

STUDENT SENTIMENT ABOUT TEACHER AND LEADER
EFFECTIVENESS

By

CHRISTOPHER D. POSEY

Bachelor of Arts in Church Ministries
Southeastern University
Lakeland, Florida
1993

Master of Business Administration
Oklahoma State University
Stillwater, Oklahoma
2007

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2022

STUDENT SENTIMENT ABOUT TEACHER AND LEADER
EFFECTIVENESS

Dissertation Approved:

Dr. Katherine Curry

Dissertation Adviser

Dr. Ed Harris

Dr. Jam Khojasteh

Dr. Tom Brown

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Katherine Curry, whose patience and enthusiasm seemed to be limitless over the course of this program. She has been more than just an advisor and educator to me; she has been an inspiration in academia. Her advice and encouragement have truly carried me through to the completion of this pursuit. I am sincerely grateful for her kindness, help, direction, and mentorship. I would also like to thank the members of my dissertation committee: Dr. Tom Brown, who has served as my outside member and whose influence during my master's studies still impacts me daily in my career; Dr. Ed Harris, who was my first contact in this program and my initial influence in starting down this life-changing path; and Dr. Jam Khojasteh, who has patiently worked with me both in class and during the writing of this dissertation to help me wrangle an academic discipline for which I do not have a natural inclination.

Thank you to my family for supporting me in this endeavor. Without question, the sacrifices a student makes are far from the only required sacrifices in a pursuit such as this. The completion of this milestone would not have been possible without their love and support.

Thank you to the teachers and administrators of the district that participated in this study for helping me frame my research and for extending to me the opportunity to conduct my research. Thanks also to the students in the same district who permitted me to interrupt their studies for a few brief moments to learn more about an issue that is extremely important to me.

Special thanks to Dr. Marla Robinson, who for years has served as a wise sounding board within my doctoral cohort, and to Ashley Keener, who invested much time in helping me devise and articulate essential, foundational, analytical themes of this study.

Name: CHRISTOPHER D. POSEY

Date of Degree: JULY, 2022

Title of Study: STUDENT SENTIMENT ABOUT TEACHER AND LEADER
EFFECTIVENESS

Major Field: EDUCATIONAL LEADERSHIP AND POLICY STUDIES

Abstract: Despite the fact that instruments measuring student sentiment about teacher effectiveness have been shown to be reliable and valid, many instruments used by administrators to evaluate teachers simply do not capture student perceptions of teacher effectiveness. One such instrument is the Teacher Leader Effectiveness (TLE) instrument. A study that examines student perceptions of effective teaching, especially as these perceptions relate to traits measured by the TLE, would make important contributions to the literature pertaining to student feedback and secondary teacher effectiveness, and it may inform evaluation processes as districts begin to adopt student feedback instruments as part of the overall evaluation system. The purpose of this study was to gain a better understanding of student sentiment about characteristics of effective teachers for the purposes of informing a survey instrument that captures student perspectives. Additionally, this study compared student identified characteristics with indicators of teacher effectiveness identified in the TLE instrument. Finally, student sentiment about effective teaching practices were considered in tandem with administrator sentiment about the same practices. This study was designed to measure student sentiment about teacher effectiveness qualities assessed through the TLE. Survey data was used to gain a better understanding of student insights on 18 of the 20 dimensions measured on the TLE rubric. Students were asked to watch a video of a teacher and assess the teacher's teaching effectiveness. Administrators at the same school were asked to watch the same video. Students and administrators evaluated the teacher using a modified version of the TLE. Descriptive statistics from all surveys were examined, and similarities and differences between student and administrator scores on the second survey were analyzed. While population sizes posed a challenge to generating statistically significant results, several interesting findings were presented.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION	1
Statement of Problem	4
Statement of Problem	5
Theoretical Framework	6
Research Questions	7
Methodology Summary	7
Importance of the Study	8
Scope of Study.....	9
Definition of Terms	10
Summary	11
II. REVIEW OF LITERATURE.....	13
Introduction	13
History of Teacher Evaluation	13
The Insufficiency of Current Teacher Evaluation Methods	21
The Evolution of the TLE Evaluation Instrument.....	22
Elements of the TLE.....	26
Theoretical Framework: Cotton's Conceptualization of Teacher Effectiveness.....	32
CCTE, the TLE, and Today's Students.....	38
III. METHODOLOGY	41
Research Design.....	41
Research Population	45
Instrumentation and Researcher Position in the Study.....	45
Research Procedures.....	46
Sample Selection	47
Phase One – Student Survey	47
Phase Two – Student and Administrator Evaluation of Online Teaching Sample.....	48
Data Analysis	49
Assumptions	49
Limitations	49
Summary	52

Chapter	Page
IV. FINDINGS.....	53
Data Supporting Research Questions 1 and 2	54
Data Supporting Research Question 3.....	56
Examination of the Data.....	58
Comprehensive Results	64
V. CONCLUSIONS.....	65
Summary of Key Findings	66
Findings.....	66
RQ1 and RQ2.....	66
RQ3.....	70
Findings Overview	73
Suggestions for Future Research.....	75
Implications for Future Research, Policy, and Practice.....	76
REFERENCES	79
APPENDICES	94
APPENDIX A: Survey Items.....	94
APPENDIX B: IRB Approval.....	100

LIST OF TABLES

Table	Page
1. Elements of the TLE	31
2. Cotton's Attributes Keyed to the TLE.....	39
3. Student Response Means for Survey #1	54
4. Student and Administrator Descriptive Statistics for Survey #2	58

LIST OF FIGURES

Figure	Page
1. TLE Evaluation System Timeline (McGee, 2013)	24
2. Survey #2 – Student Means vs. Administrator Means	74

CHAPTER I

INTRODUCTION

Ellen Moir, Brock Prize and Laureate, noted during her keynote speech at the 2014 Brock Prize Symposium held at The University of Oklahoma that, “We have a moral obligation to students to help them soar from early on” (Moir, 2014). Essential in this effort to “help students soar” is the presence of a qualified teacher in every classroom (Darling-Hammond, 2000). This understanding is supported by findings in the literature that suggest that teachers are the most influential factor in promoting student success (Klem & Connell, 2004; Furrer, Skinner, & Pitzer, 2014; Del Siegle, Rubenstein, & Mitchell, 2014). Additionally, in order to help students soar, teachers must continue to improve their skills and learn to adjust to changing student needs and demographics (Daggett, 2014).

Most often, rigorous systems of accountability that include professional teacher evaluation for instructional improvement are utilized to evaluate teacher performance (Tschannen-Moran, Hoy & Hoy, 1998; Tschannen-Moran and Hoy, 2001). Modern teacher evaluation systems in the US, although generally utilized summatively as a measure of teacher performance, are often developed with the primary goal of enhancing teaching effectiveness for enhanced student outcomes (Ellett & Teddlie, 2003). Scholars in the field of supervision, however, argue that, because of the evaluative component of most systems, few accountability systems actually result in improved teaching performance (Glickman, Gordon, & Ross-Gordon, 2018). Another important factor in considering the practical efficacy of teacher evaluation systems in the U.S. is that most evaluation systems are developed by researchers or companies

with little consideration of student, or client, perceptions of teacher qualities or characteristics that lead to enhanced learning (Chen & Hoshower, 2003).

Evaluation of teaching performance is considered a professional responsibility of trained administrators, and student voice is often left out of evaluations of teaching effectiveness. Unlike most businesses, many states lack uniform teacher accountability standards that tie “client-based” evaluation to tenure (Doherty & Jacobs, 2013). Performance-based evaluations utilized in the corporate world, however, focus necessarily on the client, for it is the conversion and retention of clients that keep businesses afloat. In contrast, although state and federal funding is allocated according to average daily attendance of students, the voice of the educational client, the student, is often excluded from considerations pertaining to teacher evaluation.

Accountability standards such as those defined by No Child Left Behind (NCLB) were designed to ensure the competence and abilities of classroom teachers. At its most foundational level, NCLB defined highly qualified teachers as having a bachelor's degree, full state certification or licensure, and the ability to prove that they know each subject they teach (US Department of Education, 2004). Such standards focus on teacher education and certification – undeniably important aspects of high-quality teachers – yet, while standards are in place to govern the delivery of a quality educational product, research indicates that students are at times inexplicably underserved (Cook-Sather, 2010; Lawrence & McCollough, 2004). Many methods of evaluation are seen as providing little more than subjective principal feedback based on an incomplete view of a teacher’s true abilities (Ingle, Rutledge, & Bishop, 2011). Few formal teacher evaluation tools go so far as to consider the opinions of students with regard to teacher effectiveness.

Evidence suggests that educational systems are beginning to borrow from human resource counterparts in the corporate world and are adopting an evaluation system that relies to

some extent on the satisfaction of clients, the *students*, with regard to the product/service provided (Mahatmya, Brown, & Johnson, 2014). This trend does not suggest that the fate of a teacher's career would lie solely in the hands of his or her students any more than a corporate service provider's career would rest entirely in the hands of a single customer and at the exclusion of consideration of other key performance indicators. This trend merely suggests that the opinion of one's client is beginning to be recognized as a key element of success in education.

Harris, Ingle, and Rutledge (2014) report that educators (including principals) across the country do, at times, receive informal feedback from students and parents. While this feedback may be more candid and possibly, at times, even more accurate than that derived from formal teacher evaluation systems, this feedback that comes "straight from the horse's mouth" is often undocumented, unregulated, at best unverifiable, and at worst, an unreasonable tirade taking place in the midst of anger (Goodboy, 2011). However, the ineffectiveness of purely informal communication does not minimize the need for a focus on the student client in educational settings. On the contrary, "[i]t is essential to remember that, ultimately, the best interests of the school's primary clients, the students, must be protected" (Stronge, 1995) and acknowledged. Client feedback in any formal evaluation system is important, even essential, yet the vast majority of districts in the Oklahoma educational system continue to invoke appraisal systems that focus on isolated administrator observations and exchanges that occur outside the context of active, natural teaching. In sharp contrast, Mahatmya, Brown, and Johnson (2014) view teachers as "service-sector professionals, just like physicians, attorneys, and mental health workers, who must cooperate with clients to achieve success."

