban school district with 70 schools, about 48,000 students, and 3,000 teachers to explore the relationship between teacher evaluation scores and student achievement. They found that standard-based teacher evaluation scores had a positive relationship with student achievement (Milanowski, Kimball, & White, 2004). Higher correlations were found between teacher evaluation scores and student performance in language arts and math (Milanowski, Kimball, & White, 2004). The study then controlled English proficiency, special education status, and ethnicity and found that the relationship between teacher evaluation

scores and student performance persisted after accounting for student characteristics commonly associated with student achievement (Milanowski, Kimball, & White, 2004). These findings, largely similar to those found in the researchers' previous study (Kimball, White, Milanowski, & Borman, 2014; Milanowski, 2004), suggest that evaluation scores from well-designed teacher evaluations tools have the potential to identify teachers whose students are achieving higher test scores.

Milanowski and colleagues (Milanowski, 2004; Milanowski, Kimball, & White, 2004) findings support the predictive validity of teacher evaluation ratings from observational protocols, but their research does not address conditions and characteristics of evaluation use that explain differences in how teachers perceive the usefulness of the evaluation system. It is difficult to imagine that teacher evaluation can achieve its promise without teachers buying into its usefulness. This is an important point because it is ultimately up to teachers themselves, not the evaluation tool, if they are to improve.

Improved teaching benefits from instructional frameworks that are used to facilitate regular interactions around teaching and learning (Wise, Darling-Hammond, McLaughlin, & Bernstein, 1984). This is to claim that the context matters for professional growth. To illustrate, there are different affective consequences between receiving advice on how to improve in an environment where the teacher feels supported and trusts his or her leader to provide authentic tools for improvement versus being given the same advice in an environment where the teacher fears punitive consequences (Fullan, 2011; Hoy & Sweetland, 2001; Wise, Darling-Hammond, McLaughlin, & Bernstein, 1984). As Fullan argues, tools used to punish teachers and schools are the wrong drivers of improvement. Punitive use of structure seems to have adverse effects. Fullan (2011) discusses what he

refers to as, " …'wrong drivers' … a deliberate policy force that has little chance of achieving the desired result" (p. 3). Among these drivers is using measures that punish teachers and schools (Fullan, 2011).

Current research on performance-based evaluation has not examined the influence of school context on the use and perceived utility of new evaluation processes. Simply focusing on the predictive validity of teacher ratings, or on the fidelity of implementation, ignores the socio-cultural context affecting the use of formal teacher evaluation (Goe & Croft, 2009; Hallinger, Heck, & Murphy, 2014). Decades of research on performance-based evaluation tools lack research on teacher perception of the usefulness of the evaluation system (Goe & Croft, 2009; Hallinger, Heck, & Murphy; 2014; Sweetland & Hoy, 2000; Wise, Darling-Hammond, McLaughlin, & Bernstein, 1985; Wolf et al., 1997). More research is needed in this regard.

General organizational literature suggests that enacted structure has consequences for employee attitudes and behavior (Alder & Borys, 1996; Dalton et al., 1980; Hoy & Sweetland, 2001). Evidence is clear that employees benefit from structure, rules, leadership, and procedures. The nuance lies in how employees perceive the structure, rules, and procedures that affects their satisfaction and performance (Finlay et al., 1995; Hackman, Oldhma, Janson, & Purdy, 1975; Organ & Greene, 1981; Shepard, 1970; Sims, Keller, & Szilagyi, 1976; Stone & Porter, 1975). The general organizational research has been more recently translated into schools to explain the conditions necessary for teachers and students to flourish (Alder & Borys, 1996; Hoy & Sweetland, 2001). Evidence from school studies corroborates general organizational research. Teachers are generally happier, more engaged, and perform better when formal structures are seen as

helpful and supportive (Forsyth, Adams, & Hoy, 2011). A lingering question that this study addresses is whether or not formal structures can influence how teachers experience performance evaluation. Does the application of formal structures influence the usefulness of teacher evaluation?

Statement of Purpose

The purpose of this study was to see what effects, if any, school structure has on the perceived usefulness of teacher evaluations. The lenses of bureaucracy theory and classical management theory explain conditions in which structural features of schools can enhance the usefulness of teacher evaluation. There are strong bodies of research on both the consequences of structure on employee satisfaction and productivity as well as the fidelity of performance-based evaluation systems (Goe & Croft, 2009; Hackman, Oldhma, Janson, & Purdy, 1975; Hallinger, Heck, & Murphy, 2014; Organ & Greene, 1981; Shepard, 1970; Stone & Porter, 1975; Wolf et al. 1997). Structure contributes to employee affective behavior. When employees feel increased job satisfaction, this has positive effects on organization productivity and profits (Hackman, Oldhma, Janson, & Purdy, 1975). In contrast, employees who feel limited job satisfaction are less productive and profits suffer (Hackman, 1980).

General organizational literature explains that context matters for worker performance and productivity (Finlay et al., 1995; Hackman, Oldhma, Janson, & Purdy, 1975; Organ & Greene, 1981; Shepard, 1970; Sims, Keller, & Szilagyi, 1976; Stone & Porter, 1975). While the perception of tools in these various environments is not explicit in the research, this study will use bureaucracy theory and classical management theory to explain that it is logical to posit that the same tool or action applied in various

environments has different consequences based on the structure. In connecting back to evaluation systems, the literature supports the argument that the same evaluation tool applied in different environments will produce inconsistent results because context matters in regards to employee experiences with formal structures like teacher evaluations.

Even with the proliferation of research surrounding organizational structure, no research could be found that explores the structure-related conditions and characteristics that explain teacher perceived usefulness of evaluation systems. There is a strong body of research on the fidelity of performance-based evaluation tools (Danielson, 2007; Koretz, 2008; Sanders & Horn, 1998; Webster & Mendro, 1997). The literature explores the algorithms, implementation technique, and content of the evaluation tool (Goe & Croft, 2009; Hallinger, Heck, & Murphy, 2014) without consideration for the consequences of the context in which the tool is implemented.

This study attempts to bridge the gap in the literature by examining the effect of formal school structure on teacher perceived usefulness of the performance evaluation. Specifically, the study examines the extent to which enabling formalization, enabling centralization, and trust in the principal explain differences in teacher perceptions of evaluation. The importance of this study can be seen within three current realities. First, states and districts have spent much time and resources on new performance evaluation frameworks without strong and conclusive evidence on their effectiveness in improving teaching and learning (Hallinger, Heck, & Murphy, 2014; Wise, Darling-Hammond, McLaughlin, & Bernstein, 1985). Second, research shows that the most important school-related factor to student achievement is teacher quality (Daley & Kim, 2010; Darling-

Hammond, 1997; Darling-Hammond, 2000; Wright, Horn, & Sanders, 1997), and understanding how to provide quality through evaluation can make the process more meaningful (Kimball, White, Milanowski, & Borman, 2014; Milanowski, 2004). If teachers are dismissing a tool that can improve student achievement, then the monetary and human capital investments in the tool meant for self-improvement are mitigated. Third, school structures are malleable. Educators control the degree to which structures are used to support useful teacher evaluation. If it is determined that formal structures matter, educators and policy makers can shift investments to fostering environments that create positive conditions for implementation of teacher evaluation.

In short, this research adds to the literature by examining the effects of formal structure on teacher experiences with the evaluation process. This has especially meaningful consequences for schools and policy makers as they continue to invest in creating systems that improve teaching quality and student achievement. The application of formal school structure may just be a mechanism by which to improve the utility of teacher evaluation.

Definition of Terms

Enabling School Structure

Enabling school structure looks at the relationship between structure and teacher's perceptions of structure in a school context (Hoy & Sweetland, 2001). Enabling structure reflects teacher perceptions of how bureaucracy is used to regulate teaching and learning. Its hierarchy helps rather than hinders and provides a decision-making framework where principals and faculty work collectively across recognized boundaries. Rules and regulations provide flexible guides for problem-solving that support employees rather

than using hierarchy as a vehicle to enhance power, punishment, or constraints (Hoy & Sweetland, 2001). Enabling school structure is operationalized as its two conceptually distinct components: formalization and centralization (Adler & Borys, 1996; Hoy & Sweetland, 2001). Formalization looks at the rules and processes within an organization. Centralization looks at the hierarchy of leadership. Conceptually, these are two distinct factors that derive from bureaucracy theory, thus they were measured as distinct constructs (Forsyth & Adams, 2014).

Faculty Trust in Principal

Faculty trust in principal measures the quality of the relationships between faculty and the principal. More specifically, it is measured based on the concept of trust developed by Hoy and Tschannen-Moran (1999) that defines trust as one party's willingness to be vulnerable to another party based on the confidence that the latter party is benevolent, reliable, competent, honest, and open (Hoy & Tschannen-Moran, 1999). Thereby, faculty trust in principal is operationalized as faculty perceiving the principal as benevolent, reliable, competent, honest, and open. Higher principal trust indicates that faculty respect and trust the principal's leadership.

Performance-Based Teacher Evaluation

Performance-based teacher evaluations are designed to assess the quality of teacher performance on one or multiple important aspects of teaching to determine teachers' knowledge and skills as they are used in practice (Coggshall, Max, & Bassett, 2008). The district in which data were derived used a performance-based evaluation for teachers. Fifty percent of the evaluation rating of all classroom teachers is based on a qualitative observation process, and fifty percent is based on a quantitative component

(Tulsa Public Schools, 2012). To meet the qualitative evaluation component, teachers are observed at a minimum of two times and in some cases three times by a qualified and certified administrator (Tulsa Public Schools, 2012). Teachers must receive notification of the observation. The evaluation tool contains twenty domains and each domain includes a definition of "effective" teaching which is a score of three on a five-point scale (Tulsa Public Schools, 2012). Teachers receive comments for scores falling at the low or high range and can submit a written rebuttal to any observation comments and request an additional observation (Tulsa Public Schools, 2012). The quantitative component is made up of two parts: current year student survey data and value added estimates (Tulsa Public Schools, 2012). Student survey data is adjusted for the grade level taught by the teacher to determine whether a teacher's performance is statistically average or significantly above or below average. The value added estimates are up to three-year averages using the teacher's overall weighted average (Tulsa Public Schools, 2012).

Teacher Perceived Usefulness of Evaluation Tool

This concept is operationalized as teacher perception of their understanding of the implementation and value of the evaluation rubric and process (Tulsa Public Schools, 2012). Teacher perceived usefulness of the evaluation tool is an aggregate of teachers across multiple schools in the district in which data were derived (Tulsa Public Schools, 2012). In this study, the terms teacher perceived usefulness of the evaluation system and teacher perceived usefulness of the evaluation tool are used interchangeably.

Teacher Trust in District Administration

This concept measures teacher perception of the district's willingness to be vulnerable to another party based on the confidence that the district is benevolent,

competent, reliable, open, and honest (Hoy, 2002; Hoy & Tschannen-Moran, 1999; Tschannen-Moran & Hoy, 1998). More specifically, it assesses faculty perceptions of the degree to which the district administration is aware of relevant issues, organized, committed, and supportive of teachers' autonomy and professional growth.

Analytic Technique

The data for the empirical investigation were hierarchically structured with teachers nested in schools. The primary interest was to determine the degree to which school formal structures shaped teacher perceived utility of performance evaluation. Thus, HLM was used to analyze variance around teacher perceived utility of evaluation. A conventional model building process in HLM 7.0 was used to test the three hypotheses. First, an unconditional null model was run to decompose variance in each dependent variable to within- and between-school factors. Second, a random coefficient regression was modeled to test the effects of teacher characteristics on their perceptions of the performance. Finally, a random effects ANOVA model was used to test the hypotheses while controlling for school conditions and teacher characteristics.

Hierarchical Linear Modeling (HLM) was designed to correct for the ecological fallacy of drawing inferences at the individual level from group level data or the atomistic fallacy of making group level generalizations from individual level data (Diez Roux, 2002; Raudenbush & Bryk, 2002). This occurs by decomposing variance in a dependent variable to individual and group factors then proceeding to account for this variance at the respective levels (Vogt, 2007). Like all parametric statistics, HLM analysis in this study was based on three primary assumptions. First, residuals are uncorrelated and have

constant variance. Second, student-level data were collected and measured without error. Third, level-1 errors are independently and normally distributed with a common variance.

Limitations

Limitations exist in all research, and this study was no exception. Three limitations should be considered when thinking about the evidence and conclusions presented in this research.

First, the research was non-experimental, meaning that some variables that could confound results were not controlled for in the study. Steps were taken to control for alternative explanations for teacher differences in perceived usefulness of performance evaluation in the statistical models, but it remains that factors other than the variables of interest could be contributing to differences in teacher perceptions.

Second, only about 7 percent of the variance in perceived usefulness of evaluation existed at the school level. It is important to account for the school-level variance as this study did, but with nearly 93 percent of the variance determined to exist across individual teachers there were many factors unique to individual teachers that this study did not account for.

Third, data come from one school district with one common evaluation framework and process. The homogenous sample could limit variability in how principals conduct evaluations, and in turn, how teachers perceive the usefulness. Limited variability would affect the estimated relationships in the analytical models.

Organization of the Dissertation

Chapter 1 explains the significance of this work in relation to the current public school reform landscape as related to the research on performance-based evaluation in

creating meaningful tools for teacher professional development (Goe & Croft, 2009; Hallinger, Heck, & Murphy, 2014; Sweetland & Hoy, 2000). A statement of problem, definition of terms, limitations of the study, and assumptions were also presented.

Chapter 2 provides a review of literature in which theories that provide the lens through which to pose and answer the hypotheses are defined and explained: bureaucracy theory and classical management theory. Key tenets and concepts are defined and described: formalization, centralization, performance-based evaluation, and enabling school structure. These theories, their tenets, and key concepts lay the framework for the hypothesis and research.

Chapter 3 presents the hypotheses and rationale. Bureaucracy theory and classical management theory are used as the lens to explain the hypothesized relationship between school structure and teacher perception of the usefulness of the evaluation system.

Chapter 4 presents the methods used to analyze the data. The research context, research design, and evaluation tool are explained. The data source and measures are described and analytical techniques are explained with justification for their use.

Chapter 5 presents the results of the study. Results include findings from descriptive data and estimates from the multi-level models are presented. Results from individual teacher and school level descriptives, correlation analysis, and Hierarchical Linear Modeling are presented.

Chapter 6 provides a discussion and summary of the findings. This section restates each hypothesis, explains data pertaining to each claim, states whether the data supports or disputes the claim, and makes an argument as to why the data supports or

disputes each claim. The chapter provides an explanation for the findings based on theoretical and speculative analysis.

Chapter 2: Review of Literature

Performance-Based Evaluations

One of the most significant findings in recent education research is that teacher effectiveness is a major determinant of student academic achievement (Darling-Hammond, 2003, 2000, 1996; Jacob & Lefgren, 2008). In response to such findings, states and school districts have made substantial monetary and capital investments to improve teacher quality through the use of teacher evaluations (Hallinger, Heck, & Murphy, 2014; Wise, Darling-Hammond, McLaughlin, & Bernstein, 1985). The assumption is that teacher evaluations weed out under-performing teachers, provide feedback to all teachers, and establish an outcome-based culture in schools; thereby, improving teacher quality and in turn student learning (Hallinger, Heck, & Murphy, 2014).

Interest in teacher evaluation as a policy solution emerged in the United States in the early 1980s when the National Commission on Excellence in Education published *A Nation at Risk: The Imperative for Educational Reform* (1983). Several of the recommendations in *A Nation at Risk* stated the need for rigorous teacher evaluations to better differentiate among levels of teacher quality and to ensure teachers were achieving high educational standards. Teacher evaluation tools developed in response to the publication were high-inference methods used to confirm that procedural conditions were met in the classroom (Hallinger, Heck, & Murphy, 2014). The evaluations were largely designed as tally forms or checklists and rarely relied on student achievement data to measure teacher effectiveness (Sweetland & Hoy, 2000).

In the past decade, increased accountability and education research showing that effective teachers are a key determinant in student achievement has led to a redesigned evaluation process that places greater emphasis on improving teacher quality as measured through student test scores (Hallinger, Heck, & Murphy, 2014). There are continuing debates about the extent to which teacher effectiveness literature has accurately identified characteristics of effective teaching and how well the teacher effectiveness literature frames subsequent development of teacher evaluation systems (Hallinger, Heck, & Murphy, 2014). Nevertheless, districts have invested heavily in experimenting with various teacher evaluation tools that measure effective teaching (Hallinger, Heck, & Murphy, 2014; Wise, Darling-Hammond, McLaughlin, & Bernstein, 1985).

Experimentation has included principal observations; analysis of classroom artifacts such as student work and lesson plans; teaching portfolios; staff interviews; and teacher, parent, and student surveys (Goe & Croft, 2009; Wolf et al., 1997). Additionally, states and districts have tied performance pay and differential compensation to teacher evaluation systems (Goe & Croft, 2009). Implementation of these features is not as widespread as the use of performance-based evaluations (Coggshall, Max, & Bassett, 2008), but they remain a feature that gets a lot of attention.

Recent teacher evaluation systems use a performance-based approach that includes multiple measures of teacher performance and a range of evidence to demonstrate teacher effectiveness (Coggshall, Max, & Bassett, 2008). Evidence on teacher effectiveness comes from classroom observations, portfolios, lesson plans, and student growth measures such as value added estimates (Coggshall, Max, & Bassett, 2008). Districts have been using performance-based teacher evaluations for the past

decade, and researchers now have the longitudinal data to evaluate the effectiveness of this school improvement strategy (Hallinger, Heck, & Murphy, 2014).

Proponents see performance-based evaluation tools as an effective way to fix problems with the previous teacher evaluation system. Such problems include narrow measures of teacher performance (Coggshall, Max, & Bassett, 2008), evaluation instruments that fail to differentiate among effective and ineffective teachers (Daley & Kim, 2010), and weak correlations between evaluation criteria and student achievement (Kane, Rockoff, & Staiger, 2008). Performance-based evaluation tools offer three primary advantages over the evaluation protocols of the past. First, performance-based evaluation tools more precisely define instructional practices and behaviors that contribute to student learning (Goe & Croft, 2009). Examples include defining how effective teachers prepare lessons, manage student behavior, assess student learning, deliver instruction, and create professional cultures (Goe & Croft, 2009). Second, evaluation tools provide clear and consistent expectations for teachers that reduce role ambiguity, establish a coherent framework for instructional practices, and encourage purposeful reflection in areas that have potential to enhance student learning (Daley & Kim, 2010; Kane, Rockoff, & Staiger, 2008). Third, performance-based evaluation tools include multiple performance categories that allow administrators to rate teachers by different levels of effectiveness for each standard (Goe & Croft, 2009).

Many performance evaluation systems also include a value added measure that compares the current test scores of a teacher's students to the test scores of a comparable group of students (Koretz, 2008; Sanders & Horn, 1998). Value added measures seek to isolate the contribution that each teacher makes to the test scores of students (Koretz,

2008). Most models are based on variance between predicted and actual student test performance after controlling for student characteristics (Koretz, 2008). Value added data are analyzed in a two- or three-level analysis where the former considers students as nested within classrooms and the latter considers teachers as nested within schools (Hallinger, Heck, & Murphy, 2014; Kane & Staiger, 2008). Value added measure were designed to inform employment decisions, professional development, merit pay, and individualized teacher improvement plans (Coggshall & Max, 2008; Toch & Rothman, 2008; Wise, Darling-Hammond, McLaughlin, & Bernstein, 1984).

Opponents see performance-based evaluation systems as wrought with inconsistencies in the design and statistical models (Goe & Croft, 2009; Hallinger, Heck, & Murphy, 2014). Four main criticisms have been leveled against value added estimates in particular. First, value added models, especially the three-level analysis, are difficult to implement due to a complex statistical model and inability to apply to students who have multiple teachers contributing to their learning (Hallinger, Heck, & Murphy, 2014; Webster & Mendro, 1997). Second, value added models yield inconsistent patterns of results for individual teachers, thereby calling into question their validity for the purpose of teacher performance evaluation and employment decisions (Goe & Croft, 2009; Hallinger, Heck, & Murphy, 2014). Third, value added performance is affected by the nonrandom assignment of students to teachers (Goe & Croft, 2009; Jacob & Lefgren, 2008). According to Goe and Croft, "Because teachers are not randomly assigned to schools, and students are not randomly assigned to teachers, it is difficult to sort out how much student achievement growth is attributable solely to teachers' efforts and how much is attributable to other factors not included in the statistical model" (2009, p. 4). This calls

into question the fairness of using value added measures of student learning in teacher evaluations. Fourth, value added ratings are unable to separate out the many alternative explanations of student achievement, thereby providing a distorted measure of an individual teacher's effectiveness (Haertel, 2013; Hallinger, Heck, & Murphy, 2014; Sanders, Wright, & Horn, 1997).

In conflict with the critical claims, evidence also exists that supports the use of value added estimates (Harris, 2011). Goe and Croft (2009) found that effective teachers are effective with students of all achievement levels, regardless of the level of heterogeneity in classrooms. Their findings show that effective teachers can overcome school contextual factors such as the nonrandom placement of students in classrooms. Goe and Croft's (2009) argument that homogeneity and heterogeneity of student ability levels within classrooms are not major concerns in assessing teacher effectiveness is comforting for those developing teacher evaluation systems (Sanders, Wright, & Horn, 1997) because it discredits the claims that value added measures would need to control for nonrandom placement of students. If effective teachers are effective regardless of the students in their classrooms, controlling for the ability of the students assigned to them is unnecessary in measuring teacher effectiveness.

Similar to the mixed evidence found with value added estimates, the research on performance-based evaluation tools, specifically teacher observations, is not conclusive or definitive. Jacob and Lefgren (2008) analyzed how well principals can distinguish between more and less effective teachers through observations. They found that principals are good at observing teachers who produce the largest and smallest academic

achievement gains, but principals have far less ability to distinguish between teachers in the middle (Jacob & Lefgren, 2008).

The research concludes that principals are unable to consider all of the necessary details during an observation to make an accurate evaluation of teacher effectiveness (Jacob & Lefgren, 2008). One explanation may be that principals base their observation scores on overall classroom achievement more so than individual student achievement. Principals inaccurately or incompletely recall objective, individual measures of student performance resulting in the overall classroom achievement score as being a more accurate predictor of a strong teacher observation score than student achievement (Jacob & Lefgren, 2008). A second explanation may be that a principal recalls the previous teacher observation, possibility forgetting the nuances of the observation, and this general overall feeling of the teacher as being effective or ineffective heavily influences the subjective evaluation (Jacob & Lefgren, 2008). Finally, Harris (2011) finds that teacher observations are prone to inconsistencies and often vary based on factors that may not have anything to do with teaching performance.

Simply stated, the existing research on evaluation systems is mixed (Hallinger, Heck, & Murhy, 2014). Proponents of value added measures argue that this objective tool removes the subjectivity from teacher evaluations of the past (Koretz, 2008), and opponents argue that inconsistencies in the design and statistical models promote decisional errors that have harmful and far reaching consequences (Goe & Croft, 2009; Hallinger, Heck, & Murphy, 2014). Researchers further critique value added measures as being difficult to implement, unstable measures in any grade level where students have multiple teachers attributing to their learning, and unable to control for classroom

heterogeneity (Haertel, 2013; Hallinger, Heck, & Murphy, 2014; Kane & Staiger, 2008; Sanders, Wright, & Horn, 1997). This makes them unreliable tools for making decisions on individual performance.

Jacob and Lefgren's (2008) analysis on the effectiveness of subjective observations suggests that there is a relatively weak relationship between principal observations of teachers and objective performance (Jacob & Lefgren, 2008). Principal observations produce a binary conclusion; a teacher is either good or bad. Yet, the majority of teachers fall in the middle of the spectrum making principal observations an inconclusive measure for evaluating the majority of teachers. Similar to research on value added models, observation tools produce inconclusive evidence on their effectiveness for evaluating teacher performance. Mixed results of both the subjective tools of performance-based evaluations – observations – and the objective tools – value added models – support the argument that it is not the evaluation framework that matters as much as it is the use of the framework.

If performance-based evaluation is truly meant to build or develop teacher capacity, then teacher's perception of the usefulness seems to depend heavily on context. Consider, as an example, going to the doctor's office and receiving the results from a heart monitor. This is a snapshot of your heart's health. Consider that the measurements indicate that your cardiovascular health needs improvement. If the patient trusts the doctor and trusts the measurement, they will be more likely to acknowledge the problem and buy into the treatment plan. Similarly, evidence from the evaluation process is more likely to be acted on if teachers find value in the information. This study seeks to

determine if aspects of the formal school structure may contribute to a more useful evaluation process.

Next, the literature review shifts to bureaucracy theory and classical management theory to formulate an argument for how school context may influence the perceived effectiveness of an evaluation tool. The argument largely rests on enabling school structure. In schools with enabling structure where employees feel that rules and procedures build a professional culture, promote dialogue, and stimulate problem-solving, employees may view the evaluation as a means for growth (Hoy & Sweetland, 2001). In contrast, teachers in schools with a hindering structure may see the evaluation tool as coercive and punitive.

Formal Structures of Organizations

Formal structure is necessary for creating effective and well-operating systems (Hoy & Sweetland, 2001). Structure determines how roles, power, decision-making, and responsibilities are assigned and how information flows between levels of management (Dalton et al., 1980; Hoy & Sweetland, 2001). As important as the role of formal structures are, the question remains: What type of structure is best? This question poses a myriad of conflicting and paradoxical responses that are contingent on the task complexity of the organization (Dalton et al., 1980; Kohn, 1971). The connection between structure and task complexity has powerful consequences for employee behavior and organization productivity (Organ & Greene, 1981; Pheysey, Payne, & Pugh, 1971; Prichard & Karasick, 1973; Stevens, Philipsen, & Diedericks, 1992; Zeitz, 1983, 1984). Bureaucracy theory offers insight into the use of formal structures in guiding organizational behavior.

Bureaucracy Theory

Bureaucracy theory explains the role of formal structure in organizations (Gouldner, 1954). The most fundamental research on organizational structure comes from Weber's (1978) treatise on bureaucracy theory where he describes two fundamental features of bureaucracy that are central to the discussion of structural control: formalization and centralization (Gouldner, 1954). Formalization and centralization refer to the rules, regulations, and authority structure that define work processes in organizations (Gouldner, 1954).

Formalization is the extent to which the organization has a codified system of rules, regulations, policies, and procedures. Gouldner's (1954) analysis of organizations identified two types of formalization: punishment-centered and representative. Punishment-centered procedures are determined unilaterally by those in power and are used to control those who deviate from the rules (Gouldner, 1954). Representative formalization is more democratic and broadens decisional authority to organizational actors that represent different role groups (Gouldner, 1954). Representative procedures are used as guides to decision-making and action, not fixed responses that are applied unilaterally and unconditionally. Applied to teacher evaluation, representative formalization would seek to use evaluation as a structure to develop teachers and to improve their capacity to meet student needs (Hoy & Sweetland, 2001) whereas a punishment-centered application would sanction teachers who do not meet performance expectations.

Adler and Borys (1996) would later use the terms coercive and enabling formalization to analyze the structural properties of organizations. In their research,

coercive formalization referred to rules and procedures meant to punish subordinates (Adler & Borys, 1996). Such rules tend to hinder productivity, alienate employees, and force compliance to mandates (Adler & Borys, 1996). Enabling formalization provides employees with the professional discretion to make decisions based on the unique circumstances of the situation or problem. This stands in contrast to rules used to coerce individuals into specific behaviors and actions. Applied to teacher evaluation, the notion of coercive and enabling structures suggests that positive experiences with evaluation processes has less to do with the framework itself and more to do with the use of the tool by school administrators.

Centralization is a second dimension of formal organizational structure described in bureaucracy theory (Gouldner, 1954). Centralization defines the authority and control structure of the organization. Low centralization diffuses power and decision-making to organizational actors across different levels of the system whereas high centralization concentrates decision-making at the top where mandates flow unilaterally downward to employees (Hall, 1968). Centralization, much like formalization, can be applied in ways that constrain behavior or expand discretionary authority to more people. Hindering centralization refers to hierarchy that impedes rather than helps solve problems (Hoy & Sweetland, 2001). When outcomes are not met, hindering centralization responds with heightened supervision, standardizing work, and tightening control (Organ & Greene, 1981). In contrast, enabling centralization helps participants solve problems by defining boundaries while retaining distinct roles based on an individual's expertise (Hoy & Sweetland, 2001). Enabling centralization is said to be flexible, agile, cooperative, and collaborative (Hoy & Sweetland, 2001).

So, what does the evidence suggest about the use of formalization and centralization? There is actually a great deal of conflicting conclusions represented in various studies spanning multiple contexts and several decades. On one hand, evidence supports the positive effects of highly structured, bureaucratic environments (Organ & Greene, 1981; Pheysey, Payne, & Pugh, 1971; Pritchard & Karasick, 1973; Stevens, Philipsen, & Diedericks, 1992; Zeitz, 1983, 1984). Kohn (1971) found that individuals who work in organizations with formal hierarchal structures performed more complex tasks that resulted in increased productivity. Organ and Greene (1981) posit that formalization reduces alienation by providing direct, explicit, and precise expectations of each employee's responsibility within the organization and eliminates role ambiguity. These studies attribute the positive effects of structured environments to a decrease in employee role ambiguity resulting in increased employee satisfaction (Kohn, 1971; Organ & Greene, 1981; Seitz, 1984; Stevens et al., 1992).

Conversely, evidence supports the harmful consequences of highly structured environments. Research shows that highly structured environments are associated with lower job satisfaction, role ambiguity, and alienation because employees see their behavior as subject to organization control more so than their professional expertise (Alder & Borys, 1996; Hage, 1965; Hall, 1968; Lam, 2005). Highly structured environments can produce conformity, lack of communication, and stifle innovation (Hoy & Sweetland, 2001). Researchers have found adults to be frustrated with unresponsive structures and the ridged rules of a high structured environment (Hoy & Sweetland, 2001). Lam supports the claim that highly structured environments produce conditions not conducive of employee learning, arguing, "adults are more motivated to learn when

organizational conditions favors individuals to work and learn from one another on a continuous basis" (Lam, 2005, p. 390).

The relationship between organizational structure and employee performance highlights two distinct points about formal organizational structures. Without any level of formalization, employees cannot identify their role within the organization and this affects attitudes, behavior, and job performance (Dalton et al., 1980). Conversely, medium to high degrees of formalization that limit job scope can result in boredom, job dissatisfaction, and low productivity (Dalton et al., 1980). Dalton et al. (1980) suggests that there is an ideal balance between the negative and positive effects of bureaucracy in any organization. Finding an ideal balance is significant for the efficiency and effectiveness by which organizations deliver value to different organizational actors.

Mixed findings on organizational structure have led some scholars to conclude that the right structure depends on the complexity of the task (Griffin, 1991). A highly formalized environment may be suitable for routine, algorithmic tasks, but tight controls for non-standard tasks that require differentiation and adaptation to changing circumstances are likely to stifle performance and alienate those who work in such conditions (Hage, 1965; Thompson, 1961). This is to suggest that generalizations about structure can be reckless without considering the task context and complexity of work.

Task Context and Formal Structure

Task context extends the discussion of formal structures by explaining that the complexity of the task (Hall, 1968; Thompson, 1967) and attributes of organizational structure (Hall, 1968) influence how employees perceive the decisions that come from the structure within which they work. While bureaucracy theory looks at the organization

as a whole, research on task contexts and performance shows that different tasks within the same organization can be perceived at different places on the bureaucracy spectrum (Alder & Borys, 1996; Hall, 1968; Thompson, 1967). Similarly, differences in core tasks have consequences for the perception of organizational structure (Hall, 1967). Research into task context is pertinent to the discussion in that it seeks to find an ideal fit between task complexity and organizational structure.

Thompson (1967) shows that perception of structure is contingent not on the job as a whole, but rather, on the specific tasks within the profession. Thompson (1967) and later Alder and Borys (1996) argue that many organizational models do not account for different tasks and their complexities. They further argue that negative experiences with formal structures are often due to a misalignment of task requirements and organizational design. Employees tend to react positively when high levels of formalization are associated with routine tasks and when low levels of formalization are associated with non-routine tasks (Alder & Borys, 1996). Effective bureaucracy is dependent on the alignment between organizational tasks and the needs of employees (Hall, 1968). This research is significant in that previous scholars studied the aggregate of tasks related to the job, and relatively few studies controlled for task routineness.