When invoking the feedback of the student client in the formal teacher evaluation process, two truths should be front-of-mind. First, "[b]ecause job performance must be reflected in behavior in order to be evaluated, this step involves the identification and selection of behaviors (i.e., performance indicators) that are reflective of key professional responsibilities"

(Stronge, 1995). Performance indicators such as behavior, competence, and delivery are characteristics most evident to a teacher's students.

Second, a client-sensitive evaluation model in public schools should be used in tandem with other evaluation instruments ('MET' Made Simple, 2012). While student sentiment may be considered to be critical in assessing teacher effectiveness, this consideration should exist within a broader, comprehensive battery of evaluations gleaned from a variety of evaluation sources. Teachers should not need to worry about placing their future in the hands of embittered adolescents who may be looking to take revenge for the *F* received on the research paper that was due (but not submitted) the week before prom. While student insight may be an essential component of a comprehensive teacher evaluation instrument, evaluators need also consider the many other aspects of teaching in creating and implementing a sound and thorough method of evaluation ('MET' Made Simple, 2012).

Statement of the Problem

Recent trends in teacher evaluation systems across the United States indicate a belief that students have a unique perspective into the performance and effectiveness of their teachers (Cook-Sather, 2002). Aleamoni (1981, 1987, 1999) attributes the value of student insight into teacher performance evaluation to the student's unique rapport with the teacher and the unique communication level between student and teacher. Peterson (2000) conceptualizes students as "stakeholders" and "consumers" of good teaching, noting the value of their contribution to teacher evaluation that comes from having closely and recently observed their teachers.

As teacher evaluation systems across the country have begun to acknowledge the value of student feedback as part of formal teacher evaluation (Wilkerson, et al., 2000; Millman & Darling-Hammond, 1990), instruments and procedures specifically designed to gather and analyze student feedback have been created. Many of these instruments yield high internal

consistency reliabilities (Aleamoni, 1999; Arubayi, 1987; Costin, Greenough, & Menges, 1971; Marsh, 1984; Peterson, 2000), indicating that the assessment of student sentiment about teacher effectiveness can indeed be a legitimate contributor to the teacher evaluation process.

Despite the fact that instruments measuring student sentiment about teacher effectiveness have been shown to be reliable and valid (Shevlin, Banyard, Davies, & Griffiths, 2000), many instruments used by administrators to evaluate teachers simply do not capture student perceptions of teacher characteristics necessary for teaching effectiveness. Most of the time, student instruments are designed with preconceived indicators of teaching effectiveness (Campbell, Kyriakides, Muijs, & Robinson, 2004). One such teacher evaluation instrument that does not incorporate student sentiment about teacher effectiveness is the Teacher Leader Effectiveness (TLE) evaluation instrument, used extensively in the state of Oklahoma and elsewhere in the United States. A study that examines student perceptions of effective teaching, especially as these perceptions relate to traits measured by the TLE, would make important contributions to the literature pertaining to student feedback concerning secondary teacher effectiveness, and it may inform evaluation processes as districts begin to adopt student feedback instruments as part of the overall evaluation system.

Statement of Purpose

In most corporate contexts, client input into employee evaluation is an essential part of organizational growth and evolution. Such evaluations enable organizations to adopt best practices, improve performance, and increase productivity (Zhu, 2014). While this principle is relatively self-evident, the standardization of client-focused evaluations, particularly in an educational context, is elusive. Much has been written on styles and modes of teacher evaluation. Much research exists on the validity of teacher evaluation methods – even methods that in some way include student feedback. However, little research includes consideration of student

sentiment about characteristics of effective teachers or relationships between the evaluation instrument being utilized in these cases and student perceptions of teacher effectiveness. Even less research focuses on how such a system may complement existing teacher evaluation protocols such as the TLE.

The purpose of this study is to gain a better understanding of student sentiment about characteristics of effective teachers for the purposes of informing a survey instrument that captures student perspectives. Additionally, this study compared student identified characteristics with indicators of teacher effectiveness identified in the TLE instrument. Finally, student sentiment about effective teaching practices were considered in tandem with administrator sentiment about the same practices.

Theoretical Framework

The theoretical framework that undergirds this quantitative study is based on the research of Kathleen Cotton and the Northwest Regional Education Lab. Cotton's research (2000) focused on the identification of core contextual and instructional factors common to teachers who were perceived to be "effective." From these two domains, Cotton (2000) further identified specific traits of effective teachers – 10 contextual attributes and five instructional attributes. These attributes and the relevant body of research conducted by Cotton, et al., shall, herein, be referred to as Cotton's conceptualization of teacher effectiveness (CCTE).

Nearly a decade after Cotton published her seminal work, a school district in Oklahoma, Tulsa Public Schools (TPS), in an attempt to improve its formal teacher evaluation system, assigned a task force to synthesize various teacher evaluation documents and studies. Once completed, a smaller task force then created a rubric that ultimately embodied the primary principles of effective teaching as identified by Cotton ("Tulsa Public," 2011). TPS stakeholders organized the traits and attributes identified in CCTE into five unique domains: classroom

management, instructional effectiveness, professional growth and continuous improvement, interpersonal skills, and leadership. Once identified, these five domains were then scrutinized and further synthesized into 20 different dimensions that make up the whole of the TLE rubric that was used at the time of this study (“Tulsa Public,” 2011).

Cotton’s contextual and instructional attributes serve as the foundation of the TLE rubric domains and dimensions, and relevant attributes will be observed in the context of the theoretical framework of this dissertation.

Research Questions

The following research questions will be pursued over the course of this study, utilizing the methodology described below:

- What teacher characteristics do students identify as being necessary for effective teaching?
- To what extent do these student-identified characteristics align with characteristics outlined in the TLE?
- Are there similarities in student and administrator sentiment about effective teaching practices?

Methodology Summary

This study was designed to measure student sentiment about teacher effectiveness qualities assessed through the TLE. Survey data was used to gain a better understanding of student insights on 18 of the 20 dimensions measured on the TLE rubric.

At the outset of this study, a list of questions pertaining to teacher effectiveness characteristics included in 18 of the 20 TLE dimensions was generated. Students were then surveyed on a 5-point Likert scale about how strongly they felt characteristics in each question

truly contribute to the effectiveness of teachers. Students also had the opportunity to write in any additional effective teaching characteristics about which they felt strongly, but which were not included in the 18 questions. Descriptive statistics were examined, and traits with high mean scores were placed into a new survey that was given to the same group of students. Write-in responses were also appended to the new survey.

Students were then asked to watch a video of a teacher who does not teach in their school district. Students were asked to assess the teacher's teaching effectiveness using the modified TLE-based evaluation instrument that evolved from the first student survey. Administrators at the same school were also asked to watch the same video the students were asked to watch. Administrators were then asked to complete the same modified survey that students were asked to complete. Similarities and differences between student and administrator scores were analyzed.

Importance of the Study

This study will be significant to theory in that it will serve as a catalyst for the continued engineering of valid, reliable, student-inclusive teacher feedback analyses. It will fill a gap in the literature that pertains to indicators of teaching effectiveness, as perceived by students, especially as they relate to teacher effectiveness principles strongly supported by students and assessed by the TLE.

This study is significant to theory in that it further frames students as end-users and consumers of an educational "product." In addition, the study advances the theoretical undertones of Kathleen Cotton's research, presenting themes from her research as theoretical foundations that can fortify future study on the topic.

At its most basic, this study is significant to practice in that it will impact and inform the methods used in the formal evaluation of secondary school teachers. It serves as a practical reminder that the most immediate client of every teacher is the individual student. This study

intends to lay groundwork for evaluation systems that genuinely value and include the voice of students. Ideally, it provides a non-threatening, intuitive method of including students in the teacher evaluation process. Findings from this study serve to revise existing evaluation instruments or perhaps even create new instruments.

Scope of Study

This study analyzes secondary student sentiment about characteristics of effective teachers and compares those sentiments with effectiveness principles that are specific to the TLE instrument. More specifically, teacher characteristics that were identified were compared with teacher effectiveness principles articulated in four of the five domains evaluated using the TLE. These domains are Classroom Management, Instructional Effectiveness, Interpersonal Skills, and Leadership.

Items that fall outside the scope of this study include the analysis of student sentiment about teacher effectiveness characteristics and qualifications specific to the TLE domain of Professional Growth and Continuous Improvement. Also, while it does reside within an analyzed TLE domain (Instructional Effectiveness), the dimension of Current State Standards does not fall into the scope of this research and is not addressed due to the limits and locality of the issue. Additionally, the study does not consider sentiment among elementary school students, nor does it consider sentiment of students at the university level. The study does not include perceptions of effective teacher characteristics and practices with any evaluation instrument other than the TLE.

Definition of Terms

Classroom Management

In the context of this study, *classroom management* is an effectiveness domain that includes the dimensions of preparation, discipline, building-wide climate responsibility, lesson planning, assessment practices, and student relations (“Teacher and Leader,” 2015).

Contextual Attributes

Contextual attributes, as identified by Kathleen Cotton (2000), pertain to “strong administrative leadership and a schoolwide focus on learning.”