Hall's (1968) research shows that within similar professional occupations employees perceive dimensions of bureaucracy differently based on various tasks. Hall studied twelve occupations and found perceptions of formal structures change based on the unique bureaucratic dimension tested. Occupations, as argued by Hall, can be categorized as autonomous, heteronomous, and departments. Within these three general types of occupations Hall measured how employees perceived hierarchy of authority,

division of labor, presence of rules, procedural specifications, impersonality, and technical competence. He found that within similar organizations, an employee may look at one task and perceive the organization as highly bureaucratic whereas a different task within a similar organization may result in the employee perceiving the organization differently. Hall (1968) shows that employee perception of the organization is contingent on the various tasks within the organization (1968), supporting the argument that effective structure depends on organizational tasks and their degree of complexity.

In summary, not all formal structures are the same, nor do they produce similar results. Effective structure depends on alignment with task contexts. Complex tasks, like teaching, benefit from professional structures that provide discretion to employees and facilitate collective problem solving. Simple tasks completed by following linear steps or processes benefit from clear expectations and standardized processes (Forsyth, Adams, & Hoy, 2011). The concept of enabling school structure emerged from evidence describing how formal structures can be applied in ways that maximize individual and group performance.

Enabling School Structure

The concept of enabling school structure emerged from bureaucracy theory and empirical studies on the alignment between task context and employee performance. Recall that bureaucracy theory defines organizational structures as necessary for coordinating work processes in ways that increase operational efficiencies (Alder & Borys, 1996). Formalization and centralization are two bureaucratic features of organizations that are central to enabling school structure (Gouldner, 1954). Formalization is the extent to which the organization has a codified system of rules,

regulations, policies, and procedures. Centralization defines the authority and control structure of the organization (Gouldner, 1954). Formalization varies along a spectrum from enabling or hindering. Enabling structures help members of the organization to achieve their duties and responsibilities, and hindering structures create barriers to that end (Sinden, 2004). For example, an enabling structure would be a policy or procedure that ensures quality and success within a particular job. A hindering structure would be some type of rule or procedure that has no bearing on the accomplishment or failure of work but still needs to be completed for some reason. At worst, hindering structures may outright prevent an employee from being able to complete essential functions (Adler, 1999; Adler & Borys, 1996; Sinden, 2004; Sinden, Hoy, & Sweetland, 2001, 2004).

Enabling school structure is a relatively modern concept that situates bureaucracy theory within an education context (Hoy & Sweetland, 2001). Enabling school structure reflects teacher perceptions of how the formal structure is used to regulate teaching and learning. As the name implies, it represents formal structures that enable teachers and school members to work at peak performance levels. Teacher perceptions of formal school structures vary along a continuum from enabling to hindering (Forsyth, Barnes, & Adams, 2006). Enabling structure promotes problem solving, provides flexible guides that can adhere to the unique context to help employees solve dilemmas, and fosters authentic leadership (Adler & Borys, 1996; Hoy 2002) whereas hindering structures foster increased supervision and standardization of processes and outcomes (Sinden, Hoy, & Sweetland, 2001).

Enabling structure does not mean that schools are devoid of formal regulations or centralized authority. Instead, the continuum from enabling to hindering depends on the

use of rules, regulations, and formal authority as experienced by school members (Sinden, Hoy, & Sweetland, 2004). Three normative features of schools define organizational structures that enable quality school performance: trusting relationships, adapting rules and procedures to their unique contexts, and an authentic leader (Hoy, 2002).

Enabling school structures encourage trusting relations among teachers and between teachers and the principal. Such structures facilitate telling the truth and make it unnecessary to attempt to explain negative performance or hide deficiencies in work output (Hoy, 2002). This in turn can reduce conflict among school professionals. Trusting relationships exist in enabling schools regardless of school size and economic status (Forsyth, Adams, & Hoy, 2011). As theoretically expected, enabling school structure therefore correlates with collegial trust (Hoy, 2002). This evidence within the research shows how enabling school structures create, or at the very least are strongly correlated with, trusting environments.

Enabling school structures have agile processes that allow teachers to adapt and adjust their practices to fit the need and situation. Sinden, Hoy, and Sweetland (2001) in their descriptive study of six schools identified as having enabling structures found that teachers said the school has common sense procedures. Common sense procedures are seen as providing structure that is needed for both students and teachers (Sinden, Hoy, & Sweetland, 2001). Because the rules were considered common sense, flexibility was seen as a tenet in the application of rules and regulations surround teaching and learning. There was also a tacit understanding among faculty that exceptions to the rules were made only for good reasons (Sinden, Hoy, & Sweetland, 2001). Additionally, teachers

demonstrated a positive perception of the rules that resulted in a positive perception of leadership and work processes (Sinden, Hoy, & Sweetland, 2004).

The converse of this relationship is also true. With enabling structure falling on one end of the continuum and hindering structures falling on the other (Sinden, Hoy, & Sweetland, 2001), the positive benefits described within an enabling school structure are mirrored by negative results in a hindering school structure. The research supports this conclusion. Rules and procedures applied in a hindering structure are seen as punitive with the intent for authority to control and discipline (Sinden, Hoy, & Sweetland, 2001), as opposed to shape and transfer knowledge. Within such contexts, employees can often resist or fracture from the administration because the rules force them to comply with artificial standards rather than serve the needs of their students (Sinden, Hoy, & Sweetland, 2001). By its very definition, these structures are hindering.

Another way that enabling school environments affects teacher perception is in relationship to how teachers view authenticity of the leader. Sinden, Hoy, and Sweetland (2001) found that one of the most consistent aspects of the schools with enabling structures was the authenticity of their leader. This was defined within the research as an ability to openly treat teachers as experts in their field. Teachers often mentioned appreciating that they were treated like professionals and that this mutual regard for professionalism existed among the teachers for each other and between administrators, leading teachers to feel as though they were regarded as experts (Sinden, Hoy, & Sweetland, 2001). One respondent in the Sinden, Hoy, and Sweetland study shared that when the principal publically stated that the teacher was the expert in the classroom, it increased her credibility and effectiveness with students.

Teachers felt that this authenticity attributed to an open atmosphere where they were unafraid to make tough decisions and were confident in enforcing discipline (Sinden, Hoy, & Sweetland, 2001). Principals said they treated teachers as experts and were supported first while asking for explanations later (Sinden, Hoy, & Sweetland, 2001). From this perspective, the effects of enabling school structure appear to be reciprocal; essentially, it is not merely the teacher who changes as a result of the structure, but the administration as well, even if it is not a fully conscious change.

The concept of enabling structure applied to teacher evaluation suggests that when teachers experience formal structure as enabling quality performance, they are more likely to find utility in the evaluation process. A hindering environment, in turn, would seem to engender beliefs that the evaluation processes is too controlling and restrictive. The research would suggest that enabling school environments would encourage teachers to view performance evaluations as a tool that can adhere to the unique contexts of their environment as opposed to encouraging conformity to false ideals that may not have any value to improved teaching and learning (Adler & Borys, 1996; Hoy 2002). If teachers view evaluation as useful for professional growth, then it would stand to reason that they would also feel autonomy to make decisions and adjustments to their own teaching that would be in line with feedback received from their principals or supervisors. Put simply, an enabling environment would be more conducive to using formal evaluation as a tool for growth inducing strategies meant to support teacher development. A growth-oriented focus would seem to encourage teachers to be more engaged in the process.

Summary of Bureaucracy Theory and Enabling School Structure

Bureaucracy theory defines organizational structures as necessitating rules that define how employees make decisions (Gouldner, 1954). More specifically, a clearly defined set of roles and jobs, a hierarchy of authority, and standardized processes are fundamental to structure (Gouldner, 1954; Weber, 1978). Bureaucracy theory claims that hierarchy is necessary but how hierarchy is implemented has implications for employee satisfaction in the workplace (Dalton et al., 1980; Ford, 1973; Pierce & Dunham, 1976; Weed, 1971).

Research on enabling structure shows that when hierarchy is perceived as a flexible framework to solve problems, there are positive consequences for the company and employee (Hoy & Sweetland, 2001). School staff that perceive rules as a flexible framework for decision-making show greater trust and respect for their principal (Sinden, Hoy, & Sweetland, 2004). Teachers demonstrate a positive perception of the rules that results in a positive perception of leadership and work processes in schools (Sinden, Hoy, & Sweetland, 2004). Formalization and centralization are necessary. It is not the fact that rules and hierarchy exist within the organization that affects the employee perception of the organizational structure. Rather, it is how employees perceive the rules and structure that affects their work (Sinden, Hoy, & Sweetland, 2001, 2004).

Next, the literature review turns to classical management theory to establish an additional connection between formal structures and useful performance evaluation. Classical management theory provides a different lens than bureaucracy theory to explore the relationship between employee perception of organizational structure and employee affective behaviors. Specifically, classical management theory explains how leaders in

formal positions of authority influence the mindsets and action of their direct reports. Having a lens to view the relationship between structure and employee affective behaviors has implications for this research in that how teachers perceive the school structure has consequences for how they feel or perceive the environment.

Classical Management Theory

As previously argued with the evidence on bureaucracy theory, schools need formal structures to deliver valuable learning experiences to students. Sometimes formal structures can get in the way of good work, but it is also true that rules, regulations, and hierarchy enable quality performance (Organ & Greene, 1981; Pheysey, Payne, & Pugh, 1971; Pritchard & Karasick, 1973; Stevens, Philipsen, & Diedericks, 1992; Sweetland & Hoy, 2001; Zeitz, 1983, 1984). The question behind this research addresses the application of formal structures in ways that facilitate effective performance. Evidence from classical management theory provides additional insight into how organizational arrangements affect employee attitudes and behavior.

Classical management theory was introduced in the late 19th century and became widespread in the first half of the 20th century as organizations tried to address issues of industrial management, including specialization, efficiency, higher quality, cost reduction and management-worker relationships (Hackman, 1980). A key tenet of classical management theory is the division of hierarchy and labor (Hackman, 1980). Generally, organizations operate with three distinct management levels: top management, middle management, and supervisors (Hackman, 1980). Management levels and responsibilities need to be clear and well defined for operations to run smoothly. Labor is divided by

projects broken down into smaller tasks that are easy to complete and accompanied by clearly defined expectations. Workers narrow their field of expertise to specialize in one area. This is believed to increase productivity and efficiency (Hackman, 1980). A breadth of studies grew out of this theory to determine the merits of management on employee job satisfaction and productivity (Hackman, Oldhma, Janson, & Purdy, 1975; Organ & Greene, 1981; Shepard, 1970; Stone & Porter, 1975). Two strands of this research are informative for this study: 1) the effects of task organization on job-satisfaction and 2) leadership authority and legitimacy.

Task Organization and Job Satisfaction

Evidence behind classical management theory suggests that higher levels of job variety are associated with high levels of job satisfaction (Organ & Greene, 1981). Research on the relationship between low job variety and job satisfaction found that as tasks become more automated, employees experience greater alienation from their work and feel less included in the transformational process of their job (Organ & Greene, 1981). Shepard (1970) found that the highest percentages of unsatisfied workers were those exposed to high degrees of rote tasks. Other researchers explored the converse relationship: increased job variety and employee satisfaction. Stone and Porter (1975) studied five individuals at 16 different jobs and found satisfaction with work to be significantly related to task variety and autonomy. Hackman, Oldhma, Janson, and Purdy (1975) studied heterogeneous samples of keypunch operators and found that employees with greater job variety were more satisfied. The results showed a reduction in the number of operators needed, errors made, and employee turnover (Hackman, Oldhma,

Janson, & Purdy, 1975). The company also reported tens of thousands of dollars in cost savings (Hackman, Oldhma, Janson, & Purdy, 1975).

Other researchers looked at the relationship between job autonomy, variety, and ambiguity on job satisfaction. Sims, Keller, and Szilagyi (1976) found that job autonomy and variety were strongly associated with job satisfaction and were negatively associated with role ambiguity (Sims, Keller, & Szilagyi, 1976). Finlay et al. (1995) controlled for job variety, job ambiguity, and job autonomy and found that each had statically significant independent influences on levels of reported job satisfaction (Finlay et al., 1995). While there is no research that explicitly links job satisfaction or affective behaviors to employee perception of organizational features, classical management theory makes a strong case that task environments affect affective beliefs and behaviors of employees (Finlay et al., 1995). In short, higher levels of job autonomy, higher levels of job variety, and lower levels of job ambiguity are associated with higher job satisfaction (Finlay et al., 1995).

Inferring from the above evidence, a case can be made that characteristics of the task environment have consequences for teacher attitudes and perceptions of performance evaluation. Task environments that are narrowly defined and routinized would seem to elicit negative beliefs about formal structures, like the evaluation system; whereas, environments that promote flexibility, adaptability, and variety would seem to engender positive experiences. It seems logical to posit that if employees are satisfied with how the task environment is organized, they would be more likely to hold positive views about structural features like performance evaluation.

Leadership Authority and Legitimacy

Classical management theory explains that organizations need some level of hierarchical decision-making (Hall, 1968). When leaders with formal decision making responsibilities are perceived favorably, employees tend to approve of decisions and feel a heightened sense of well-being and attachment to the organization (Keyes, Hysom, & Lupo, 2000). In turn, outcomes like increased profits, heightened customer satisfaction and loyalty, increased productivity, and improved employee retention increase (Keyes, Hysom, & Lupo, 2000). In short, employee and organizational performance is partly the result of managers who have achieved legitimate authority (Keyes, Hysom, & Lupo, 2000). This has implications for how perceptions of school principals influence affective teacher behavior.

Occupancy of a formal position, like the principalship, does not automatically ensure a leader's legitimacy in that position (Kanter, 1977; Reskin & Ross, 1992; Zelditch & Walker, 1984). Legitimacy is determined based on the leader's action and sources of legitimacy. Zelditch and Walker (1984) discuss the three independently operating sources of legitimacy: authorization, endorsement, and propriety. Authorization is the leader gaining support from individuals higher in an organizational hierarchy. Endorsement involves receiving support from peers and subordinates. Propriety refers to whether an individual approves of the norms. As such, propriety support necessitates a measure of fairness because individuals are more likely to approve of norms that are fair (Zelditch & Walker, 1984).

Legitimate leaders receive support from their peers, subordinates, and supervisors (Keyes, Hysom, & Lupo, 2000); in turn, legitimate leaders are likely to generate trust and

personally experience more positive emotions (Ford & Johnson, 1998; Johnson & Ford, 1996; Keyes, Hysom, & Lupo, 2000; Ridgeway, 1989; Zelditch & Walker, 1984). Legitimate leaders share information, provide subordinates autonomy over their work, and yield influence to their subordinates (Keyes, Hysom, & Lupo, 2000). As a result of providing subordinates opportunities for autonomy and input, subordinates are likely to experience more positive emptions when working for a leader who has established legitimacy in his/her abilities to lead the organization (Keyes, Hysom, & Lupo, 2000).

Evidence in classical management theory show that legitimate leaders promote a host of positive outcomes and feelings in employees that result in positive organizational performance (Hackman, 1980; Hackman, Oldhma, Janson, & Purdy, 1975; Organ & Greene, 1981; Shepard, 1970; Stone & Porter, 1975). When employees have increased job autonomy, increased job variety, and decreased role ambiguity, employees have greater job satisfaction and view their leader as legitimate (Finlay et al., 1995). Perceiving the leader as legitimate results in employees feeing a heightened sense of well-being and attachment to the organization (Keyes, Hysom, & Lupo, 2000), resulting in better worker satisfaction and loyalty, increased productivity, and improved employee retention (Keyes, Hysom, & Lupo, 2000). Legitimacy established by a school principal holds promise as an essential condition for useful teacher evaluation.

Conclusion to the Review of Literature

The lenses of bureaucracy theory and classical management theory show how domain specific tasks can be viewed differently under various organizational structures. Bureaucracy theory explains the effects of structure in an organization. Its two facets –

formalization and centralization – describe how authority gets carried out within organizations and how work is regulated (Gouldner, 1954). Researchers would later use the terms coercive and enabling formalization and centralization (Adler & Borys, 1996). Hierarchy exists in all organizations; the difference between organizations is in how leadership uses rules and authority to achieve strategic objectives.