Instructional Attributes

Instructional Attributes, as identified by Kathleen Cotton (2000), pertain to the cycle of development, execution, and revisiting of classroom lessons and instruction.

Instructional Effectiveness

In the context of this study, *instructional effectiveness* is an effectiveness domain that includes the dimensions of literacy, current state standards, involvement of all learners, explanation of content, clear instruction and directions, modeling, monitoring, adjustments made based upon monitoring, closure, and student achievement (“Teacher and Leader,” 2015).

Interpersonal Skills

In the context of this study, *interpersonal skills* are an effectiveness domain that includes the dimension of effective interpersonal skills (“Teacher and Leader,” 2015).

Leadership

In the context of this study, *leadership* is an effectiveness domain that includes the dimension of professional involvement and leadership (“Teacher and Leader,” 2015).

Teacher and Leader Effectiveness

In the context of this study, Teacher and Leader Effectiveness (or “TLE”) refers not to the general principle of the ability of a teacher to perform his or her prescribed function of teaching with skill and accuracy, but instead to the evaluation instrument by the same name that is utilized in educator evaluations.

Summary

Chapter I introduces the value of student sentiment pertaining to teacher effectiveness, particularly with regard to the foundational principles of teacher effectiveness articulated in the TLE instrument. This observation evolves three questions to be answered in this study. The first question examines characteristics of effective teachers from the viewpoints of their students. The second is similar to the first; however, it examines key qualifications of effective teachers, again from the viewpoint of their students. The third question considers similarities and differences between student perceptions of teacher effectiveness and administrator perceptions of teacher effectiveness.

Chapter II of the study provides a review of the literature on teacher evaluation history, background, and methods, with specific focus on methods that incorporate student feedback. It acknowledges problems frequently associated with evaluation instruments that include student feedback and follows with a sort of counterpoint that examines the validity and reliability of such instruments. The chapter culminates in a demonstration of the value of student feedback in the realm of teacher evaluation and the theoretical framework that undergirds the study as a whole,

providing a thorough examination CCTE as the theoretical framework for the study. As a part of this examination, CCTE is keyed to the TLE instrument currently administered in public schools by evaluating administrators. CCTE is shown to be a foundational body of research for the TLE. In addition, the value of principles brought forth in CCTE and furthered in TLE are examined in the context of valuable student feedback.

Chapter III describes the research design and methods utilized in this study. Justification for the choice of methods is presented. Also included in this chapter is a description of choice and use of strategies and tools for gathering and analyzing the data. Relevant limitations and assumptions of the study are also acknowledged in this chapter.

Chapter IV presents findings from the descriptive statistics pertaining to the study of concepts in the TLE as they align with student sentiment about effective teaching. It also reveals similarities and differences between student perceptions of effective teaching and administrator perceptions of effective teaching.

Chapter V discusses findings through the lens of CCTE. The chapter concludes with implications for practice and research.

CHAPTER II

REVIEW OF LITERATURE

Introduction

A review of the relevant literature helps to expose an anomaly that exists in current educational accountability standards pertaining specifically to teacher evaluation. It also reveals how an assessment of student sentiment about the Teacher Leader Effectiveness (TLE) evaluation instrument may provide a dimension that is missing from existing teacher evaluation tools.

History of Teacher Evaluation

The development of effective, formal teacher evaluation systems is a relatively new science, yet the desire for highly effective teachers in American schools has existed since colonial times. There were no portfolio requirements or nervous meetings in the principal's office during these times. Instead, a system of teacher inspections designed to evaluate a teacher's effectiveness was conducted by local communities to ensure adherence to religious mores (Jewell, 2017). The so-called "standards movement" is said to have begun more than a century ago in 1894, with the release of the National Education Association's Report of the Committee of Ten (Klock, 2010). The report was one of the first that defined educational standards in terms of measurable student behaviors and was the beginning of a national emphasis on education.

The desire for a strong, national education program continued, and in 1937, politician and American educational reformer, Horace Mann, lobbied the Massachusetts legislature to create a state board of education (Jewell, 2017). This push included the institution of "normal schools,"

which were later put in place to provide formal teacher training – an educational model that combined administrative oversight and teacher observation and feedback (Jewell, 2017). This model laid the foundation for formal teacher evaluation in the US.

Education legislation remained a responsibility of the individual states until around 1965 with the evolution of the Elementary and Secondary Education Act (ESEA), which brought about the modern standards-based reform movement. ESEA was introduced to serve disadvantaged children in the US, and it was ESEA that enlarged the federal role in education (Jewell, 2017). Title I, Part A of the ESEA “provides financial assistance to local educational agencies (LEAs) and schools with high numbers or high percentages of children from low-income families to help ensure that all children meet challenging state academic standards (“Improving Basic,” 2015). These LEAs and schools, in order to receive federal funding, were required to participate in standardized testing (Ravitch, 2000). Funding was provided through four formulas that were based on poverty estimates and the cost of education in each state:

- Basic Grants provide funds to LEAs in which the number of children counted in the formula is at least 10 and exceeds 2 percent of an LEA's school-age population.
- Concentration Grants flow to LEAs where the number of formula children exceeds 6,500 or 15 percent of the total school-age population.
- Targeted Grants are weighted so that LEAs with higher numbers or higher percentages of children from low-income families receive more funds.
- Education Finance Incentive Grants (EFIG) distribute funds to states based on factors that measure a state's effort to provide financial support for education compared to its relative wealth; and the degree to which education expenditures among LEAs within the state are equalized (“Improving Basic,” 2015).

Beyond the funding benefits provided by ESEA, the Act also served to challenge educators to rethink current processes and to develop new theory and practice both in classroom teaching and in the administration of teacher evaluation. The movement took hold post-Sputnik, in the 1970s and 1980s with the minimum-competency testing movement, which was spurred by a desire on behalf of disenchanted parents who felt that their children were performing poorly on standardized tests because of a lack on the part of public schools to teach “the basics” (Koretz & Deibert, 1995). Schools were encouraged to remove “frills” and return to an intensive focus on reading, writing, and arithmetic, with emphasis on concrete educational outcomes (Brookheart, 2013).

In the 1970s, Madeline Hunter devised a seven-step model of mastery learning, which eventually evolved into a teacher evaluation platform in a number of states (Marzano, 2011). Hunter’s model included script-taping and a lesson design model that encouraged quantifiable teacher growth and learning (Marzano, 2011). Later in that decade, amidst a backdrop of significant changes in the US, the federal government further introduced itself into the educational process by creating the US Department of Education. Officially instituted in 1979, the Department endeavored to create educational equality for all students and to encourage local control over educational standards, including teacher evaluations (Stallings, 2002).

While the National Council for Accreditation of Teacher Education (NCATE) had been around since 1954, it was during the 1980s that many states began to adopt standards set forth by the Council (Bales 2006). This marked the beginning of a shift of control of accreditation and certification from state-level accountability to federal-level accountability. This was an important step in consolidating recruitment, program approval, licensing, and in-service professional development policies across state lines (Bales, 2006). Previously, program “approval and licensing policies were so diverse across state lines that the quality of teachers prepared in one state could not be compared with those prepared in another state” (Bales, 2006). The NCATE-

prompted shift from state- to federal-level accountability opened the door to address this problem. NCATE later merged with the Teacher Education Accreditation Council (TEAC), to form the Council for the Accreditation of Educator Preparation (CAEP).

In 1983, a major milestone in the history of teacher evaluation came about. Amidst growing concern about the condition of the educational institution in America, President Ronald Reagan's National Commission on Excellence in Education published a report that examined the education system, especially as it pertained to the preparation of the American workforce. This report, entitled *A Nation at Risk*, unearthed “disturbing inadequacies” (“National Commission,” 1983) in the condition of education in the US. The report emphasized the need for “higher academic standards, increased student course requirements, a longer school day, and significant changes in the training and retention of teachers” (Thomas & Brady, 2005). In addition to encouraging a boost in teacher pay, the report recommended tying salaries, promotion, tenure, and retention decisions to effective evaluation systems (Muldoon, 2009). Conclusions and recommendations made in *A Nation at Risk* led to national uniform teacher licensing measures proposed later in the 1990s.

Contributing to educational reforms of the 1980s, a new model designed to measure teacher effectiveness was developed: Value Added Models (VAM). VAM positioned student test performance as an indicator of teacher effectiveness by examining academic growth that students made from one year to the next (Rivkin, et al., 2005). William Sanders first pioneered and applied VAM to individual teachers beginning in the mid-1980s to prove that teacher effectiveness could be measured using student test data (Hill, 2000). In 1994, Congress adopted the reauthorization of the ESEA, known as Improving America’s Schools Act (IASA). In support of VAM, IASA required each state to develop academic standards to be measured through annual objective testing. Schools that failed to increase student achievement, as evidenced by this annual testing, were penalized (Ryan, 2004).

In 1996, another milestone in teacher evaluation evolved. In continued support of the use of VAM for teacher evaluation, a significant piece of research, *Enhancing Professional Practice: A Framework for Testing*, was published by Charlotte Danielson. This report examined 76 unique components of effective teaching. *Enhancing Professional Practice* contributed to the history of teacher evaluation by providing an extensive ranking system for teachers, using the descriptors, *unsatisfactory, basic, proficient, and distinguished* (Marzano, 2011). “The level of specificity supplied in the Danielson model provided the foundation for the most detailed and comprehensive approach to evaluation to that time” (Marzano, 2011).

In 1998, soon after the publication of *Enhancing Professional Practice*, the state of California moved forward with making performance assessments for teacher certification mandatory.