Evidence in bureaucracy and classical management theories have implications for the current study in showing that the context of the environment matters in how employees perceive their organization (Alder & Borys, 1996; Dalton et al., 1980; Griffin, 1991; Hall, 1968; Thompson, 1967). The literature does not specifically address the relationship between employee perceptions of organizational features and usefulness of teacher evaluation. That stated, it seems reasonable to believe that when the evaluation tool is implemented in a coercive environment, the tool negatively affects employees' affective behaviors by stifling their creativity with a rigid and punitive framework. Conversely, in an environment that promotes creativity and dialogue, employees could view the evaluation tool as a means of stimulating problem-solving. Bureaucracy theory and classical management theory lay out the argument that effectiveness of the evaluation tool is contextually based. That is, teacher perceptions of the evaluation process depend on the structure and context in which the evaluation is used to improve teaching practice.

Chapter 3: Rationale and Hypotheses

The literature review presented evidence that leads to the argument underlining the hypotheses for this study. First, evidence on the effectiveness of new performance-based teacher evaluation is mixed (Coggshall, Max, & Bassett, 2008; Daley & Kim, 2010; Hallinger, Heck, & Murphy, 2014; Kane, Rockoff, & Staiger, 2008; Koretz, 2008; Murphy 2013; Sanders & Horn, 1998). Second, evidence in bureaucracy theory explains how formal structures, like an evaluation tool, may be used in ways that increase worker satisfaction and performance (Alder & Borys, 1996; Finlay et al., 1995; Hoy & Sweetland, 2001). Third, evidence in classical management theory singles out the role of managers and leaders in generating working conditions that support employees and enhance their attitudes and performance (Hackman, 1980). These three strands of research combine to suggest that performance-based teacher evaluation can have differential effects on teachers depending on characteristics of the formal organizational context.

Performance-based evaluations are a necessary element of the formal school structure. Schools are required by law to evaluate teachers on an annual basis. Evaluation, however, should achieve more than simple regulatory compliance. It should be used as a process to provide instructional feedback, to distinguish among different levels of effectiveness, to make explicit professional standards, and to be used for improvement purposes (Hallinger, Heck, & Murphy, 2014). Teachers deserve evaluation protocols and processes that set forth professional expectations and provide accurate performance feedback. The challenge is in creating an environment where teachers perceive the evaluation process as useful. Implementation of new performance-based

evaluation systems affect how teachers experience and perceive the tool, yet effective implementation remains inconsistent across many schools (Hallinger, Heck, & Murphy, 2014).

Evidence on enabling structures explains why some teachers may experience teacher evaluation as effective and useful whereas others may not find much value in the process. The problem is not with formal structure, but rather the use and application of formal structure to coordinate work processes (Adler, 1999). When teachers experience formal rules and regulations as enabling effective performance, they generally have better attitudes toward teaching, are more likely to persist in challenging tasks, view leadership favorably, and experience a more professionally supportive environment (Hoy & Sweetland, 2000, 2001; Tschannen-Moran, 2009). Rules and regulations experienced as hindering have the opposite effect; they can evoke cynicism, engender alienation, and breed discontent among organizational members (Adler, 1999; Adler & Boyers, 1996; Sinden, Hoy, & Sweetland, 2004).

Given the above evidence, it seems reasonable to believe that teachers who perceive rules and regulations as enabling would view teacher evaluation as a tool that can support their professional growth. Continuing with this reasoning, teachers who perceive rules and regulations as enabling are likely to be more receptive to the principal's use of evaluation as a means of instructional improvement. Conversely, negative experiences with formal rules and regulations may engender beliefs that the evaluation process is too controlling and punitive. Thus, it is predicted that,

H1: Teachers who perceive the school as having an enabling formalization will perceive the evaluation tool as effective.

Centralization is a second dimension of the formal school structure that has likely consequences for the perceived usefulness of teacher evaluation. Centralization exists in the formal authority structure of the school and school system (Hoy & Sweetland, 2000, 2001). A highly centralized school confines decisions and influence to the formal authority of administrators; whereas, low centralization diffuses influence and decisions to school members with relevant knowledge and expertise to solve problems and issues they encounter (Sinden, Hoy, & Sweetland, 2004). Enabling centralization supports teachers in solving problems, addressing issues, planning for the future, and facilitating effective work. In contrast, hindering centralization impedes problem solving by relying on external controls, like close supervision, threats, and punitive consequences, to manage teaching and learning (Hoy & Sweetland, 2000, 2001).

Similar to formalization, it is reasonable to believe that teachers who experience centralization as enabling their growth and development as professionals would view teacher evaluation in a favorable light. Negative views of the hierarchy, on the other hand, would seem to imply that evaluation is used more to control teacher behavior rather than to support improvement. Additionally, a negative perception of authority would seem to present a psychological barrier that would also affect the perceived usefulness of the evaluation framework. Thus, it is predicted that,

H2: Teachers who perceive the school as having an enabling centralization will perceive the evaluation tool as effective.

Classical management theory has been used to examine how organizational structures, division of labor, and leadership contribute to employee affective beliefs and behaviors (Hackman, 1980). Evidence indicates that in organizations where employees have increased job autonomy, increased job variety, and decreased role ambiguity, employees have greater job satisfaction and view their leader as legitimate (Finlay et al., 1995). Additionally, legitimate leaders are more effective in their role, establish stronger relationships, and support a more positive and productive work environment (Hackman, Oldhma, Janson, & Purdy, 1975; Organ & Greene, 1981; Shepard, 1970; Stone & Porter, 1975). A legitimate leader is someone who is proficient in the skills necessary to successfully fulfill their responsibilities, views the processes and procedures at their disposal as tools to adhere to context-specific situations, and takes risks with open and transparent communication, confident that their actions will be met in positive ways (Finlay et al., 1995). This suggests that leaders who establish legitimacy can be more effective in their use of formal structures to spur better outcomes.

In this way, a legitimate leader embodies the key tenets of trust. They are open, honest, competent, reliable, and benevolent (Hoy & Tschannen-Moran, 1999). Trustworthy leaders establish strong commitment to organizational visions, they are more effective in implementing improvement strategies, they build supportive relationships, and they elevate the collective performance of the organization (Tschannen-Moran, 2014). Specifically in schools, faculty trust is positively associated with school effectiveness (Forsyth, Adams, & Hoy, 2011; Hoy, Tarter, & Wiskowskie, 1992; Tarter, Sabo, & Hoy, 1995), student achievement (Goddard, Tschannen-Moran, & Hoy, 2001),

positive school climate (Tschanen-Moran & Hoy, 1998), and principal authenticity (Hoy & Kupesmith, 1984).

It is reasoned here that trust creates the psychological safety needed for teachers to engage authentically in the evaluation process. Extending evidence in classical management theory, school principals who are perceived as trustworthy are likely to have established the legitimacy needed to work effectively with teachers during the evaluation process. Positive interactions in the context of evaluation, as facilitated by trust, can function as lubricant for a more effective evaluation experience. Thus, it is predicted that,

H3: Teachers in schools with high faculty trust in principal will perceive the evaluation tool as useful.

Chapter 4: Methods

Research Design

The study involves survey research and seeks to use a causal-correlational design to measure the relationship between enabling school structure, principal trust, and teacher perceived usefulness of the evaluation framework. The data for the empirical investigation were hierarchically structured with teachers nested in schools. The primary interest was to determine the degree to which formal school structures shaped teacher perceived utility of performance evaluation. Thus, Hierarchical Linear Modeling (HLM) was used to test three hypotheses.

The HLM analysis followed a conventional model building process. First, an unconditional null model was run to decompose variance in each dependent variable to within- and between-school factors. Results of the null model were used to calculate the Intraclass Correlation Coefficient (ICC). Second, a random coefficient regression was modeled to test the effects of teacher characteristics on their perceptions of performance. Finally, a random effects ANOVA model was used to test the hypotheses while controlling for school conditions and teacher characteristics. Predictor variables for this final model were entered in a stepwise manner with statistically significant variables retained and included in a final combined model.

District Context

The school district is located in a city with a metropolitan population of approximately 950,000 residents. At the time of the study, the district served approximately 42,000 students across 88 sites. Of the 42,00 students, approximately 31 percent identify as African-American, 29 percent as Caucasian, 25 percent as Hispanic, 8

percent as Native American, and 2 percent as Asian. Eighty three percent of the students qualified for the federal lunch subsidy. Nearly 2,400 teachers are employed in the district. Teachers average 10 years of teaching experience and approximately 25 percent of teachers hold advanced degrees. At the time of the study, the evaluation tool had been used in the district for four years (Tulsa Public Schools, 2012).

Evaluation Tool

The performance-based evaluation system developed by the district relies on a standards-based evaluation instrument, classroom observation, principal or assistant principal feedback, multiple performance levels to differentiate teacher effectiveness across standards, and value added data linked to teachers. Teacher evaluation ratings result from a yearlong observation and evaluation process. Probationary teachers receive four classroom observations ranging from 20-30 minutes and two formal evaluations while non-probationary teachers have two classroom observations for 20-30 minutes and one formal evaluation. Feedback conferences follow observations and the summative evaluation. At the time of this study, value added estimates were not calculated in the composite teacher effectiveness ratings but were provided to teachers (Tulsa Public Schools, 2014).

The classroom observation and evaluation instrument includes five domains: classroom management, instructional effectiveness, personal growth, interpersonal skills, and leadership professional practice. Teachers receive a score ranging from 1-5 (1= ineffective, 2 = needs improvement, 3 = effective, 4 = highly effective, and 5 = superior) on 20 indicators of effective teacher behavior (Appendix B). Each domain is weighted for the composite teacher effectiveness score. Classroom management consists of 30 percent

of the composite score, instructional effectiveness 50 percent, performance growth 10 percent, interpersonal skills 5 percent, and leadership 5 percent.

Data Source

Data were collected by the Oklahoma Center for Education Policy (OCEP) as part of its ongoing study of school and district capacity. OCEP collected teacher data with electronic surveys emailed to certified teachers in 71 schools during the 2013-2014 academic year. Teachers within each school were randomly sampled and assigned to either survey form A or form B. Teacher perceptions of the evaluation process come from survey form A, and teacher perceptions of school structure and principal trust were measured in survey form B. This separation between predictor variables and the criterion variable helps to address common measurement bias found in survey research. The response rate for faculty completing survey A was 65 percent and 62 percent for those randomly assigned to complete survey B.

Measures

Enabling School Structure

Enabling school structure was operationalized as its two conceptually distinct components: formalization and centralization (Hoy & Sweetland, 2001). Formalization consists of rules and regulations within an organization. Centralization involves the leadership hierarchy in the school and how it is used to control teacher behavior (Adler & Borys, 1996; Hoy, 2002; Sinden, 2001). Conceptually, these are distinct, yet related, dimensions of the formal organizational structure of schools. As such, they were operationalized as two variables (Forsyth & Adams, 2014).

The enabling formalization scale includes six items with a Likert response set ranging from "strongly disagree," coded as 1, to "strongly agree," coded as 6. Sample items include: "In this school, red tape is a problem," "Administrative rules in this school enable authentic communication between teachers and administrators," and "Administrative rules help rather than hinder."

The enabling centralization scale consists of 6 items with a response set ranging from "strongly disagree," coded as 1, to "strongly agree," coded as 6. An exploratory factor analysis was performed on the centralization items. Principal axis extraction was used with no rotation. Results show that the items load strongly on one factor loadings ranging from 0.73 - 0.84 (Appendix A). The survey asked respondents six questions to be answered on a scale of 1-6 with higher numbers indicating more often. Examples of survey questions that respondents were asked to gauge teacher perception of centralization included six questions to be answered on a scale of 1-6 with higher numbers indicating stronger agreement:

1. The administrative hierarchy of this school enables teachers to do their job.
2. The administrative hierarchy obstructs student achievement.
3. The administrative hierarchy of this school facilitates the mission of this school.
4. The administrative hierarchy of this school obstructs innovation.
5. In this school the authority of the principal is used to undermine teachers.
6. The administrators in this school use their authority to enable teachers to do their jobs.

Faculty Trust in Principal

Faculty trust in principal is operationalized as faculty perceiving the principal as benevolent, reliable, competent, open, and honest (Hoy & Tschannen-Moran, 1999). Higher principal trust indicates that faculty respect and trust the principal's leadership. The faculty trust in principal scale includes six items with a Likert response set ranging from "strongly disagree," coded as 1, to "strongly agree," coded as 6. An exploratory factor analysis was performed on the faculty trust in principal items. Principal axis extraction was used with no rotation. Results show that the items load strongly on one factor loadings ranging from 0.73 - 0.95 (Appendix A). Examples of survey questions that respondents were asked to gauge teacher perception of faculty trust in principal included seven questions to be answered on a scale of 1-6 with higher numbers indicating stronger agreement:

1. The teachers in this school have faith in the integrity of the principal.
2. Teachers in this school trust the principal.
3. The principal in this school typically acts in the best interest of teachers.
4. The principal of this school does not show concern for the teachers.
5. Teachers in this school can rely on the principal.
6. The principal in this school is competent in doing his or her job.
7. The principal doesn't tell teachers what is really going on.

Teacher Perceived Utility of Performance Evaluation

Teacher perception of evaluation tool is operationalized as teacher perception of their understanding of the implementation and value of the evaluation rubric and process. The teacher perceived utility of the performance evaluation scale includes six items with a Likert response set ranging from "strongly disagree," coded as 1, to "strongly agree," coded as 6. An exploratory factor analysis was performed on the faculty trust in principal items. Principal axis extraction was used with no rotation. Results show that the items load strongly on one factor loadings ranging from 0.53 - 0.84 (Appendix A). Examples of survey questions that respondents were asked to gauge teacher perceived utility of performance evaluation included nine questions to be answered on a scale of 1-6 with higher numbers indicating stronger agreement:

1. I understand the 4 domains of the TLE rubric.
2. Evaluation domains and dimensions were made clear in a pre-observation conference.
3. The evaluation process helped me develop as a teacher.
4. I am confident the evaluation process fairly reflects my teaching effectiveness.
5. Face to face feedback from the evaluation was provided after each observation.

Control Variables

Teacher and school control variables were included in the model so to account for plausible rival explanations. Teacher controls included the total years of teaching experience, the number of years in their current school, gender, and free and reduced

price lunch rate. The teacher's total years of teaching experience was operationalized as the total number of years in the teaching profession. Number of years in current school was operationalized as the number of academic years a teacher had worked in their current school. Gender was dummy coded as 1 for female and 0 for male. School controls included the percentage of students in the school that qualify for the government subsidized free or reduced price lunch program.

Analytical Technique

The data for the empirical investigation were hierarchically structured with teachers nested in schools. The primary interest was to determine the degree to which school formal structures shaped teacher perceived utility of performance evaluation. Thus, Hierarchical Linear Modeling (HLM) was used to analyze variance around teacher perceived utility of evaluation. HLM assumes that predictors have a linear relationship with the dependent variable, that level one and level two errors are normally distributed and uncorrelated across levels, level one error is constant, and predictors at the highest organizational level are independent (Woltman et al., 2012).

A conventional modeling building process in HLM 7.0 was used to test the three hypotheses. First, an unconditional null model was run to decompose variance in each dependent variable to within- and between-school factors. Results of the null model were used to calculate the Intraclass Correlation Coefficient (ICC). ICC estimates the percent variance at the group level and individual level. Second, a random coefficient regression was modeled to test the effects of teacher characteristics on their perceptions of the performance. Finally, a random effects ANOVA model was used to test the hypotheses while controlling for school conditions and teacher characteristics. Predictor variables for

this final model were entered in a stepwise manner with statistically significant variables retained and included in a final combined model.

The ICC is calculated by dividing the variance components between groups by the total variance in the sample. The calculation shows that 23 percent of the variance in teacher perceived usefulness of the teacher evaluation tool was attributed to enabling formalization. The ICC calculation shows that 21 percent of the variance in teacher perceived usefulness of the teacher evaluation tool was attributed to enabling centralization and that 32 percent of the variance in teacher perceived usefulness of the teacher evaluation tool was attributed to faculty trust in teacher. These variances were statically significant.