In 2000, Cathleen Cotton, in conjunction with the Northwest Regional Educational Lab, published her seminal work, *The Schooling Practices that Matter Most*. In her publication, Cotton identified 15 core contextual and instructional factors, or attributes, that purportedly enable students to learn successfully (Cotton, 2000). These factors of teaching practice, which are articulated elsewhere in this study, serve as foundational, guiding principles for other formal teacher evaluation systems, including the Teacher Leader Effectiveness rubric used to evaluate public school teachers in Oklahoma at the time of this writing (TLE, 2015). Cotton’s research, particularly her core contextual and instructional factors for teaching, contributed directly to the TLE. “Its rubric assesses many of the contextual factors identified in the Cotton paper” (“Tulsa Public,” 2015).

Building upon the foundation of the Improving America’s Schools Act (IASA), the No Child Left Behind Act (NCLB) was enacted in 2002 by President George W. Bush. IASA had been a reauthorization of ESEA in 1994. It specified what students should be able to know and do

in reading and mathematics and required that states develop standards based on these specifications. Success in teaching these standards was measured through annual objective testing at least once in each of three grade spans: grades 3–5, 6–9, and 10–12 (Taylor, O’Day, & Le Floch, 2010). In addition to instituting accountability measures for teacher effectiveness nationwide, NCLB attempted to provide equal educational quality to poor and minority students (Ryan, 2004); specifically, English-language learners, students in special education, and poor and minority children, whose achievement, on average, trails their peers. Failing to bring these students to a proficient level, a level determined by each state individually, would result in a loss of federal Title I funding (Klein, 2015).

In the summer of 2009, President Barak Obama announced the largest ever federal competitive investment in school reform: Race to the Top (RTT) (“Race to,” 2009). One primary goal of RTT was to provide “viable approaches to measure the effectiveness of teachers, provide an effectiveness rating to each individual teacher, and use those ratings to inform professional development, compensation, promotion, tenure, and dismissal” (Muldoon, 2009). The \$4.35 billion measure challenged states to focus on four central areas of school improvement and reform:

- Adopting internationally benchmarked standards and assessments that prepare students for success in college and the workplace;
- Recruiting, developing, retaining, and rewarding effective teachers and principals, especially where they are needed most;
- Building data systems that measure student success and inform teachers and principals about how they can improve instruction; and,
- Turning around the lowest-achieving schools (“Race to,” 2009).

States were rewarded for past accomplishments and incentivized to make future improvements (“Race to,” 2009). A state would be rewarded if its longitudinal data system included the following 12 elements from the American COMPETES Act of 2007:

1. a unique statewide student identifier that does not permit a student to be individually identified by system users;
2. student-level enrollment, demographic, and program participation information;
3. student-level information about the points at which students exit, transfer in and out, drop out, or complete P-16 education programs;
4. the capacity to communicate with higher education data systems;
5. a state data audit system that assesses data quality, validity, and reliability;
6. yearly test records of individual students' performance on NCLB-required tests;
7. information on students not tested by grade and subject;
8. a teacher identifier system with the ability to match teachers to students;
9. student-level transcript information, including information on courses completed and grades earned;
10. student-level college readiness test scores;
11. information on the extent to which students transition successfully from secondary school to postsecondary education, including whether students enroll in remedial coursework; and
12. all other information necessary to address alignment and adequate preparation for success in postsecondary education (Lohman, 2010).

In 2013, the Bill and Melinda Gates Foundation introduced a project that encouraged continued reform in teacher evaluation procedures nationwide. The Measures of Effective Teaching project (MET) proposed several alternatives to weighting evaluations so that teachers would not “focus too narrowly on a single aspect of effective teaching and neglect its other

important aspects” (Cantrell and Kane, 2013). For example, if it is desirable for students to master a broad set of learning objectives – a set that goes beyond the objectives assessed during a high-stakes standardized test – MET provides other venues to identify teacher practices that produce valued outcomes.

Following closely on the heels of RTT and MET was the Every Student Succeeds Act (ESSA), the latest reauthorization of ESEA. Signed by President Barak Obama in 2015, ESSA shifted the focus of measuring student success from uniform national testing to state-driven assessments (“Every Student,” 2015). This measure addressed concerns on behalf of teacher, parents, and students that high-stakes testing was insufficient to comprehensively address teacher performance (Au & Gourd, 2013). The Act intended to:

- include multiple forms of student assessment;
- promote state-driven standards;
- provide intervention and funding for the lowest-performing schools;
- allow for state determination and creation of evaluation systems;
- generate programs to reward effective teachers;
- increase the number of STEM teachers; and,
- provide resources to encourage data-driven systems and creative approaches to education (“Every Student” 2015).

The Act requires that all states adopt standards-based accountability systems to monitor school performance. Performance standards are tied to specific assessments that serve as measuring tools. These tools become a primary basis for reporting student performance, as well as school and district performance (Koretz & Deibert, 1995). ESSA ultimately replaced NCLB and emphasized the provision of equal learning opportunities to all students (“Every Student,” 2015).

While many teacher accountability models are currently in place that are intended to fulfill requirements outlined by NCLB (Towles-Reeves, Kleinert, & Muhomba, 2009) and, subsequently, ESSA, the two pieces of legislation had significant differences. NCLB brought with it an unprecedented level of federal involvement in education (Au & Gourd, 2013). In addition, NCLB introduced a new norm that emphasized the standards movement and high-stakes testing (Au & Gourd, 2013). ESSA resulted from dissatisfaction on behalf of educators over what was perceived as an over-emphasis of a single measure of accountability (Smith & Kovacs, 2011). The Act required using *multiple* measures of student assessment, and to permit students to demonstrate academic achievement in ways other than high stakes exams (“Every Student,” 2015). As a result, teacher evaluations grew in both scope and relevance as teachers gained enhanced responsibility for student outcomes (Fennell, 2016).

The Insufficiency of Current Teacher Evaluation Methods

While federal, state, and local teacher accountability standards are clearly in place across the US, educational reform at times produces disappointing results (Clark & Astuto, 1994). Formal teacher evaluation tools have earned a reputation for failing to provide technical or instructional substance to teachers in need of development (Little, 2009). Many are seen as providing little more than subjective principal feedback based on an incomplete view of a teacher’s true abilities (Ingle, Rutledge, & Bishop, 2011). For example, some teacher evaluations involve two or three pre-announced observations by a principal or administrator. The curriculum and interaction observed during these few observations accounts for only a fraction of the teacher’s teaching portfolio for the school year. In addition, because the observation is pre-announced, teachers may modify their teaching norm, leaving the observer with a skewed idea of the teacher’s true teaching abilities.

Additionally, the standards movement has come under criticism in recent years. Some standards are said to repress creativity, social interaction, and reflection (Brott, 2016; Saltman, 2017), and an emphasis on math and literacy has created a two-dimensional curriculum focus on math and language arts – sometimes even to the extent of stealing away class focus that had been previously allotted to staples like social studies and science (Winstead, 2011). Such issues of narrowing curricula, in addition to the reduction of student assessment to a series of multiple-choice questions on state-mandated standardized tests that determine school funding, have resulted in a great deal of concern about the level to which teachers are able to meet their students' educational needs (McReynolds, 2006). This standards movement has narrowed the curriculum, enhanced stress for students and teachers by placing an enormous emphasis on student performance on high stakes evaluations, and created complexity in teacher evaluation systems by directly relating teacher effectiveness and student performance.

Teacher evaluation tools that are directly tied to student performance, and especially to student performance on standardized tests, have been directly criticized for the following reasons:

- Statistical evidence does not indicate a correlation between VAM and improvement on student standardized tests.
- While much of the enthusiasm behind VAM is based on the notion it parallels the evaluation methods used in private sector, the measurement of private sector employee performance almost never depends on a single, narrow quantitative measure.
- Standardized tests now in use are both imperfect and incomplete. Their measurements of student achievement are not without question. (Baker, et al., 2010).

The Evolution of the TLE Evaluation Instrument

As the standards movement evolved, teacher evaluation was viewed as insufficient, and because teaching and testing were becoming more standardized, educational leaders and policy

makers emphasized a need for a more standardized, state-wide approach to teacher evaluation (Hamilton, Stecher, & Yuan, 2008). The development of this plan re-emphasized a more structured approach to teacher evaluation across the State. In response, in 2009, one district, Tulsa Public Schools (TPS), began development of a new formal teacher evaluation system. Development consisted of a two-part process, the first of which involved national evaluation experts, TPS teachers, curriculum specialists and principals. These stakeholders were charged with the review of existing teacher evaluation documents and research studies. The group was to make principle-based recommendations on their findings in order to carve the rough structure of TPS's new teacher evaluation system. Once this list of principles was identified, it was then passed to a smaller team that was tasked with creating a correlating evaluation framework/rubric. The creation of this rubric was the second part of TPS's evaluation system development process ("Tulsa Public," 2011). On December 15, 2011, the Oklahoma State Board of Education named the TLE Observation and Evaluation System as the presumptive default for teacher evaluations. During the pilot year of implementation (2012-2013), districts could choose from three teacher evaluation frameworks:

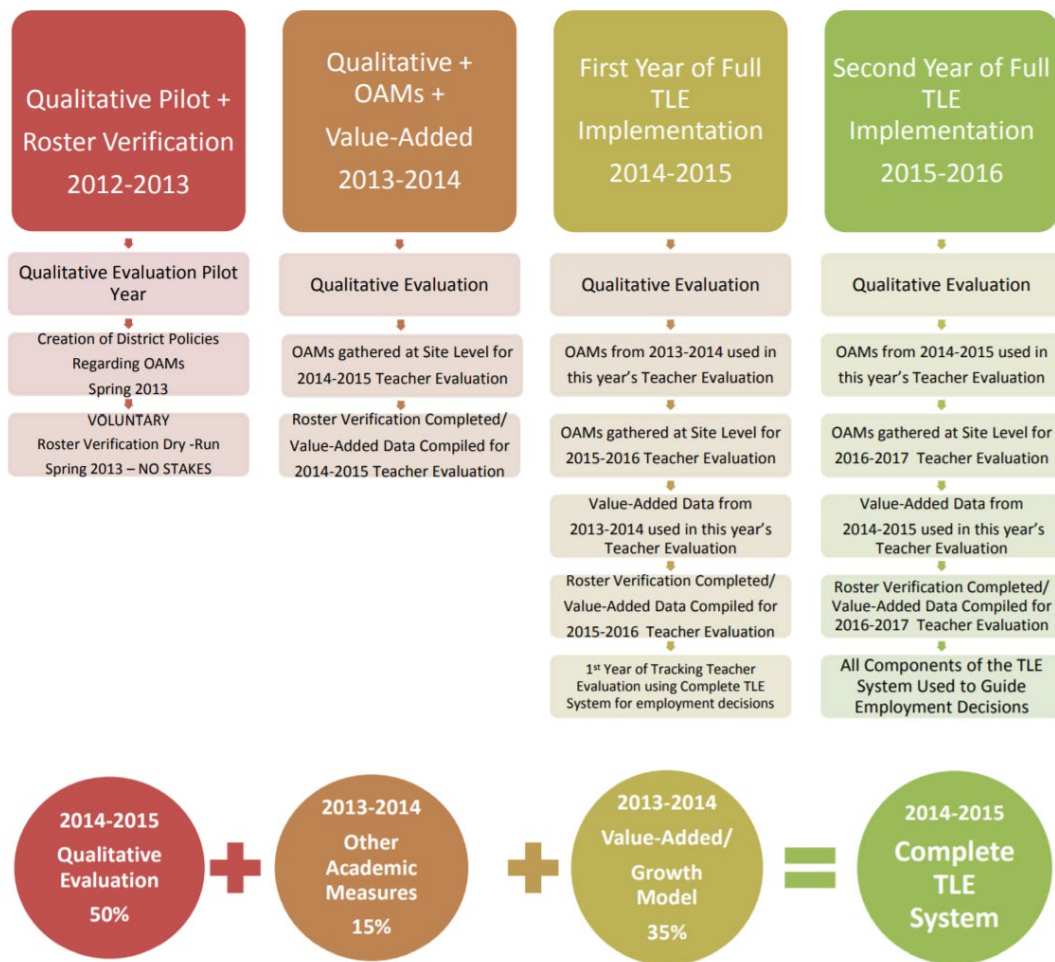
- Tulsa's TLE Observation and Evaluation System
- Marzano's Causal Teacher Evaluation Model
- Danielson's Framework for Teaching

During the pilot period, districts selected and implemented one of these three frameworks. Then, throughout the pilot year, districts provided input and feedback regarding selected frameworks. Teacher and leader evaluations obtained during the pilot year were utilized to gather preliminary data concerning the usefulness of the tool for evaluation across districts. This data was not formally used in teacher evaluation processes and did not count against teachers or leaders during the 2012-2013 school year. The data obtained during the pilot year was used by districts to establish baselines and offer guidance as Oklahoma schools moved forward with permanent

implementation during the 2013-2014 school year (“Teacher and Leader,” 2011). In a split vote, the TLE Commission recommended the TLE Observation and Evaluation System as the default teacher qualitative assessment tool (“Oklahoma Teacher,” 2011). Implementation was ultimately bumped one year to the 2014-2015 school year based on requests of stakeholders. Figure 1 provides an overview of the implementation of the TLE model in the state of Oklahoma.

Figure 1

TLE Evaluation System Timeline (McGee, 2013)



According to the original statute, the TLE system contained seven evaluation components:

1. A five-tier scale which rates teachers and administrators from *ineffective* to *superior*.
2. Annual evaluations to provide feedback to improve student learning.
3. Comprehensive remediation plans and instructional coaching for teachers who receive a rating of *needs improvement* or *ineffective*.
4. A provision that 50% of the ratings of teachers and leaders shall be based on quantitative components, while the other 50% must be based on qualitative components.
 - 35% of the quantitative component must be based on student academic growth using multiple years of standardized test data, and 15% must be based upon other academic measures.
5. An evidence-based qualitative assessment tool to evaluate teachers that includes "observable and measurable characteristics of personnel and classroom practices that are correlated to student performance success."
6. An evidence-based qualitative assessment tool to evaluate administrators that includes "observable and measurable characteristics of personnel and classroom practices that are correlated to student performance success."
7. An assessment using objective measures of teacher effectiveness including student performance on unit or end-of-year tests must be used to evaluate those teachers in grades and subjects for which there is no state-mandated testing measure to create a quantitative assessment for the quantitative portion of the TLE system ("Oklahoma Teacher," 2011).

Qualitative measures of the TLE include observable characteristics of teacher performance that are assessed using the TLE rubric. *Quantitative measures* consist of multiple years of student standardized test data ("Oklahoma Teacher," 2011). *Other academic measures* are "additional alternative instruments ensuring a robust teacher evaluation, capturing unique facets of effective teaching, and reflecting student academic performance impacted by the teacher" ("TLE Commission," 2012).

With regard to teacher assessment, TLE distinguishes between an “evaluation” and an “observation.” According to TLE supplementary documentation, “an observation is when an administrator observes activities within an educational setting” (“Teacher and Leader,” 2012). Teachers in Tulsa who are evaluated using the TLE rubric are required to be observed twice. An evaluation is the “combination of the current year’s observations and quantitative analysis of student achievement, assess an educator’s effectiveness” (“Teacher and Leader,” 2012). Probationary teachers are to be evaluated two times per year using the TLE rubric, and career teachers must be evaluated once per year (“Teacher and Leader,” 2012).

While the teacher assessment is a more standardized form of teacher evaluation, efforts have been made to enhance validity and reliability of the TLE. Concerning reliability, new administrators must attend a three-day training to learn how to administer the TLE. More specifically, new administrators undergo calibration training, and they are educated about the Tulsa TLE Model overview and background, the TLE Rubric, the observation and evaluation process, and the use of rubric indicators (“Tulsa TLE Model,” n.d.). Upon completion of their training, administrators must then pass both a written examination as well as an interrater reliability examination, which qualifies them to be preliminarily certified to evaluate teachers. This certification is valid for two years. (“Teacher and Leader,” 2012). TLE implementation is supported by funds allocated through a formula based on a district’s average daily attendance (“Oklahoma State Board,” 2011).

Elements of the TLE

The TLE evaluation rubric is organized into five unique domains: classroom management, instructional effectiveness, professional growth and continuous improvement, interpersonal skills, and leadership (see Table 1 below). These five domains are further divided into 20 different dimensions that make up the whole of the TLE rubric. Teacher performance is

rated on a scale of 1-5, and teachers receive a corresponding delineation of “Ineffective” (1), “Needs Improvement” (2), “Effective” (3), “Highly Effective” (4) or “Superior” (5).

Four of the five domains originally created by the TPS stakeholders are readily visible to students in a traditional classroom: Classroom Management, Instructional Effectiveness, Interpersonal Skills, and Leadership. More than simply being readily apparent to students, the manifestation of these domains through each teacher’s daily practice results in behavior that may actually be evaluated by both administrators and students. The domain of professional growth and continuous improvement includes assessment of activity and behavior that is not readily visible to students in a way that is directly tied to the domain, and is, therefore, excluded from this study.

Classroom Management. The TLE domain, “Classroom Management” comprises 30% of the total domain emphasis of the TLE. It consists of six dimensions: Preparation, Discipline, Building-Wide Climate Responsibility, Lesson Plans, Assessment Practices, and Student Relations.

According to the TLE Rubric (2015), a teacher who is “superior” in the dimension of Preparation plans for and executes instructional strategies, has long and short-term instructional plans, plans consistently, and has materials and equipment ready at the start of a lesson.

A superior disciplinarian clearly defines, monitors and effectively manages student behavior, engaging students in expectations of discipline and, as necessary, curtails misbehavior in a prompt and consistent manner (“TLE,” 2015).

A step beyond discipline, the TLE (2015) holds teachers responsible for conditions of the larger school environment, even outside of their classrooms. A third dimension within the domain of Classroom Management is that of building-wide climate responsibility. Perhaps somewhat more ambiguous than many other dimensions on the TLE, building-wide climate responsibility

speaks generally to a larger group of school-wide conditions, including project management, discipline and orderly behavior, and school leadership.

On the topic of lesson planning, the TLE (2015) places increased value on those plans that are consistently developed on time (or in advance) and in a collaborative manner. The ideal lesson plan is one that may be used as a grade level model yet is flexible and malleable enough to be revised readily based on relevant student data and performance.

At the other end of the learning cycle is assessment. The TLE (2015) deems assessment practices that “guide and support differentiated instruction” that are fair, that are collaborative, that provide practical, actionable feedback and that does so very soon after the assessment is administered, and that are communicated to students regularly, with plenty of opportunity for course correction when needed, as superior (TLE, 2015).

The final dimension of the Classroom Management domain pertains not to specific instruction, but to the teacher’s interactions with his or her students. The TLE domain of Student Relations emphasizes the optimization of the learning environment “through respectful and appropriate interactions with students” (“TLE,” 2015). Such interactions do not occur in a vacuum, but are explicitly connected to learning, as teachers are to convey “high expectations for students and an enthusiasm for the curriculum” (“TLE,” 2015). Research indicates that respectful and appropriate teacher interactions with students improve learning outcomes (Edmonds, 1979; Jennings and Greenburg, 2009; Louis & Smylie, 2016).