Unconditional Model (Null Model)

Level I: Teacher Evaluation = $\beta 0 + r$

Level II: $\beta 0 = \gamma 00 + u$

Random Coefficient Regression Model

Level I: Teacher Evaluation = $\beta 0+\beta 1$(female.)$+\beta 2$(years in current school)$+r$

Level II: $\beta 0 = \gamma 00$

$\beta 1 = \gamma 10$

$\beta 2= \gamma 20$

Random Effects ANCOVA Model

Level I: Teacher Evaluation = $\beta 0+\beta 1$(female.)$+\beta 2$(years in current school)$+r$

Level II: $\beta 0 = \gamma 00 + \gamma 01$ (school FRL rate) $+ \gamma 02$ (percent non-minority) $+ \gamma 03$ (formalization) $+ u0$

$\beta 1 = \gamma 10$

$\beta 2 = \gamma 20$

Post Hoc Analysis

Very few teachers have the ability to unilaterally choose the evaluation tool. Such decisions are most often handed down from district leaders to principals to implement. Arguing that teachers know that principals are implementing an evaluation protocol that they themselves did not select, it is logical to assume that teachers would view the district as the authority in this context thereby making the significance of individual teacher trust in district administration a plausible factor in the perceived usefulness of the evaluation tool. High trust in district administration makes teachers more open to change and willing to embrace processes that are designed to elevate teaching practice. Low trust, in contrast, prevents teachers from risking vulnerability and can evoke cynicism toward district goals and strategies.

The post hoc analysis examined the relationship at the individual-level between individual teacher trust in the district and teacher perception of the evaluation tool. It was important to seek explanations for individual teacher factors that explain their experiences with the evaluation process. Teacher trust in district administration is an individual teacher belief that affects teacher attitudes and behavior. Trust in district administration signals a type of affective endorsement for policy tools used by the central office to improve teaching and learning. Consistent with classical management theory, when leadership advances formal policies perceived as legitimate, employees tend to approve and embrace formal structures that can enhance their work (Hall, 1968). Thus, the purpose of the post hoc was to test the relationship between teacher trust in district administration and teacher perceived usefulness of the evaluation tool.

Measure and Analysis

The teacher trust in district administration scale measures teacher perception of the district's willingness to be vulnerable to another party based on the confidence that the district is benevolent, competent, reliable, open, and honest (Hoy, 2002; Hoy & Tschannen-Moran, 1999; Tschannen-Moran & Hoy, 1998). More specifically, it assesses faculty perceptions of the degree to which district administration is aware of relevant issues, organized, committed, and supportive of teachers' autonomy and professional growth.

Individual teacher trust in district administration is the unit of analysis for this study. Thus, the data requires a multilevel or hierarchically structured analysis to measure the effects of district conditions on individual teacher perceptions of the evaluation process. Levels of grouped data are a commonly occurring phenomenon. In the education sector, data are often organized at the student, classroom, school, and district levels. Analysis of hierarchical data takes into account that contexts at any particular level influence the data at other levels. To best account for the hierarchy, HLM was used to analyze variance in the outcome when the predictor variables are at varying hierarchical levels (Woltman et al., 2012). A conventional modeling building process in HLM 7.0 was used to test the hypothesis.

Random Effects ANCOVA Model

Level I: Teacher Evaluation = $\beta 0+\beta 1$(female)$+\beta 2$(years in current school)$+$ $\beta 3$(teacher trust in district administration) $+ r$

Level II: $\beta 0 = \gamma 00 + \gamma 01$ (school FRL rate) $+ \gamma 02$ (percent non-minority) $+ \gamma 03$ (formalization) $+ u0$

$\beta 1 = \gamma 10$

$\beta 2 = \gamma 20$

Chapter 5: Results

The empirical part of the study set out to test three hypotheses about the relationship between features of formal school structure and teacher perception of performance evaluation. Results of the empirical tests are presented in this section. The section begins with descriptive statistics and correlation results that describe the sample of teachers and schools and the bivariate relationships among the teacher level variables. Next, results of the HLM analysis are reported to test the hypotheses. The chapter concludes with results of the post hoc analysis.

Teacher and School Level Descriptives

Descriptive data are reported in Table 1. For teacher level data, the sample represents teachers who have taught between one and thirty years with the average length of time being 12.96 years. The average length of time teachers had been in their current school was about half as long, 6.15 years with a minimum of 1 year and a maximum of 30 years. Nine percent of teachers in the sample were National Board Certified, and 85 percent were female. Teacher perceived usefulness of the teacher evaluation system is measured on a Likert-scale ranging from "strongly agree" to "strongly disagree" with higher scores indicating responses of greater agreement. The average response was 4.10 with a minimum of 1.14 and a maximum of 6.00. This is an aggregate of responses from nine questions that respondents answered. An average, aggregate response of 4.10 shows that teachers "somewhat agree" that the Teacher Leadership Effectiveness tool is useful.

For school level data, the average Free or Reduced Lunch (FRL) rate, a commonly used proxy of school poverty, was 86 percent, with a minimum of 17 percent

and a maximum of 100 percent. The average percent of non-minority students was 36 percent with a minimum of 6 percent and a maximum of 79 percent.

Faculty trust in principal, enabling formalization, and enabling centralization are an aggregate of questions that respondents answered on a Likert-scale ranging from "strongly disagree" to "strongly agree." The average response for faculty trust in principal was a 4.45 on a range from 1.90 to 5.60. The average response for enabling formalization was 4.09 on a range from 2.20 to 5.17 showing that teachers "somewhat agree" that the school fosters enabling conditions. The average response for enabling centralization was 4.40 on a range from 2.46 to 5.63, showing that teachers "somewhat agree" that the school fosters conditions of enabling centralization.

Table 1: Descriptive Statistics for Teacher and School Characteristics

	Mean	SD	Min	Max
Teacher Level				
Years Experience	12.96	9.08	1.00	30.00
Years in Current School	6.15	6.45	1.00	30.00
National Board Certified	0.09	0.29	0.00	1.00
Female	0.85	0.36	0.00	1.00
TLE Usefulness	4.10	0.99	1.14	6.00
School Level				
FRL Rate	86.38	20.32	17.00	100.00
Percent non-minority	36.02	18.27	6.00	79.00
Index Score	61.22	18.31	8.00	102.00
Faculty Trust in Principal	4.45	0.81	1.90	5.60
Enabling Formalization	4.09	0.60	2.20	5.17
Enabling Centralization	4.46	0.65	2.46	5.63

Note. n=71 schools, n=572 teachers

Correlation Results

Table 2 reports results of the bivariate correlation analysis of teacher level variables. These data show the strength of the relationship between teacher characteristics and perceived usefulness of the evaluation tool. The results were used to determine if teacher characteristics should be entered in the HLM models.

There was a statically significant relationship between female and teacher perception of the evaluation system ($r = .10$, $p<.05$). Meaning that, on average, female teachers had a slightly higher perception of the favorableness of the evaluation tool than male teachers. Although statistically significant, the strength of the relationship was small. There was also a statically significant relationship between years in current school and perceived usefulness of evaluation ($r = -.102$, $p<.05$). The relationship was negative, indicating that the more years teachers have been in the school, the lower their perception of the evaluation system. Overall, the correlation results show that gender and years in the current school had statistically significant relationships with teacher perceptions of performance evaluation. Female had a positive association whereas years in the current school had a negative association. Although neither relationship was very strong, they were statistically significant, providing enough evidence to include both variables in the HLM analysis.

Table 2: Correlation Results for Teacher Level Variables

	TLE Useful.	NBC	Years in Current School	Years Exp.	Female
TLE Usefulness	1	-.002	-.102*	-.040	.102*
National Board Certified		1	.005	.014	-.018
Years in Current School			1	.533*	.042
Years Experience				1	.055
Female					1

Note. *p-value<.05 ** p-value<.01

HLM Results

Table 3 reports the within school and between school variance for teacher perceived usefulness of the evaluation system. These variance components show the degree to which differences in teacher perceived usefulness of the evaluation tool can be attributed to school and teacher level factors where Level-1, *r* is teacher level, and INTRCPTI, *u0* is school level. The between school differences in teacher perceived usefulness was statically significant ($\chi^2 = 111.84$, p<.01). The variance components were used to calculate the ICC. The calculation shows that 7 percent of the variance in teacher perceived usefulness of the teacher evaluation tool was attributed to school differences. This 7 percent variance attributed to schools was statically significant.

Table 3: Teacher and School Factors of Perceived Usefulness of TLE

Random Effect	Standard Deviation	Variance Component	*d.f.*	*x squared*	*p*-value
INTRCPTI,*u0*	0.27234	0.07417	67	111.84531	<0.001
Level-1, *r*	0.95887	0.91942			
ICC	.07				

Table 4 presents the result of a random effect ANCOVA that was used to test the three hypotheses. The school level predictor variables were entered individually in a step-wise pattern. Statistically significant variables were retained and included in a combined model. This approach shows the unique effect of each variable when all the variables were combined in the same model.

Model 1 includes the teacher and school control variables. At the teacher level, female had a statistically significant relationship with perceived usefulness ($\beta_1 = .24$, $p<.01$) and so too did years in the current school ($\beta s = -.12$, $p<.05$). The school level controls of FRL rate and percent non-minority were not related to perceived usefulness.

In model 2, enabling formalization was added to the model. Results show a statistically significant relationship with perceived usefulness ($\gamma_3 = .14$, $p<.05$). As the perceived formalization of the school structure increases by one standard deviation, the perceived usefulness of the teacher evaluation system increases 0.14 standard deviations. Perceived enabling formalization explains approximately 25 percent of the 7 percent school level variance in perceived usefulness of the teacher evaluation system.

In model 3, enabling centralization was added. Results show that centralization had a statistically significant relationship with teacher perceived usefulness of the

evaluation tool (γ_3=.17, p<.05). As the perceived centralization of the school structure increased by one standard deviation, the perceived usefulness of the teacher evaluation system increased by 0.17 standard deviation. Perceived enabling centralization explained about 38 percent of the 7 percent school level variance in teacher perceived usefulness of the evaluation tool as being attributed to variables.

Faculty trust in principal was added in model 4. Results show that faculty trust in principal had a statistically significant relationship with perceived usefulness (γ_3 = .13, p<.05). As the perceived centralization of the school structure increased by one standard deviation, the perceived usefulness of the teacher evaluation system increased by 0.13 standard deviation. Of the 7 percent variance in teacher perceived usefulness of the teacher evaluation tool, 25 percent was attributed to faculty trust in principal.

The combined model includes each of the predictor variables along with the school and teacher level controls. When formalization, centralization, and faculty trust in principal are in the model together, centralization had the strongest unique effect on perceived usefulness of performance evaluation (γ_4 = .31, p<.05). Enabling formalization and faculty trust in principal were not statistically significant. It is important to point out intriguing changes to the parameter estimates in the combined model. Notice that the effects of enabling formalization and principal trust changed from positive to negative. Also, notice that the effect of enabling centralization increased from .17 to .31. These changes to the parameter estimates point to potential multi-collinearity among centralization, formalization, and trust. The strong associations among these is likely affecting the estimated relationships in the combined model.

Table 4: HLM Results from the Random Effects ANCOVA

Fixed Effects	Model 1	Model 2	Model 3	Model 4	Combined Model
Intercept	.01 (.05)	.01(.05)	.01 (.05)	.01 (.05)	.01 (.04)
FRL	-.00 (.08)	.05 (.08)	.07 (.08)	.03 (.08)	.07 (.08)
Percent Non-Minority	.08 (.08)	.06 (.07)	.07 (.07)	.05 (.07)	.08 (.07)
En Formalization	-	.14 (.06)*	-	-	-0.12 (.15)
En Centralization	-	-	.17 (.06)*	-	0.31 (.16)
Faculty Trust in Principal	-	-	-	.13 (.05)*	-0.03 (.12)
Female slope	.24 (.12)*	.22 (.11)*	.24 (.11)*	.24 (.11)*	.26 (.11)*
Years in School slope	-.12 (.04)**	-.12 (.04)**	-.12 (.04)**	-.12 (.04)**	-.12 (.04)**
S-level Variance Explained		25%	38%	25%	37%

Note. *p-value<.05 ** p-value<.01. All estimates are standardized with a mean of 0 and a standard deviation of 1.

Post Hoc

Results of the post hoc analysis are presented in Table 5. Teacher trust in district administration was added to the model along with the teacher control variables, school controls, and enabling centralization. Enabling centralization was retained as it was the strongest school level predictor of teacher perceived usefulness of performance evaluation.

It is interesting to first note the percent of explained school level variance. With teacher trust in district administration included at level one, 75 percent of the school level variance was explained. This was an increase of about 37 percent from the best fitting model from the original analysis (model three), suggesting that there may be some grouping effects associated with teacher trust in district administration. The unique effect of district trust was statistically significant and strong ($\beta_3 = .44$, $p<.01$). A one standard deviation increase in district trust was associated with nearly a half standard deviation increase in perceived usefulness of performance evaluation. By itself, district trust explained approximately 19 percent of the teacher level variance in the perceived usefulness of teacher evaluation.

Table 5: HLM results from the Post Hoc Analysis

Fixed Effects	Model 1
Intercept	
TTDA	0.44 (0.4)**
FRL	0.01 (0.07)
Percent Non-Minority	0.02 (0.06)
Female	0.14 (0.10)
Years in School	-0.07 (0.04)*
Enabling Centralization	0.15 (0.07)
Teacher Level Variance Explained	19%
School Level Variance Explained	75%

Note. **p*-value<.05 *** p*-value<.01; TTDA = Teacher Trust in District Administration. Variables in the analysis were standardized to a mean of 0 and a standard deviation of 1.

Chapter 6: Discussion

This study sought to address a gap in the literature by testing the relationship between features of school structure and teacher perceived usefulness of performance evaluation. Informed by evidence in bureaucracy and classical management theories, the study proposed that perceived effectiveness of teacher evaluation was not based on the quality of the evaluation tool itself, but rather on the organizational structures within which the tool was used. As such, the empirical analysis tested the influence, if any, of school structure on the perception of the effectiveness of the performance-based teacher evaluation tool. Now, results of the study are explained through the lenses of bureaucracy theory and classical management theory.

Formalization and Teacher Evaluation

It was predicted there would be a positive relationship between enabling formalization and perceived usefulness of the evaluation system. The hypothesis had moderate support. When enabling formalization was entered into the model with teacher and school controls, it had a statistically significant relationship with perceived usefulness of performance evaluation. Results showed that formalization explained 25 percent of the school level variance in teacher perceptions of evaluation. The effect of enabling formalization faded in the combined model when it was considered alongside enabling centralization and trust in principal. This result should not be mistaken to mean that enabling formalization has no relationship to favorable perceptions of evaluation when taking into account other school factors. Rather, it is more likely the case that the high correlation between enabling formalization and centralization affected the estimated relationships in the combined model.

To understand why enabling formalization has a relationship to perceived usefulness of teacher evaluation, it is necessary to return to bureaucracy theory. Bureaucracy theory argues that organizations function at optimal levels when formal structures organize and guide the actions of organizational actors (Gouldner, 1954; Weber, 1978). That is to say that every organization needs some level of formalization. Alder and Borys (1996), and later Hoy and Sweetland (2001, 2004), revised this basic argument to suggest that efficient and effective organizational performance occurs when formal rules and regulations provide adequate discretion in how work gets accomplished and problems get solved. This stands in contrast to rules and regulations that hinder, or get in the way of, quality performance.

Teacher evaluation is part of the formal school structure. It is necessary for effective and efficient school performance, but its use can vary from the enabling end of the continuum to the hindering. As results in this study indicate, teacher perception of the school environment as fostering conditions of enabling formalization positively correlates with teacher perception of the evaluation tool as useful. This finding has implications for research and practice by suggesting that rules and regulations used to support teacher growth and performance provide a nurturing environment for teacher evaluation to be used as a support mechanism.

Logic dictates that teachers want feedback without fear of punitive action. From this standpoint, teachers who perceive their school environment to be one in which following or not following a specific rule or set of rules would determine employment status would be concerned about any evaluation that determined whether or not those rules had been achieved. Furthermore, teachers may take issue with the interpretation of

the evaluation, especially if the criterion used to judge them was ambiguous or subjective. In contrast, teachers in an environment that enables growth are disposed to interactions and routines oriented toward continuous improvement. The evaluation framework fits in such an environment for it directs improvement conversations, and provides valuable evidence, in instructional areas that affect student learning. There is a clear difference between using formal evaluation as an external control and using it to provide meaningful information on teaching and learning processes. The latter use tends to align with an enabling environment whereas the former reflects a hindering one.

Enabling Centralization and Teacher Evaluation

It was hypothesized that there would be a relationship between enabling centralization and perceived usefulness of the teacher evaluation. The data show that the average perceived level of centralization within the schools was 4.40 on a scale of 2.46 to 5.63. This would fall in the effective range. Higher scores indicate greater agreement with statements gauging the level of perceived centralization within the school. The evidence showed that centralization had a statically significant relationship to teacher perception of the usefulness of the evaluation tool ($\beta = .17$, $p < 0.05$). Enabling centralization explained about 38 percent of the school level variance in perceived usefulness of the teacher evaluation system.

To understand the relationship between enabling centralization and formal teacher evaluation, it is necessary to return to the literature. Bureaucracy theory explains effective organizational performance as a function of structural features that coordinate decisions and actions of organizational actors (Alder & Borys, 1996; Hoy & Sweetland, 2001, 2004). Enabling centralization reflects an environment where decisional authority

comes from expertise and is situated within professionals who are closer to the core tasks (Hoy & Sweetland, 2001). Hindering centralization, in contrast, confines the hierarchy of decision making to formal positions that are often removed from the primary processes of the organization (Alder & Borys, 1996; Sinden, Hoy, & Sweetland, 2004).

Results of this study suggest that teachers are more likely to perceive the teacher evaluation as useful when school principals create environments where decisional authority resides in a professional culture. In such a context, the teacher evaluation would be used as the basis for professional conversations centered on teacher development and improvement. Such an environment establishes a degree of psychological safety that opens teachers up to critical feedback. Receptiveness to information about instructional weaknesses or challenges is a prerequisite for teachers to find the evaluation process meaningful. Such a climate is more likely found in schools with enabling centralization (Adler & Borys, 1996; Hoy, 2002; Hoy & Sweetland, 2001; Sinden, Hoy, & Sweetland, 2004).

Centralization perceived to hinder performance has the opposite effect. In such an environment, teachers are defensive, protective, and fearful that authority will be used in ways that harm their growth and development (Hoy, 2002). Not surprising, teachers are not as open and receptive to information, ideas, or criticisms when the formal authority structures is perceived to constrain performance (Hoy, 2002; Hoy & Sweetland, 2001). This certainly has consequence for the usefulness of teacher evaluation. Negative experiences with the hierarchy of the school would seem to engender beliefs that the evaluation is meant to undermine and control teachers rather than be a tool meant to improve and inform.

Trust in Principal and Teacher Evaluation

It was hypothesized that there would be a positive relationship between teachers who perceived their environment as fostering a culture of trust in the principal and teachers who perceived the evaluation system as useful. The data support the hypothesis that trusting relationships between teachers and the principal influence teacher perceptions of evaluation. Trust accounted for approximately 25 percent of the school level variance in teacher perceptions. An explanation of this finding comes from evidence in classical management theory related to the legitimacy of leaders.

Classical management theory explains that when employees have increased job autonomy, increased job variety, and decreased role ambiguity employees have greater job satisfaction and tend to view their leader as legitimate (Finlay et al., 1995). Within this context, a legitimate leader is someone who is proficient in the skills necessary to successfully fulfill her responsibilities, views the processes and procedures at her disposal as tools to adhere to context-specific situations, and takes risks with open and transparent communication confident that her actions will be met in positive ways (Finlay et al., 1995).

In many ways, a legitimate leader embodies the key tenets of trust, making trust an essential relational condition for effective teacher evaluation. The evidence suggests that in school environments where teachers trust the principal, teachers perceive the evaluation tool as useful. This relationship exists largely because teachers who trust their principal believe the principal has their best interest at heart (Tschannen-Moran, 2014). A lack of trust, on the other hand, raises doubt about principal intent, leading to self-protective mindsets and behaviors that function to buffer teachers from perceived harmful

effects with the evaluation process (Forsyth, Adams, & Hoy, 2011; Hoy & Sweetland, 2001). Without trust, it is hard to envision a scenario where teacher evaluation can be used to improve teaching and learning. This leads to the post hoc evidence on the relationship between teacher trust in district administration and teacher evaluation.

Trust in District Administration and Teacher Evaluation

The purpose of the post hoc was to test the relationship between teacher trust in district administration and teacher perceived usefulness of the evaluation tool. The data strongly support the hypothesis that trusting relationships between teachers and the district administration influence teacher perceptions of evaluation. District trust explained approximately 19 percent of the teacher level variance in the perceived usefulness of teacher evaluation. It is interesting to note with teacher trust in district administration the percent of explained school level variance increased by about 37 percent from the original analysis, suggesting that there may be some grouping effects associated with teacher trust in district administration.

Classical management theory provides the best lens for an explanation of this finding (Organ & Greene, 1981; Pheysey, Payne, & Pugh, 1971; Pritchard & Karasick, 1973; Stevens, Philipsen, & Diedericks, 1992; Sweetland & Hoy, 2001; Zeitz, 1983, 1984). Classical management theory posits that when leadership advances formal policies perceived as legitimate, employees tend to approve and embrace formal structures that can enhance their work (Hall, 1968). This would mirror earlier arguments regarding teacher trust in the principal. From the lens of classical management theory, then, when teachers perceive the district administration as legitimate, teachers would be predisposed to a favorable view of the performance evaluation framework (Finlay et al., 1995).

Characteristics of legitimate leaders align with key tenets of trust. In turn, this would imply that trust in the governing structures, whether those were local as in the case of the principal or more remote as in the district, is an essential relational condition for effective teacher evaluation (Finlay et al., 1995; Tschannen-Moran, 2014).

A plausible explanation for the data showing a stronger relationship between teacher perception of district administration as compared to teacher perception of principal may be that teachers know principals are implementing an evaluation tool they themselves did not select. As such, the degree to which teachers trust where the evaluation came from, in this case, district administration, has a stronger relationship on their perceived usefulness of the evaluation. Again, this harkens back to the idea of perceived legitimacy of leadership structures.

Consideration for future study may be the relationship, if any, between principal trust in district administration and the teacher evaluation tool. The degree to which the principal perceives district administration as legitimate may have bearing on how the principal implements the evaluation tool in their building. It would certainly appear reasonable that the extent to which a principal perceives the district leadership structures as legitimate would have some kind of effect on either the perception or implementation of the teacher evaluation tool, or at least on the teacher perception of the district.

Teacher Characteristics and Teacher Evaluation

The analysis also examined the relationship between individual teacher characteristics and perceptions of the evaluation process. The primary teacher characteristics included whether or not teachers had achieved National Board Certification (NBC), number of years in the school, and gender. Even though the effects

of teacher characteristics were either non-existent or small, it is worth commenting on these findings.

The research showed that NBC teachers are no more likely to perceive the evaluation tool as useful than teachers who are not nationally certified. NBC is a highly rigorous process that takes tremendous time and energy to complete. It could be suggested that teachers willing to go through this ambitious process are highly effective teachers because the same characteristics necessary to be a highly effective teacher are necessary for completing the rigorous National Board Certification process. If we assume that teachers obtain this certification for intrinsic motivations, it would seem natural to posit that teachers who obtain their National Board Certification would be interested in continued self-improvement, professional development, and growth. However regardless of NBC, teachers are no more likely to perceive the tools within the school system as useful.

An explanation for the lack of a statistically and practically significant relationship comes from evidence supporting the hypotheses. If perception of the usefulness of the evaluation tool is truly a product of environment, individual teacher characteristics such as gender, age, and National Board Certification should not matter to the process or to teacher perception. While it would seem logical to assume that teachers who have gone through rigorous evaluation processes would be better able to determine the effectiveness of future processes, it appears that this factor had no meaningful or significant outcome in this case.

The evidence did show a statistically significant and negative relationship between number of years in the school and perceived usefulness of teacher evaluation.

This suggests that being in the same school for a longer period of time was associated with negative or ambivalent perceptions of teacher evaluation. It may be that teachers who have been in the teaching profession for many years become accustom to routine processes and procedures. They are familiar with how the processes and procedures are implemented and the consequences on them individually. This familiarity eliminates ambiguity and uncertainty. Constant change in schools, with little to no meaningful results, naturally creates cynicism and apathy toward anything new. This would apply as much to teacher evaluation as it would to a new reading intervention.

It also may be the case that the significant and negative relationship between years in current school and TLE's usefulness is the result of longer tenured teachers at a school experiencing the TLE as ineffective in identifying low- and high-performing teachers. If a tenured teacher perceives another teacher as ineffective but they are still employed at the school, tenured teachers would begin to view the tool as ineffective in weeding out low-performing employees. The converse could also be true. Repeated instances of discrepancies between teacher perception of a colleague's performance and the evaluation score would result in an employee perceiving the tool as ineffective because their experiences do not align with the tool.

Another plausible reason for the small negative relationship is that tenured teachers at a school may be more self-actualized so the need for external tools to evaluate performance is viewed as unnecessary, especially when the evaluation does not align with teacher perception. These teachers have spent more years immersed in the school landscape. Even if these teachers do not seek out opportunities, the requirements of their job expose them to experiences where they begin to determine their effectiveness. Either

through comparison between and among colleagues, professional development, conferences, or annual evaluations, tenured teachers may be more aware of their professional strengths and areas of improvement. If the evaluation tool is not nuanced enough to identify professional strengths or shows a perceived strength as an area of improvement, a tenured teacher could view the tool as ineffective. Disconnect between teacher perception of him/herself and the evaluation tool could result in the negative relationship between the teacher and TLE usefulness.

The small positive relationship between female teachers and perceived usefulness of the evaluation system is interesting and difficult to make sense of because the primary focus in this study was on school structure and not gender differences. One possibility that cannot be ruled out is the relatively small number of male teachers in the sample. Approximately 85 percent of teachers identified as female, meaning that the significance of a man's negative perception of the evaluation system would be muted by the larger sample size.

Implications for Leadership Practice

Although school systems across the country have made significant investments in new performance-based evaluation tools and processes, considerable evidence shows continued problems with how these new systems function (Hallinger, Heck, & Murphy, 2014). Reasons for the inability of teacher evaluation to accomplish broad-based goals vary from problems inherent with the observational protocols, flaws with the metrics, misguided performance assumptions, and implementation challenges (Polikoff & Porter, 2014; Whitehurst, Chingos, & Lindquist, 2014). This study was conceived in order to

understand if and how organizational structures in schools can contribute to more useful teacher evaluation.

Evidence supports the hypotheses that formal structures like formalization and centralization, and informal conditions, like trust in the principal, are associated with teacher perceptions of the evaluation process. Given evidence in bureaucracy and classical management theory, these findings make sense and can be explained by the importance of organizing teaching and learning in ways that balance structure with professional discretion. Next, two implications for how evidence in this study relates to administrative practice are advanced.

First, schools should use resources to increase enabling formalization, enabling centralization, and faculty trust in the principal. Existing research explains how these three conditions affect other aspects of school performance by facilitating collective problem solving, fostering commitment to common goals, and promoting instructional creativity (Hoy & Sweetland, 2001; Tschannen-Moran, 2009). This study adds to the body of evidence by linking enabling structures to useful teacher evaluation. While the application of formalization and centralization may not be a panacea for teacher evaluation problems, they provide a starting point that can be achieved with little to no financial resources.

Second, teacher evaluation is a required structural feature of schools. It can be used in effective or ineffective ways. And as demonstrated in this study a favorable or unfavorable experience can depend in part on the larger formal environment of the school. In an enabling environment, where teachers perceive rules and procedures as adaptable to their unique needs, evaluation can be experienced as supporting teacher

growth (Adler & Borys, 1996; Hoy & Sweetland, 2001). Teachers need and can benefit from formal structures, but problems tend to arise when formal structures constrain the ability of educations to adapt and grow.

These implications for practice have potentially significant ramifications in Oklahoma where this research was conducted. This research shows that there are cost effective ways to invest in the growth and development of our teachers that could have implications for improved student learning, primarily by ensuring that new tools and approaches are not dismissed but instead are used to help bolster teacher outcomes. Additionally, the investment in the performance and professional growth of our state's teachers could have far reaching implications for their satisfaction and morale.

Conclusion

A significant takeaway from this research is that organizational structure has implications for current practice and future research. The data presented within this study support the theoretical connections within the literature in that conditions of enabling centralization, enabling formalization, and trust are necessary for teachers to perceive the evaluation tool as useful.

If elements of structure and trust can ultimately affect the perception of the success or failure of an evaluation system, it would behoove educators, policy makers, and districts to be more mindful in their approach to teacher and leader effectiveness systems in the future. A great deal of time, money, and human capital is invested in these types of programs, but it may well be that these investments alone are not enough to ensure the perception of success within a system. From this perspective, the implementation of evaluation systems and the cost associated with them becomes a

matter of value. This is not to say that structure should be the end goal instead if there is a strong evaluation system, backed by clear research and practice, those that implement it should be especially mindful of the effects of the systems into which these evaluation systems will be introduced.

While this research stops short of a causal relationship between perception of the tool and ultimate change in teacher performance as a result of those perceptions, it would be logical that ongoing discussions about the perceived usefulness of evaluation systems, as opposed to incremental progress to make them better, is ultimately not in the best interest of teachers or administrators. Furthermore, it creates policy discussions that are then not centered on improving schools and instead on creating tools. While organizational structure alone may be an incomplete mechanism to ameliorate performance problems in urban schools, coupled with strong tools, the structure may create conditions in which teachers and administrators are more open to the idea of growth and improvement. At the very least, it appears they will be more open to the idea that these tools can produce positive results.

Future research could examine whether or not perception affects the measured usefulness of the tool: Do positive perceptions of the tool correlate to better student outcomes? Are principals whose staff perceives the tool as useful more likely to be happy with their placement? Are there connections between perception of the tool's usefulness and other measures of organizational satisfaction? This research does not make the leap between teacher perception of the usefulness of the evaluation tool and effectiveness. Further research could analyze this, including the relationship between

teacher level characteristics and perceived usefulness of both school structure and perceived usefulness of the evaluation tool.

While the perception of the tool has plenty of implications, the district should only focus on structure once it has a tool that it knows will work. Put another way, people may disengage from effective tools in the wrong structure, but the right structure does not mean that the wrong tool gets better; it is the perception of the tool's usefulness that improves. This research looked solely at the relationship between perception of structure and perception of one tool – the teacher evaluation tool. It is logical to conclude that the perception of structure has consequences for how all tools and processes within the structure are perceived. In this way, perception of structure has implications beyond this research that extend to how all rules, policies, and practices are perceived and their usefulness.

References

Adler, P. S., & Borys, B. (1996). Two types of bureaucracy: Enabling and coercive. *Administrative Science Quarterly*, 61-89.

Borman, G. D., & Kimball, S. M. (2005). Teacher quality and educational equality: Do teachers with higher standards based evaluation ratings close student achievement gaps?. *The Elementary School Journal*, *106*(1), 3-20.

Bush, G. W. (2001). No Child Left Behind.

Cavalluzzo, L. C. (2004). Is National Board Certification an Effective Signal of Teacher Quality?. *CNA Corporation*.

Coggshall, J., Max, J., & Bassett, K. (2008). Key Issue: Using Performance-Based Assessment to Identify and Support High-Quality Teachers. *National Comprehensive Center for Teacher Quality*.

Cunningham, W.G. Gresso, D.W. (1993), *Cultural leadership*. Boston: Allyn Bacon.

Daley, G., & Kim, L. (2010). A Teacher Evaluation System That Works. Working Paper. *National Institute for Excellence in Teaching*.

Dalton, D. R., Todor, W. D., Spendolini, M. J., Fielding, G. J., & Porter, L. W. (1980). Organization structure and performance: A critical review. *Academy of Management Review*, 5(1), 49-64.

Danielson, C. (2007). *Enhancing professional practice: A framework for teaching*. ASCD.

Darling-Hammond, L. (1996). What matters most: A competent teacher for every child. *Phi Delta Kappan*, *78*(3), 193.

Darling-Hammond, L. (1997). *Doing what matters most: Investing in quality teaching*. National Commission on Teaching & America's Future, Kutztown Distribution Center, 15076 Kutztown Road, PO Box 326, Kutztown, PA 19530-0326.

Darling-Hammond, L. (2000). Teacher quality and student achievement. *Education policy analysis archives*, *8*, 1.

Darling-Hammond, L. (2003). Keeping Good Teachers: Why It Matters, What Leaders Can Do. *Educational leadership*, *60*(8), 6-13.

Darling-Hammond, Linda, Arthur E. Wise, and Sara R. Pease. "Teacher evaluation in the organizational context: A review of the literature." *Review of educational research* 53.3 (1983): 285-328.