Instructional Effectiveness. Practices demonstrating instructional effectiveness make up the largest percentage of the TLE, garnering literally half (50%) of the focus of the instrument. Dimensions within the domain of Instructional Effectiveness include Literacy, Current State Standards, Involves All Learners, Explains Content, Clear Instructions and Directions, Models, Monitors, Adjusts Based on Monitoring, Establishes Closure, and Student Achievement. While

some peripheral overlap exists between these domains and those attributed to Classroom Management, the TLE presents these domains in a uniquely instructional context.

The first domain of Instructional Effectiveness presented in the TLE is that of Literacy. The instrument exhorts the classroom teacher to embed “the components of literacy into all instructional content.” Specifically, the TLE evaluates the embedding of reading, writing, vocabulary development, spelling, and listening/speaking in all classroom curriculum. Teachers who perform in a superior manner in the domain of Literacy include, in addition to these practices, innovative use of multimedia, computer, information analysis and technology in daily lessons (TLE, 2015).

The TLE places a premium on the engagement of all students in classroom learning activities, acknowledging that the only way such engagement is possible is through diverse teaching and questioning techniques and allowing sufficient “wait time” for students to respond to questioning. The TLE encourages active learning as an essential component of student engagement, providing specific examples of cooperative learning activities, advance organizers, and researching and reporting out, each of which rests on the foundational practices articulated by Collins and O’Brien (2011) of information gathering, questioning, and thinking and problem solving.

Two TLE dimensions within the domain of Instructional Effectiveness are closely related: Explains Content and Clear Instructions and Directions. Both focus on clarity of content manipulation both ahead of and during the instruction dissemination process itself. Both dimensions attribute a variety of methods and modes (e.g., hands-on activities, body language, and thematic instruction as well as modelling and visual demonstrations) to highly effective and superior teaching practices.

Beyond providing clear instructions, teachers help students learn by modeling skills and processes. The TLE includes a dimension pertaining to this skill wherein highly effective teachers are expected to provide demonstrations of educational skills and processes proactively. The TLE then encourages teachers to establish closure by summarizing, discussing main ideas, discussing connections, or any other strategy that provides a reinforcing conclusion to the lesson. Ideally, students will, as a result, “connect the lesson to prior learning and articulate how learned skills can be used in the future” (“Teacher and Leader,” 2015).

Interpersonal Skills. While Classroom Management and Instructional Effectiveness are both essential to quality teaching, either or both of these two domains may be usurped by a teacher’s inability to interact with students appropriately; hence, the inclusion of the domain of Interpersonal Skills is included in the TLE evaluation instrument. According to the TLE rubric (2015), teachers are expected to communicate “sensitively with families and colleagues,” using diverse methods of engagement in working with these stakeholders. At the lower end of influence, the domain of Interpersonal Skills has a relative weight of 5% on the TLE.

Leadership. The final TLE domain is that of Leadership. The domain of Leadership makes up 5% of the relative weight of TLE domains. The domain of Leadership is further qualified in its only dimension, Professional Involvement and Leadership. This domain examines a teacher’s inclination to lead school or district events, to initiate important professional activities, or to lead others to challenge and reject biased, negative, or disrespectful, attitudes or practices (TLE, 2015).

Table 1*Elements of the TLE*

Weight	Domain	Dimension	Indicator
30%	Classroom Management	Preparation	Teacher plans for and executes a lesson relating to short-term and long-term objectives.
		Discipline	Teacher clearly defines and effectively manages student behavior.
		Building-Wide Climate Responsibilities	Teacher assures a contribution to building-wide positive climate responsibilities.
		Lesson Plans	Teacher develops daily lesson plans designed to achieve the identified objectives.
		Assessment Practices	Teacher acknowledges student progress and uses assessment practices that are fair, based on identified criteria, and support effective instruction.
		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.
50%	Instructional Effectiveness	Literacy	Teacher embeds the components of literacy into all instructional content.
		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.
		Involves All Learners	Teacher uses active learning, questioning techniques and/or guided practices to involve all students.
		Explains Content	Teacher teaches the objectives through a variety of methods.
		Clear Instruction & Directions	Teacher provides clear instruction and direction.

Table 1 - cont'd.

Elements of the TLE

Weight	Domain	Dimension	Indicator
		Models	Teacher demonstrates / models the desired skill or process.
		Monitors	Teacher checks to determine if students are progressing toward stated objectives.
		Adjusts Based Upon Monitoring	Teacher changes instruction based on the results of monitoring.
		Establishes Closure	Teacher summarizes and fits into context what has been taught.
		Student Achievement	Effective development and use of modified assessments and curriculum for special education students and other students experiencing difficulties in learning.
10%	Professional Growth and Continuous Improvement	Professional Learning	Uses Professional Growth as a Continuous Improvement Strategy.
		Professional Accountability	Exhibits behaviors and efficiencies associated with professionalism.
5%	Interpersonal Skills	Effective Interpersonal Skills	Effective Interactions and Collaboration with Stakeholders.
5%	Leadership	Professional Involvement & Leadership	Exhibits Positive Leadership through Varied Involvements.

Theoretical Framework: Cotton's Conceptualization of Teacher Effectiveness

One study in particular served as the foundation of the creation of the evaluation rubric that came to be known as the Teacher and Leader Effectiveness evaluation instrument. This primary study culminated in a paper produced by Kathleen Cotton and the Northwest Regional Educational Lab. It analyzed research findings on educational practices that lead to enhanced student outcomes. The purpose of the paper was to identify the core contextual and instructional factors that enable students to learn successfully (Cotton, 2000). This study and the recurring

principles that undergird it shall herein be referred to as Cotton's conceptualization of teacher effectiveness (CCTE).

Kathleen Cotton's "particularly noteworthy" ("Tulsa Public," 2011) contribution to the TLE was a framework that articulated best practices in teaching. The framework was divided into two attributes that Cotton (2000) herself designated as "most crucial": contextual attributes and instructional attributes. From these two attributes, Cotton isolated 10 individual contextual attributes and five instructional attributes, all of which ultimately contributed to the development of the five TLE domains. The attributes included the following:

Contextual Attributes

- Safe and orderly school environment
- Strong administrative leadership
- Primary focus on learning
- Maximizing learning time
- Monitoring student progress
- Academically heterogeneous class assignments
- Flexible in-class groupings
- Small class size
- Supportive classroom climate
- Parent and community involvement

Instructional Attributes

- Careful orientation to lessons
- Clear and focused instruction
- Effective questioning techniques

- Feedback and reinforcement
- Review/reteaching as needed

Cotton's contextual and instructional attributes serve as the foundation of the TLE domains and dimensions, and relevant attributes will be observed in the context of the theoretical framework of this dissertation. Contextual Attributes pertain to "strong administrative leadership and a schoolwide focus on learning," whereas Instructional Attributes pertain to the cycle of development, execution, and revisiting of classroom lessons and instruction (Cotton, 2011). Both types of attributes are grounded in classic school effects research (Brookover and Lezotte, 1979; Edmonds, 1979) and teacher effects (Brophy, 1979; Rosenshine, 1976).

Cotton (2000) acknowledges that teacher responsibilities extend within and beyond the classroom walls, noting the necessity for strong teacher leadership in order for a school to meet the needs of its students, teachers, and administrators effectively. Cotton devotes a portion of her research to acknowledging the value of the maximization of learning time, noting the valuable intersection of this learning time with other climate-enhancing traits such as maintaining a safe and orderly environment. In this context, she articulates a number of practices designed to enrich building-wide climate:

1. Allocating time for various subjects based on school and district goals and utilizing alternative scheduling practices (e.g., block scheduling) to ensure adequate time allocations for core subjects;
2. Establishing and enforcing policies regarding tardiness, absenteeism, and appropriate classroom behavior;
3. Providing learning time and help outside of regular school hours for students who need it;
4. Reviewing potential new programs and activities in terms of their likely impact on learning time;

5. Keeping loudspeaker announcements and other administrative intrusions brief;
6. Ensuring that the school day, classes, and other activities start and end on time;
7. Holding in-service activities as needed to improve staff skills in managing classrooms and increasing student time-on-task (Cotton, 2011).

While some of these items may not be under direct control of classroom teachers, they all speak to the importance of a teacher's proactive involvement in the development and maintenance of a building-wide culture that maximizes learning time. Along these lines, Cotton (2000) asserts the importance of "a warm school climate whose signature feature is a concern for students as individuals." This quality implies teacher influence beyond the classroom door and suggests a more holistic teacher presence throughout the halls of the building.

Monitoring student progress is a dimension vetted by research (Cohen, 1987; Guskey, 1994; Herman, 1992; Krug, 1992; Levine & Lezotte, 1995; O'Conner, 1995), affirmed by Cotton (2000), and included as a dimension within the domain of Instructional Effectiveness on the TLE ("TLE," 2015). Teachers are expected to "[move] throughout the room to assure optimal impact while students are working on guided practice to promote and reinforce students' progress toward the stated objectives." Cotton calls the practice of monitoring student progress a "critical [practice] in [its] own right and a "logical [manifestation] of a primary focus on student learning" ("Teacher Leader," 2015). Cotton (2000) encourages monitoring activities such as:

1. Collecting and reviewing student performance data to ensure early identification and support for students with learning difficulties;
2. Establishing and using procedures for collecting, summarizing, and reporting student achievement information; using aggregated data to determine overall performance and trends; and disaggregating data to review the performance of specific student groups;

3. Reviewing test results, grade reports, attendance records, and other materials to identify problems and taking action based on findings;
4. Reviewing assessment instruments and methods for their suitability to the students being evaluated, and making changes as needed;
5. Making summaries of student performance available to all staff for their use in planning; making periodic reports to parents and community members;
6. Using assessment methods beyond standardized achievement tests to enrich their understanding of students' progress;
7. Aligning classroom assessments of student performance with the written curriculum and actual instruction;
8. Routinely checking students' understanding by conducting recitations, checking students' work during seatwork periods, assigning and checking homework, administering quizzes, and reviewing student performance data.

A natural result of monitoring student progress is that of making adjustments in teaching based upon observations. Review and re-teaching has been shown to be “critically important” (Cotton, 2000) by researchers (Rosenshine, 1995), and they serve as essential Instructional Attributes in Cotton’s battery of teaching practices that matter most. Beyond mere reinforcement by repetition, Cotton (2000) encourages teachers to “use different materials and strategies ... than those used for initial instruction.” Teachers must remain cognizant of methods employed during the initial teaching of a concept or lesson so that they can effectively adjust teaching once problems have been identified. In other words, to effectively adjust based upon monitoring, teachers cannot merely teach on “auto pilot,” but must be engaged in the lesson themselves, just as the students are.

The natural denouement to an effective teaching cycle is that of “closure.” Cotton (2000) does not address the issue of “closure” specifically, but she does, throughout her research,

acknowledge the need for effective teachers to connect what has been learned in the past with what is currently being taught in the classroom for use in future lessons.

On the topic of a teacher's interactions with special education students and students experiencing difficulties in learning, Cotton (2000) has much to say. She takes a proactive stance, recommending "[t]aking note of students' strengths and, when feasible, offering learning activities that capitalize on those strengths." In doing so, a teacher levels the playing field for all students, regardless of their individual learning challenges. Cotton does not endorse the "watering down" of classroom questioning. To the contrary, Cotton recommends "Assuring that faster and slower learners have opportunities to respond to higher-cognitive questions," making modifications in wait-time as necessary. Continuing this concept of handling learning challenges proactively, Cotton suggests "reteaching approaches that capitalize on students' strengths as well as those that address their weaknesses." In more profound learning challenges, Cotton encourages resisting the urge to let students with learning challenges off easy; to the contrary, she endorses challenging those students to the extent that they may be challenged (which she acknowledges is different for different students), providing words of genuine encouragement when students are successful (regardless of where on the cognitive continuum "success" resides for each student), and modifying and reteaching lessons whenever necessary. On a more individual level, Cotton recommends "making adaptations to serve the needs of faster and slower learners," and further, "[d]elivering [varied] high-quality instruction to students in lower-ability groups" (Cotton, 2000).

While Cotton's research informs each individual dimension of the Tulsa TLE, when examined as a whole, her postulates create a synergy that extends beyond a list of domains and dimensions. In the context of Cotton's impact on the TLE, Aristotle's oft-repeated observation holds true: the sum of the whole is greater than its parts.

CCTE, the TLE, and Today's Students

Cotton's research may be keyed to the Oklahoma TLE both vertically, through consideration of Cotton's Instructional Attributes and her Contextual Attributes, and horizontally, based on TLE dimensions (see Table 2 below). While Cotton's research supports the Oklahoma TLE that is used to assess Oklahoma secondary school teachers, the Oklahoma TLE has not been evaluated by students. Research garnering student sentiment about the TLE may fill a gap in the research pertaining to student sentiment about the effectiveness of the evaluation instrument in capturing teaching attributes that are significant to a teacher's primary client. CCTE serves as a foundation for the TLE, and the TLE serves as a statistically reliable and valid instrument for assessing teacher performance. Additionally, it has been seen that student responses as a data source for teacher evaluation can be reliable and valid; therefore, it would be expedient to gather student feedback on teacher performance utilizing an instrument that speaks to experiences that are relevant to and explicitly observable by students in a secondary classroom and that embodies the proven principles of the TLE, which are firmly founded on Cotton's conceptualization of teacher effectiveness.

Table 2

Cotton's Attributes Keyed to the TLE

	Contextual Attributes		Instructional Attributes	
	Cotton	TLE	Cotton	TLE
Instructional Effectiveness (50% of TLE)			Careful orientation to lessons	Literacy
			Careful orientation to lessons	Involves all learners
			Effective questioning techniques	Explains content
			Clear and focused instruction	Models
	Monitoring student progress		Clear and focused instruction	Monitors
	Monitoring student progress			Adjusts based on monitoring
N/A		N/A	Establishes closure	
		Clear and focused instruction	Student achievement	
Classroom Management (30% of TLE)	Safe and orderly environment	Preparation		
	Safe and orderly environment	Discipline		
	Safe and orderly environment	Building-wide climate		
	Maximizing learning time	Lesson plans		
	Supportive classroom climate	Student relations		

Table 2, cont'd

Cotton's Attributes keyed to the TLE

	Contextual Attributes		Instructional Attributes	
	Cotton	TLE	Cotton	TLE
Interpersonal Skills 5% of TLE	Supportive classroom environment	Effective interpersonal skills		
Leadership 5% of TLE	Supportive classroom environment	Professional involvement and leadership		

CHAPTER III

METHODOLOGY

This quantitative study was designed to capture student sentiment about teacher effectiveness qualities assessed through the TLE and to identify any other teacher characteristics, as perceived by students, that are essential to teaching efficacy. Additionally, it is intended to observe similarities and differences between students and administrators regarding teacher effectiveness traits.

Survey data was used to gain a better understanding of student attitudes about 18 of the 20 dimensions measured on the TLE rubric – dimensions that have evolved from CCTE. Data from an exploration survey (“Survey #1”) will inform a modified survey (“Survey #2”) which students used to evaluate a videotaped teaching sample. Student responses to both the original survey and the videotaped teacher evaluation were compared using descriptive statistics. Further, the relationship between student opinions about specific teacher effectiveness characteristics and administrator opinions about the same teacher effectiveness characteristics were observed and analyzed using descriptive statistics.

Research Design

At the outset of this study, a modified list of the unique teacher effectiveness characteristics that are assessed using the TLE rubric was generated. The characteristics came from 18 of the 20 dimensions articulated in the TLE Observation and Evaluation Rubric for Teachers. The two dimensions that are components of the Professional Growth & Continuous Improvement domains were not used, as these dimensions pertain to continuing education activity

– activity to which most students are likely not privy. The following dimensions will be analyzed:

1. Preparation
2. Discipline
3. Building-Wide Climate Responsibility
4. Lesson Plans
5. Assessment Practices
6. Student Relations
7. Literacy
8. Adherence to Current State Standards
9. Involvement of All Learners
10. Explaining of Content
11. Clear Instruction & Directions
12. Modeling of Principles Taught
13. Monitoring of Classroom
14. Inclination to Make Adjustments Based upon Monitoring
15. Establishment of Closure
16. Student Achievement
17. Effective Interpersonal Skills
18. Professional Involvement & Leadership (“Teacher Leader,” 2015)

These 18 dimensions were summarized in an assessment instrument (the aforementioned survey) using simple sentences and language that is readily accessible to secondary students. Each summary was followed with a 5-point numeric Likert response set. Selection of 1 from the response set indicates that students feel that the summarized trait is not a true sign of teacher effectiveness whatsoever, and selection of 5 on the scale indicates that students believe wholeheartedly that the trait in question is indeed a defining characteristic of teacher effectiveness.

The survey was printed and distributed to participating secondary students during their English class. Students were asked to read each TLE-based statement about teacher effectiveness and then rate the strength of their convictions about the importance of each trait when considering teaching effectiveness using the 5-point scale. Survey responses were completely anonymous.

Once the participants completed the surveys, the surveys were collected, and scores on each of the 18 survey items were tallied and then analyzed. Mean scores were calculated for each characteristic listed by tabulating the number of times a certain characteristic is listed divided by the total number (N) of participants. This process was utilized to determine teacher characteristics that are notably important to students that appear on the TLE. Data was entered into an Excel Spreadsheet for analysis. Analysis consisted of an examination of the descriptive statistics of the survey to determine which attributes from this collection of data emerged as key indicators (as indicated with higher mean scores) of teacher effectiveness. Individual effectiveness traits that had an overall mean score of 3.5 or greater were identified and transcribed onto a second survey.

In addition to being given the opportunity to respond to the 18 Likert-scale ratings, students were also invited to *hand-write* any teacher effectiveness traits that they felt were true indicators of effectiveness but were not mentioned in any of the 18 survey questions at the bottom of the survey. The data was analyzed to identify alternate descriptions that pertain to similar

teacher characteristics because of the likelihood that students would not use entirely consistent wording to describe teacher characteristics they deem to be important. For example, it is probable that responses of “helps me learn from my mistakes” and “gives me a second chance” describe a similar teacher characteristic. Handwritten responses were summarized and transcribed onto the new survey (Survey #2) along with the individual effectiveness traits that had an overall mean score of 3.5 from Survey #1.

Following student ranking of effective teacher characteristics, a survey was developed utilizing the most important characteristics identified by students in the initial survey in both the objective and subjective portions of the survey. Students who participated in the initial survey were then asked to utilize the scale created from the results of the initial survey to evaluate an online teaching simulation. As with the first survey, students utilized a 5-point Likert numeric response set, rating the videotaped teaching experience based on their perceptions of the effectiveness of the videotaped teacher per the traits listed in Survey #2. The selection of “1” from the second survey indicated that the student felt the videotaped teacher was not effective at all in the discipline assessed in a particular question, and the selection of “5” from the second survey indicated that the student felt the teacher excelled in the discipline assessed in the particular question.

School administrators were asked to respond in the same way to the same survey (Survey #2) using the same list of effectiveness characteristics. Student and administrator responses to the second survey were observed for similarities and differences in perceptions about teacher effectiveness and were analyzed using descriptive statistics.