Dawes, John (2008). Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *International Journal of Market Research* 50 (1): 61–77.

Diez Roux, A.V. (2002). A glossary for multilevel analysis. *Journal of Epidemiology and Community Health, 56*(8), 588-594.

Finlay, W., Martin, J. K., Roman, P. M., & Blum, T. C. (1995). Organizational structure and job satisfaction do bureaucratic organizations produce more satisfied employees?. *Administration & Society,* 27(3), 427-450.

Ford, R. N. (1973). Job enrichment lessons from AT&T. *Harvard Business Review*, 51(1), 96-106.

Ford, R., & Johnson, C. (1998). The perception of power: Dependence and legitimacy in conflict. *Social Psychology Quarterly, 61,* 16-32.

Forsyth, P. B., & Adams, C. M. (2014). The school principal and organizational predictability. In D. Van Maele, P. Forsyth, and M. Van Houtte (Eds.), *Trust relationships and school life*, Springer.

Forsyth, P.B., Adams, C.M., & Hoy, W.K. (2011). *Collective trust: Why schools can't improve without it.* New York: Teachers College.

Forsyth, P. B., Barnes, L. L., & Adams, C. M. (2006). Trust-effectiveness patterns in schools. *Journal of Educational Administration,* 44(2), 122-141.

Fullan, M. (2011). Whole system reform for innovative teaching and learning. *Microsoft-ITL Research (Ed.), Innovative Teaching and Learning Research*, 30-39.

Fullan, M. (2011a). Choosing the wrong drivers for whole system reform. Seminar Series 204. Melbourne, AU: Centre for Strategic Education.

Fullan, M. (2011, October). Choosing the wrong drivers. In *Center for Strategic Education, Seminar for Studies in Education, University of Toronto, Toronto, Canada. Series Paper* (No. 204, p. 5).

Gates Foundation. (2013). Measures of effective teaching (MET). Downloaded January 14, 2013 from http://www.gatesfoundation.org/united-states/Pages/measures-of-effective-teaching-fact-sheet.aspx.

Gardner, D. P. (1983). A Nation At Risk: The Imperative For Educational Reform. An Open Letter to the American People. *A Report to the Nation and the Secretary of Education*.

Goe, L., & Croft, A. (2009). Methods of evaluating teacher effectiveness. *Washington, DC: National Comprehensive Center for Teacher Quality*.

Goldhaber, D., & Anthony, E. (2007). Can teacher quality be effectively assessed? National board certification as a signal of effective teaching. *The Review of Economics and Statistics*, *89*(1), 134-150.

Gouldner, A.W. (1954). *Patterns of industrial bureaucracy***.** Glencoe, IL: Free Press.

Goddard, R.D., Tschannen-Moran, M., & Hoy, W.K. (2001). A multilevel examination of the distribution and effects of teacher trust in students and parents in urban elementary schools. *The Elementary School Journal, 102*(1), 3-17.

Hackman, J. R. (1980). Work redesign and motivation. *Professional Psychology*, *11*(3), 445.

Hackman, J. R., Oldham, G., Janson, R., & Purdy, K. (1975). A new strategy for job enrichment. *California Management Review*, 17(4), 57-71.

Haertel, E. H. (2013). Reliability and validity of inferences about teachers based on student test scores.

Hage, J. (1965). An axiomatic theory of organizations. *Administrative Science Quarterly*, 10, 289-320.

Hall, R. H. (1968). Professionalization and bureaucratization. *American Sociological Review*, 92-104.

Hallinger, P., Heck, R. H., & Murphy, J. (2014). Teacher evaluation and school improvement: An analysis of the evidence. *Educational Assessment, Evaluation and Accountability*, *26*(1), 5-28.

Hoy, W.K. (2002), "Faculty trust: a key to student achievement," Journal of School Public Relations, Vol. 23 No. 2, pp. 88-103.

Hoy, W. K., & Sweetland, S. R. (2001). Designing better schools: The meaning and measure of enabling school structures. *Educational Administration Quarterly,* 37(3), 296-321.

Hoy, W. K., Tarter, C. J., & Witkoskie, L. (1992). Faculty trust in colleagues: Linking the principal with school effectiveness. *Journal of Research and Development in Education, 26*(1), 38-45.

Hoy, W.K., & Tschannen-Moran, M. (1999), Five faces of trust: an empirical confirmation in urban elementary schools, *Journal of School Leadership*, Vol. 9 (3), 184-208.

Jacob, B. A., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, *26*(1), 101-136.

Johnson, C., & Ford, R. (1996). Dependence, power, legitimacy, and tactical choice. *Social Psychology Quarterly, 59,* 129-139.

Kane, T. J., & Staiger, D. O. (2008). *Estimating teacher impacts on student achievement: An experimental evaluation* (No. 14607). National Bureau of Economic Research.

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*, *27*(6), 615-631.

Kimball, S. M., White, B., Milanowski, A. T., & Borman, G. (2004). Examining the relationship between teacher evaluation and student assessment results in Washoe County. *Peabody Journal of Education*, *79*(4), 54-78.

Kohn, M. L. (1971). Bureaucratic man: A portrait and an interpretation. *American Sociological Review*, 461-474.

Koretz, D. (2008). A measured approach. *American Educator*, *32*(2), 18-39.
Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. P*eabody Journal of Education*, *79*(4), 33-53.

Lustick, D., & Sykes, G. (2006). National board certification as professional development: What are teachers learning?.

Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. *peabody Journal of Education*, *79*(4), 33-53.

Milanowski, A. T., Kimball, S. M., & Odden, A. (2005). Teacher accountability measures and links to learning. *Measuring school performance and efficiency: Implications for practice and research*, 137-161.

Murphy, R. (2013). Testing teachers: what works best for teacher evaluation and appraisal.

Odden, A., & Wallace, M. (2008). How to achieve world class teacher compensation. *Indianapolis: Freeload. Popham, W.(1988). The dysfunctional marriage of formative and summative teacher evaluation. Journal of.*

Organ, D. W., & Greene, C. N. (1981). The effects of formalization on professional involvement: A compensatory process. *Administrative Science Quarterly*, 26, 237-252.

Pheysey, D. C., Payne, R., & Pugh, D. A. (1971). The influence of structure at organizational and group levels. *Administrative Science Quarterly,* 156, 61-73.

Pierce, J. L., & Dunham, R. B. (1976). Task design: A literature review. *Academy of Management Review,* 1 (4), 83-97.

Polikoff, M. S., & Porter, A. C. (2014). Instructional alignment as a measure of teaching quality. *Educational Evaluation and Policy Analysis*.

Prichard, R. D., & Karasick, B. W. (1973). The effects of organizational climate on managerial job performance and job satisfaction. *Organizational Behavior and Human Performance,* 9, 126-146.

Public School revenue Sources. (2015, May). Retrieved November 16, 2015, from http:nces.ed.gov/.

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchial linear models* (2nd ed.). Thousand Oaks, CA: Sage.

Reynolds, C. R., Livingston, R. B., Willson, V. L., & Willson, V. (2010). *Measurement and Assessment in Education*. Pearson Education International.

Ridgeway, C. (1989). Understanding legitimation in informal status orders. In J. Berger, M. Zelditch, Jr., & B. Anderson (Eds.), *Sociological theories in progress: New formulations* (pp. 131-159). London: Sage Publications.

Sanders, W. L., & Horn, S. P. (1998). Research findings from the Tennessee Value-Added Assessment System (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education*, *12*(3), 247-256.

Sanders, W. L., Wright, S. P., & Horn, S. P. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, *11*(1), 57-67.

Sato, M., Wei, R. C., & Darling-Hammond, L. (2008). Improving teachers' assessment practices through professional development: The case of National Board Certification. *American Educational Research Journal*, *45*(3), 669-700.

Shepard, J. M. (1970). Functional specialization, alienation, and job satisfaction. *Industrial and Labor Relations Review*, 23, 207-219.

Sims, H. P., Szilagyi, A. D., & Keller, R. T. (1976). The measurement of job characteristics. *Academy of Management Journal*, *19*(2), 195-212.

Sinden, J. E., Hoy, W. K., & Sweetland, S. R. (2004). An analysis of enabling school structure: Theoretical, empirical, and research considerations. *Journal of Educational Administration,* 42(4), 462-478.

Stevens, F., Philipsen, H., & Diedericks, J. (1992). Organizational and professional predictors of physician satisfaction. *Organization Studies,* 13, 35-49.

Stone, E. F., & Porter, L. W. (1975). Job characteristics and job attitudes: A multivariate study. *Journal of Applied Psychology,* 60(1), 57.

Sweetland, S. R., & Hoy, W. K. (2000). School characteristics and educational outcomes: Toward an organizational model of student achievement in middle schools. *Educational Administration Quarterly*, *36*(5), 703-729.

Tarter, C. J., Sabo, D., & Hoy, W. K. (1995). Middle school climate, faculty trust and effectiveness: A path analysis. *Journal of Research and Development in Education, 29*(1), 41-49.

Toch, T., & Rothman, R. (2008). Rush to Judgment: Teacher Evaluation in Public Education. Education Sector Reports. *Education Sector*.

Thompson Victor, A. (1961). Modern Organization.

Tulsa Public Schools (2012). *2014-2015 Master Contract Tulsa Public Schools and Tulsa Classroom Teachers Association*. Retrieved from http://www.tulsaschools.org/1_Administration/03_STAFF_MEMBERS/_DEPARTMENTS/human_capital/_documents/pdf/MasterContract.pdf.

Tulsa Public Schools (October 2014). *Frequently asked questions about value-added analysis.* Retrieved from file:///Users/elizabethvereecke/Downloads/-documents-tulsa-FrequentlyAskedQuestions-Updated_10-22-14%20(1).pdf.

Tulsa Public Schools (March 2016). *TPS seeks community input on funding priorities. http://www.tulsaschools.org/2_News/01_PUBLIC_INFO/news_item.asp?ID=15429.*

Tschannen-Moran, M. (2004, November). *What's trust got to do with it? The role of faculty and principal trust in fostering student achievement.* Paper presented at the meeting of the University Council for Educational Administration, Kansas City, MO.

Tschannen-Moran, M. and Hoy, W.K. (1998), "Trust in schools: a conceptual analysis," Journal of Educational Administration, Vol. 36, pp. 334-5.

US Department of Education (2010a) *A Blueprint for Reform*, US Department of Education, Washington, DC.

Vogt, W. P. (2007). *Quantitative research methods for professionals*. Allyn & Bacon.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, *63*(3), 249-294.

Weber, M. (1978). *Economy and society: An outline of interpretive sociology*. University of California Press.

Webster, W. J., & Mendro, R. L. (1997). *The Dallas value-added accountability system: Grading teachers, grading schools: Is student achievement a valid evaluation measure*, Sage Publications Company: Thousand Oaks, CA.

Weed, E. D. (1971). Job enrichment clean up at Texas Instruments. *New Perspectives in Job Enrichment.* New York.

White, B. (2004). The relationship between teacher evaluation scores and student achievement: evidence from Coventry, RI. *CPRE-UW Working Paper Series*. Madison, WI: University of Wisconsin-Madison, Wisconsin Center for Education Research, Consortium for Policy Research in Education, San Diego, CA.

Whitehurst, G. J. R., Chingos, M. M., & Lindquist, K. M. (2014). Evaluating teachers with classroom observations. *Brown Center on Education Policy: Brookings Institute*.

Wise, A. E., Darling-Hammond, L., McLaughlin, M. W., & Bernstein, H. T. (1985). Teacher evaluation: A study of effective practices. *The Elementary School Journal*, 61-121.

Wolf, K. (1997). Portfolios in teacher evaluation. ERIC, March 1997.
Woltman, H., Feldstain, A., MacKay, J. C., & Rocchi, M. (2012). An introduction to hierarchical linear modeling. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 52-69.

Zeitz, G. (1983). Structural and individual determinants of organization morale and satisfaction. *Social Forces,* 61, 1008-1108.

Zeitz, G. (1984). Bureaucratic role characteristics and member affective response in organizations. *Sociological Quarterly*, 25, 301-318.

Zelditch Jr, M., & Walker, H. A. (1984). Legitimacy and the Stability of Authority. *Advances in group processes*, *1*, 1-25.

Appendix A

Exploratory Factor Analysis

Faculty Trust in Principal

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Factor	Total	% of Variance	Cumulative %	Total	% Variance	Cumulative %
1	6.269	78.359	78.359	6.039	75.482	75.482
2	.455	5.687	84.046			
3	.410	5.119	89.165			
4	.291	3.636	92.801			
5	.212	2.645	95.447			
6	.168	2.095	97.541			
7	.146	1.820	99.361			
8	.051	.639	100.000			

Factor Analysis

	Factor 1
FTPrin1	.801
FTPrin2	.930
FTPrin3	.949
FTPrin4	.891

FTPrin5	.837
FTPrin6	.902
FTPrin7	.885
FTPrin8	.734

Formalization

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Factor	Total	% of Variance	Cumulative %	Total	% Variance	Cumulative %
1	3.569	59.490	59.490	3.115	51.909	51.909
2	.725	12.083	71.573			
3	.593	9.887	81.460			
4	.447	7.448	88.908			
5	.366	6.108	95.016			
6	.299	4.984	100.000			

Factor Analysis

	Factor 1
ESS1	.514
ESS2	.760
ESS5	.685

ESS7	.790
ESS9	.782
ESS10	.754

Centralization

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Factor	Total	% of Variance	Cumulative %	Total	% Variance	Cumulative %
1	4.116	68.594	68.594	3.745	62.418	62.418
2	.542	9.027	77.621			
3	.446	7.436	85.057			
4	.343	5.710	90.767			
5	.301	5.016	95.783			
6	.253	4.217	100.000			

Factor Analysis

	Factor 1
ESS3	.844
ESS4	.767
ESS6	.840
ESS8	.773
ESS11	.777
ESS12	.734

Teacher Evaluation System

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Factor	Total	% of Variance	Cumulative %	Total	% Variance	Cumulative %
1	4.291	61.294	61.294	3.889	55.560	55.560
2	.784	11.200	72.494			
3	.662	9.451	81.945			
4	.430	6.137	88.082			
5	.321	4.589	92.671			
6	.288	4.111	96.782			
7	.225	3.218	100.000			

Factor Analysis

	Factor 1
TLE1	.799
TLE2	.532
TLE4	.842
TLE5	.826
TLE6	.823
TLE7	.754
TLE9	.575

Appendix B

Classroom Observation and Evaluation Instrument

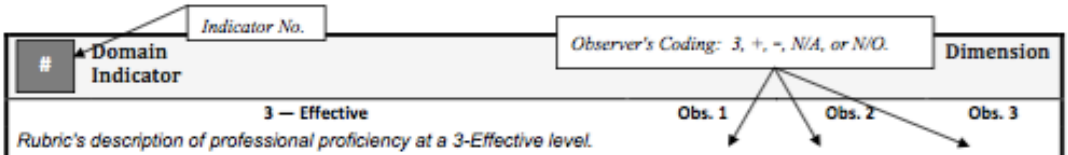
TULSA
PUBLIC SCHOOLS

TLE Observation Form
Teachers
2014-2015

Educator Name:_____ School Name_____

Evaluator Name:_____

	Obs. 1	Obs. 2	Obs. 3
Date			
Obs. Conf. Date			
Educator's Initials			
Observer's Initials			



3 — Effective	Obs. 1	Obs. 2	Obs. 3
Rubric's description of professional proficiency at a 3-Effective level.			

Note: The observation rating should reflect the evaluator's intentional study and analysis of the teacher's classroom performance and other factors that quantify the impact of the educator—up to, and including, the date of the classroom observation.

Comments:

Insert comments, dates, observation notes, evidence collected to date, etc.

***

1 Domain: **Classroom Management** Dimension: **Preparation**

Teacher plans for and executes a lesson relating to short-term and long-term objectives.

3 — Effective	Obs. 1	Obs. 2	Obs. 3
Plans for and executes instructional strategies that encourage the development of performance skills relating to short and long term objectives. Develops instructional plans that are in alignment with most current state standards and, as available and appropriate, curriculum maps and pacing guides. Plans consistently address student diversity and describe how instruction will be differentiated. Ensures materials and equipment are ready at the start of the lesson or instructional activity (most of the time).			

Comments:

2 Domain: **Classroom Management** Dimension: **Discipline**

Teacher clearly defines and effectively manages expected behavior.

3 — Effective	Obs. 1	Obs. 2	Obs. 3
Establishes, communicates, and consistently implements appropriate standards of conduct. Students are usually engaged and clear as to the expectations of the classroom, requiring few reminders relative to the age and development of the students. Monitors the behavior of students during whole-class, small group and seat work activities and during transitions between instructional			

activities.

As necessary and appropriate, stops misbehavior promptly and consistently, with a voice level / word choice suitable to the situation.

Comments:

3	Domain: **Classroom Management**	Dimension: **Building-Wide Climate Responsibilities**

Teacher assures a contribution to building-wide positive climate responsibilities.

3 — Effective	Obs. 1	Obs. 2	Obs. 3
Regularly and routinely participates in school projects and initiatives that contribute to promoting orderly behavior throughout the school.			
Follows the procedures, practices and guidelines outlined by the school, district, state and federal laws intended to keep students healthy and safe.			

Comments:

4	Domain: **Classroom Management**	Dimension: **Lesson Plans**

Teacher develops daily lesson plans designed to achieve the identified objectives.

3 — Effective	Obs. 1	Obs. 2	Obs. 3
Plans are developed consistently and on time based upon an analysis of data.			
Plans with other members of the grade-level / school planning teams (when it is an expectation of the campus).			
Provides substitute plans, classroom rosters, seating charts, behavior plans, emergency plans and identification of diverse learning groups.			

Comments:

5	Domain: **Classroom Management**	Dimension: **Assessment Practices**

Teacher acknowledges student progress and uses assessment practices that are fair, based on identified criteria, and support effective instruction.

3 — Effective	Obs. 1	Obs. 2	Obs. 3
Consistently uses assessments to evaluate student learning and guide instruction. Grading is fair and in accordance with district's grading policies. Provides adequate and timely feedback from assessment results for students to reflect and set goals. Recognizes student progress and achievement at significant intervals and encourages learning behaviors that would result in student success.			

Comments:

6	Domain: **Classroom Management**	Dimension: **Student Relations**

Teacher optimizes the learning environment through respectful and appropriate interactions with students, conveying high expectations for students and an enthusiasm for the curriculum.

3 — Effective	Obs. 1	Obs. 2	Obs. 3
Oral, written and nonverbal communications with students are considerate and respectful. Consistently conveys a generally positive view of learning and of the curriculum, demonstrating high academic expectations for most students.			

Comments:

7	Domain: **Instructional Effectiveness**	Dimension: **Literacy**

Teacher embeds the components of literacy into all instructional content.

3 — Effective	Obs. 1	Obs. 2	Obs. 3
Literacy (the practice of reading, writing, developing vocabulary, spelling, or listening/ speaking) is embedded in the lesson as a vehicle for learning the content and for demonstrating understanding. As appropriate for the content area, instruction is provided through text.			

Comments:

8	Domain: **Instructional Effectiveness**	Dimension: **Current State Standards**
Teacher understands and optimizes the delivery focus of current state standards and the expectations derived from same on student learning and achievement.		

3 — Effective	Obs. 1	Obs. 2	Obs. 3
Understands the current state standards as evidenced by use of alternate instructional strategies and modified content focus aligned with current state standards.			

Comments:

9	Domain: **Instructional Effectiveness**	Dimension: **Involves All Learners**
Teacher uses active learning, questioning techniques and/or guided practices to involve all students.		

3 — Effective	Obs. 1	Obs. 2	Obs. 3
Routinely uses strategies to ensure engagement of all students. Engages most students in active learning experiences 80 percent of the class time. Uses questioning techniques throughout the lesson, scaffolding to at least the mid-level of Bloom's taxonomy. Provides adequate wait time for student response and engagement. Engages students by incorporating their general skills and interests into the lesson.			

Comments:

10	Domain: **Instructional Effectiveness**	Dimension: **Explains Content**
	Teacher teaches the objectives through a variety of methods.	

3 — Effective	Obs. 1	Obs. 2	Obs. 3
Uses a variety of activities (e.g. modeling, visuals, hands-on activities, demonstrations, gestures, body language and thematic instruction) to support the instructional outcomes and meet varied student needs/ learning styles / multiple intelligences. Technology is included in the planning process to support instruction, and technology is used on a regular basis as an instructional tool.			

Comments:

11	Domain: **Instructional Effectiveness**	Dimension: **Clear Instruction & Directions**
	Teacher provides clear instruction and direction.	

3 — Effective	Obs. 1	Obs. 2	Obs. 3
Provides instruction, directions and procedures in a variety of delivery modes, e.g., verbal, modeling, visual, demonstration, etc., that are accurate, clearly stated / presented and relate to the learning objectives. Gives students directions for transitions and includes transitioning in the planning process to optimize academic learning time. Uses spoken and written language that is clear and correct, conforms to standard English, vocabulary, and is appropriate to students' ages and interests.			

Comments:

12	Domain: **Instructional Effectiveness**	Dimension: **Models**
	Teacher demonstrates / models the desired skill or process.	

3 — Effective	Obs. 1	Obs. 2	Obs. 3
Provides demonstrations and modeling of the desired skill or process that are clear and precise to students.			

Comments:

13	Domain: **Instructional Effectiveness**	Dimension: **Monitors**
	Teacher checks to determine if students are progressing toward stated objectives.	

3 — Effective	Obs. 1	Obs. 2	Obs. 3
When appropriate, moves to all areas of the room while students are working on guided practice to promote and reinforce students' progress toward the stated objectives. Uses different types of student response techniques, both individual / group. Uses student response techniques to check for understanding. Uses feedback from students regarding their understanding. Uses wait time of 3-5 seconds (more for more complex questions) after voicing the question for the purpose of monitoring student understanding. Provides opportunity for students to formulate more thoughtful responses and allows time for the student to consider supporting evidence.			

Comments:

14	Domain: **Instructional Effectiveness**	Dimension: **Adjusts Based Upon Monitoring**
	Teacher changes instruction based on the results of monitoring.	

3 — Effective	Obs. 1	Obs. 2	Obs. 3
Consistently monitors student involvement and makes efforts to adjust instructional plans to engage more students. Assesses mastery of the new learning to determine if independent practice or re-teaching is appropriate and makes adjustments to lessons. Reviews data from assessments to modify instruction and guide intervention strategies.			
Comments:			

15	Domain: **Instructional Effectiveness**	Dimension: **Establishes Closure**
	Teacher summarizes and fits into context what has been taught.	

3 — Effective	Obs. 1	Obs. 2	Obs. 3
Uses one or more closure strategies (e.g., summarizing, discussing main ideas or connections) to consolidate and solidify student learning and help students organize the information into a meaningful context. Connects what is learned to prior learning.			
Comments:			

16	Domain: **Instructional Effectiveness**	Dimension: **Student Achievement**		
	Effective development and use of modified assessments and curriculum for special education students and other students experiencing difficulties in learning.			
	3 — Effective	**Obs. 1**	**Obs. 2**	**Obs. 3**
	Accepts responsibility for the success of all students Modifies assessments for special education student populations in alignment with the IEPs and for other students experiencing difficulties in learning as appropriate. Provides required feedback to student, roster teacher and/or parent. Assures that all students have access to current state standards/ district curriculum.			
	Comments:			

17	Domain: **Professional Growth and Continuous Improvement**	Dimension: **Professional Learning**		
	Uses Professional Growth as a Continuous Improvement Strategy			
	3 — Effective	**Obs. 1**	**Obs. 2**	**Obs. 3**
	Participates in the required minimum hours of professional development updating their content knowledge and current professional practices.			
	Comments:			

18	Domain: **Professional Growth and Continuous Improvement**	Dimension: **Professional Accountability**		
	Exhibits behaviors and efficiencies associated with professionalism.			
	3 — Effective	**Obs. 1**	**Obs. 2**	**Obs. 3**
	Exhibits consistent reliability-based behavior patterns as evidenced by punctuality and dependability; adhering to prescribed arrival and departure times; following notification and reporting procedures for absences; complying with reporting timelines and other time sensitive info./compliance requests.			
	Comments:			

19	Domain: **Interpersonal Skills**	Dimension: **Effective Interpersonal Skills**		
	Effective Interactions and Collaboration with Stakeholders.			
	3 — Effective	**Obs. 1**	**Obs. 2**	**Obs. 3**
	Interacts with families and colleagues in a timely, consistent, positive and professional manner. Complies with school procedures for communicating with families and colleagues and makes an effort to engage them in the educational program. Collaborates appropriately and makes decisions that reflect genuine professional consideration.			
	Comments:			

20	Domain: **Leadership**	Dimension: **Professional Involvement & Leadership**		
	Exhibits Positive Leadership through Varied Involvements.			
	3 — Effective	**Obs. 1**	**Obs. 2**	**Obs. 3**
	Agrees to participate in school or district events when asked. Finds ways to contribute to the profession and follows through. Assumes a proactive role in addressing student needs.			
	Comments:			

Appendix C

2014-2015 Master Contract Tulsa Public Schools and Tulsa Classroom Teachers Association Teacher: Teacher and Leadership Effectiveness evaluation tool

When scheduling a conference for the purpose of issuing an admonishment, personal development plan, or any discipline document to be placed in the teacher's personnel file, the Administrator shall:

A. Inform the individual of the specific nature of the subject to be addressed.

B. Allow for a reasonable amount of time for the individual to secure representation, if desired.

Multiple Measure System of Evaluation

A teacher's overall evaluation for the school year shall be comprised of multiple measures of professional effectiveness—qualitative and quantitative measures. The multiple measure system of teacher evaluation, designed in collaboration with Tulsa Classroom Teachers Association, is described in Appendix G. At the conclusion of the school year, the District shall provide each teacher with access to a single report that will communicate that teacher's scores on all available qualitative and quantitative measures. The qualitative component of teacher evaluation (the observation-based teacher evaluation conducted pursuant to the Tulsa Model) is described below. TPS staff will develop a policy for board approval that incorporates the components of the multi-measure evaluation system described herein and in Appendix G. This policy shall allow TPS and TCTA flexibility to review and improve the multi-measure system as necessary, especially if there is a change in the state law governing the use and calculation of quantitative measures.

The Qualitative Teacher Evaluation Process

1. Statement of Purpose

The purpose of teacher evaluation in the District is to improve the quality of instruction. In seeking this goal, two primary objectives are acknowledged. First, the observation and evaluation of teacher performance is intended to identify the teacher's strengths and weaknesses, to agree upon strategies for reinforcing strengths and remediating weaknesses, and to follow through on the steps designed to improve the teacher's performance. The second objective of evaluation is to provide a rational basis for administrative decisions regarding continued employment.

2. Frequency of Qualitative Evaluation

Tulsa teachers will be observed and evaluated in compliance with the timelines negotiated in the TLE Observation & Evaluation Handbook for Evaluators using the Tulsa Model, (See Section 4. The TLE Observation and Evaluation Process and Timeline.)

A teacher may request a performance evaluation at any time.

3. Method of Qualitative Evaluation

The evaluator is to complete the CONFIDENTIAL EVALUATION by rating the teacher's performance on the agreed upon Evaluation Forms attached to this document as Schedule III). Each Ineffective and/or Needs Improvement rating requires that the evaluator complete a Personal Development Plan, specifying what the teacher should do to overcome that

particular inadequacy, unless a Goal Setting Form is appropriate and used by the evaluator in accordance with the TLE Observation & Evaluation Handbook for the Evaluators Using the Tulsa Model.

The Personal Development Plan form is intended primarily to supplement the CONFIDENTIAL EVALUATION, as described above. However, if the evaluator observes a single area of ineffective or needs improvement performance that needs immediate correction, the Personal Development Plan may be used to cite the deficiency and give instructions for correcting the problem. Administrators are encouraged, as the performance issues may dictate, to meet informally with a teacher before issuing a Personal Development Plan.

4. Personnel Affected

This provision shall apply to all regular employees defined by law as teachers.

Teachers on temporary contracts shall be subject to evaluation, but shall not be afforded job rights unless otherwise specified by state law.

5. Procedure for Qualitative Evaluation

All observations of the classroom teaching performance of any teacher shall be conducted openly and with the full knowledge of the teacher and all evaluations shall be made by a qualified and certified administrator. Teachers will be provided with a minimum two week period of time during which their observation will be conducted. Notification must be provided no later than the end of the teacher work day, if an observation will be conducted the following day.

No evaluation of any teacher's classroom performance shall be completed and filed unless and until the evaluating administrator shall have observed the teacher at work at least two (2) times, or three (3) times if the teacher shall promptly after the second observation request that he be observed again. No teacher shall receive adverse comments from any observer in the presence of pupils.

Each teacher shall be given a copy of any written evaluation report intended for his file and shall have the opportunity to discuss such report with the evaluating administrator. After such discussion, the teacher shall sign the report, indicating only that he has read and discussed the evaluation. The teacher's signature does not indicate agreement with the report. The teacher shall have the right to place in his file a response/rebuttal to any written evaluation within twenty (20) work days of receipt of the evaluation. The response/rebuttal shall be attached to the written evaluation. This response/rebuttal shall be submitted by the teacher to the evaluator and Human Capital and signed by the person making the original evaluation, as an acknowledgement that it has been called to his/her attention.

Any adverse evaluation of a teacher's performance placed in his file may be subject to the grievance procedure herein set forth, but only on the ground of bad faith and/or discrimination.

6. Procedure for Filing

The evaluator shall be responsible for submitting to Human Capital, as often as required, an evaluation of each teacher to be evaluated.

Evaluation records shall be confidential and access to such records shall comply with State and Federal law.

7. Provisions for Amendments

The procedures, criteria, instruments, and process of evaluation shall be subject to review and appraisal as required by law and the negotiations process. Any legislative act, State Department ruling, or court decision which makes any part of this provision unlawful will in no way invalidate the rest of this provision.

8. Detailed Specifics of the Observation and Evaluation Process

The TLE Observation & Evaluation Handbook for Evaluators using the Tulsa Model and the Rubric/Observation Forms delineate the specific steps, timelines and processes that operationalize the Rubric/Observation/Processes and Feedback/Support components. These processes include the Personal Development Plan and Goal Setting Process/Form and the intensive 2-month feedback and support process which complements and follows the Personal Development Plan for select staff members.

The TLE Observation & Evaluation Handbook for Evaluators using the Tulsa Model and the Rubric/Observation Forms (identified as process components, e.g., Sections 2 through 8) are a part of the Master Contract and subject to the negotiation's process. Said documents are incorporated by reference and will be available on the District website.

Changes, additions and/or deletions to the TLE Observation & Evaluation Handbook for Evaluators using the Tulsa Model and the Rubric/Observation Forms shall follow the established procedures of agreed upon Memorandum of Understanding in process, format and design. If such changes, additions and/or deletions occur during the term of this Contract date/time, identified errata documentation will be attached to the TLE Observation & Evaluation Handbook for Evaluators using the Tulsa Model.

Causes for Dismissal or Nonrenewal of Teachers

Dismissal and nonrenewal of teachers shall be as provided by law.

Standards of Performance and Conduct for Teachers

Teachers are charged with the education of the youth of this State. In order to perform effectively, teachers must demonstrate a belief in the worth and dignity of each human being, recognizing the supreme importance of the pursuit of truth, devotion to excellence, and the nurture of the democratic principles.

In recognition of the magnitude of the responsibility inherent in the teaching process and by virtue of the desire of the respect and confidence of their colleagues, students, parents, and the community, teachers are to be guided in their conduct by their commitment to their students and their profession.

Appendix D

IRB Approval

Institutional Review Board for the Protection of Human Subjects

Human Research Determination Review Outcome

Date: March 26, 2015

Principal Investigator: Elizabeth Brands Vereecke

Study Title: STRUCTURE AND THE TEACHER EVALUATION TOOL:WHAT REALLY MATTERS IN CREATING EFFECTIVE TOOLS FOR TEACHER DEVELOPMENT

Review Date: 03/26/2015

I have reviewed your submission of the Human Research Determination worksheet for the above-referenced study. I have determined this research does not meet the criteria for human subject's research. The proposed activity will include the researcher using existing, de-identified data to run secondary analyses. Therefore, IRB approval is not necessary so you may proceed with your project.

If you have questions about this notification or using iRIS, contact the HRPP office at (405) 325-8110 or irb@ou.edu. Thank you.

Cordially,

Lara Mayeux, Ph.D.
Vice Chair, Institutional Review Board