This information is important because it provides further understanding of whether there are commonalities between student perceptions of teacher effectiveness and principal perceptions

of teacher effectiveness on the domains of the TLE that students perceive to be important for teaching effectiveness.

Research Population

The research population for this study consisted exclusively of 9th grade students from a rural public high school in Oklahoma. Surveys were distributed to all students in their English classes in this selected high school. All students were sampled to minimize sampling error. English classes were chosen for this study because all students at the selected high school are required to take English class during their 9th grade year. Choosing all English class students should help to minimize sampling error and encourage a large enough participant pool to perform data analyses. The goal in this study was to evaluate the teacher effectiveness sentiment (especially as it pertains to principles presented in the TLE) of high school students.

The surveys were distributed to more than 90 students in the selected school. Survey #1 contained effectiveness principles corresponding to all but two items on the TLE rubric, resulting in 18 model parameters total (out of 20 total TLE rubric items) on Survey 1.

Instrumentation and Researcher Position in the Study

The researcher is familiar with the context of teacher evaluation, having been a secondary teacher at a public school for eight years. During this eight-year tenure, student voice was not solicited for the purposes of formal teacher evaluation. Nevertheless, the researcher observed anecdotally that students:

1. had opinions about the effectiveness of their teachers;
2. had unique insight into the specific teaching methods of their teachers; and,
3. were able to formulate and articulate unique and concrete opinions about their teachers' effectiveness.

In addition to involvement in public education, the researcher has also been employed in corporate environments and has been able to observe “evaluations” as a working member of both private and nonprofit sectors during those years. This professional experience provided insight into the disparity between formal employee evaluations in public schools and those conducted in a corporate context.

That being said, the selection of quantitative research and the utilization of surveys as the primary data collection instrument in this study provides a level of detachment that effectively separates the researcher from the researched (“Quantitative and Qualitative,” 2009) and helps implement a measure of objectivity.

Research Procedures

Before research was conducted, arrangements were made with the principal of the participating school district’s high school to administer a two-phase survey to students at the school. With administrator approval, the researcher asked students to voluntarily participate in the study. The invitation to participate took place initially via a letter distributed to students in their English classes. Students under the age of 18 were asked to take the letter home to parents so that parents could provide informed consent for student participation. Students were also asked to sign the consent document.

Before the survey was conducted, the researcher met with school administrators to explain the research concept, process, value, results, and privacy procedures. The researcher also met (virtually) with the participating English classes to explain the study and answer any questions that students may have had about the study.

Sample Selection

Three weeks prior to the administration of the first survey, 9th grade English teachers were asked to distribute to their students a letter that described the study and each student's potential role in the study. The letter included a consent form that parents and students were asked to complete. Students who did not complete consent forms within a week of the intended survey date were not permitted to participate in the study.

Phase One – Student Survey

A week before the administration of the survey, the researcher collected information about the numbers of participating students for each English class. Surveys were counted out and then placed in manila envelopes and labeled according to class hour. The day before the administration of the survey, the envelopes containing the counted surveys were delivered to the school and given to the English teachers.

On the day the survey took place, the English teachers were asked to allot 10 minutes for students to distribute, complete, and return the initial surveys. Students took the survey and then placed the completed surveys back into the envelope in which they arrived. The teacher sealed the envelope and, at the end of the school day, provided all (sealed) envelopes to the school principal, who then handed the envelopes containing the completed surveys to the researcher.

The researcher then organized all surveys for analysis. First, survey responses were entered into an Excel spreadsheet. Additionally, all hand-written responses were reviewed to determine if there were additional teacher characteristics that were important to students. To analyze these results, general categories of student responses were generated by the researcher as common themes emerged from the data. The additional characteristics were utilized as a component of Survey #2.

Phase Two – Student and Administrator Evaluation of Online Teaching Sample

A second survey based on two criteria gleaned from the first survey was created by the researcher. The two criteria for inclusion on Survey #2 were:

1. TLE survey items from the first survey that had a mean score of 3.5 or higher.
2. Additional “write-in” characteristics identified by students as important for teacher effectiveness.

The second survey contained the total number of items that captured the two criteria above. As with Survey #1, each survey item included in Survey #2 was measured using a 5-point Likert scale. The wording of each question was modified in a way that enabled participants to evaluate an online lesson using the survey content that captures the two criteria above while still retaining the basic skill or discipline found in the corresponding TLE dimension.

The day before the administration of the second survey, the envelopes containing the counted surveys were delivered to the school and given to the English teachers. The second surveys were then distributed to the same students who completed the first round of surveys. Additionally, the surveys were distributed to evaluating administrators at the school.

Participating students and administrators were asked to view 30 minutes of an online lesson video. The video came from the library of teaching videos available free to the public at teachingchannel.com. Students and administrators were asked to evaluate the teacher in the video using the characteristics in the revised survey.

Upon viewing the online lesson, students completed the evaluation surveys and then placed the completed surveys back into the envelope in which they arrived. The teacher sealed the envelope and, at the end of the school day, provided all (sealed) envelopes to the school principal, who then handed envelopes containing the completed evaluation surveys to the researcher.

Data Analysis

Study data were analyzed in several ways. First, descriptive statistics about student responses regarding TLE criteria that are assessed in the first survey were analyzed using Microsoft Excel. Mean scores were generated for each of these items. Items with a mean score of 3.5 or higher were retained for development of the second survey. Student responses to the open-ended question on the first survey were reviewed, and categories were developed to organize student responses that are similar (yet not worded exactly the same). These items, along with the items with a mean score of 3.5 or above (mentioned previously) were entered onto Survey 2.

The second survey was handled similarly to the first survey. Data from survey responses from both students *and administrators* were logged in Microsoft Excel. Analysis consisted of an observation of the descriptive statistics from the second survey.

Assumptions

The following assumptions were made during this study:

- Data pertaining to both administrator and student evaluation of participating teachers were collected and measured without error.
- Level one errors are independent and normally distributed with a common variance.
- Residuals are uncorrelated and have constant variance.
- Data collected across students are independent.

Limitations

One limitation in this study pertains to the initial phase of data collection. Students were asked to respond to a survey in which 18 teaching characteristics were presented. These 18 principles were directly connected to the TLE rubric. Students were led to respond directly to these specific traits rather than being given latitude to report any trait they feel is indicative of

effective teaching, free of the influence of instruments and rubrics that may have been created beforehand. While students were given the opportunity to respond subjectively at the bottom of the initial survey, the inclusion of the 18 traits may have suggested the presence of an exhaustive list of effective teaching characteristics and may have discouraged students from responding candidly (or at all) in the subjective portion. However, this decision was actually a delimitation because these parameters were established by the researcher. This decision was guided by the purpose of observing similarities and differences between administrator responses to the second survey and student responses. Therefore, designing the study in this way provided an opportunity to investigate descriptive statistics between student and administrator perceptions of teaching effectiveness.

Closely related: While the study purposed to examine student sentiment about characteristics of effective teachers, it examined these characteristics through the lens of a single evaluation instrument: the TLE. Many other evaluation instruments for teacher evaluation exist, and there are more ways to assess student sentiment about characteristics of effective teachers than by surveying them utilizing one single, specific evaluation instrument.

Along these lines, a third limitation of this study may be observed in the creation of the initial survey students were asked to take. In order to provide students with a survey that contained accessible language and meaningful principles, the 18 TLE traits about which students were asked in the initial survey were articulated differently than they were on the TLE rubric itself. This “translation” of teacher effectiveness traits may have been considered to be subjective, and some question may arise as to the complete and unaltered embodiment of the originally intended TLE principle in the student survey questions. Additionally, with regard to the subjective (“write-in”) portion of the first survey, some interpretation by the researcher was necessary for category aggregation.

One significant limitation in this study pertained to the sample size of administrators surveyed, of which there were three. Not only was it impossible to draw conclusions that were statistically significant from such a small group, but it was also impossible to reliably conduct certain statistical tests that examine variances and means of groups when one population is so small. Such tests include the correlation analysis, ANOVA, and the independent T-test. These tests would have provided greater insight into both student and administrator responses to each survey and into the relationships between student and administrator sentiment. Because the administrator group was so small, the researcher was limited only to a simple study of descriptive statistics. While these statistics are indeed telling, they are also somewhat limited in scope.

Another limitation in the study had to do with the limited scope and experience of the selected student survey population. The opinions gathered in the study accounted for a very small collection of secondary students: 9th grade English students from a single school district. The view provided by this study is a proverbial scratch in the surface of secondary student sentiment in general. To this point, it should also be acknowledged that the maturity level of student participants, especially in the face of such multifaceted traits as those presented in the TLE and the study surveys, may also be limited compared to sophomores, juniors, and seniors.

A final limitation of this study was that it was possible that those students with more involved families returned consent forms and students with less involved families did not, even if the student was capable of and interested in taking part in the study. This limitation could be important because findings in research indicate that middle class families often demonstrate higher levels of involvement in schools than those families experiencing poverty (Lareau, 2011). If student participants are heavily skewed toward one demographic group, this factor could pose a limitation on the study. Effort was made to encourage participation by all English class students, and adequate time was given for families to consider the study, ask questions, and return the consent form.

Summary

This study was designed to analyze student sentiment about 18 out of 20 of the teacher effectiveness qualities measured using the TLE instrument. To accomplish this, a sample of 9th grade English class students were asked to respond to a survey wherein they assigned a ranking to each of the 18 items in the TLE based on how strongly they felt the item was truly indicative of teacher effectiveness. Students also wrote in additional qualities on the survey. These responses were collected, and traits with mean scores of 3.5 or higher on a 5-point Likert scale were then placed on a second, follow-up survey along with “write-in” qualities.

The second survey, which surveyed only traits with mean scores of 3.5 or higher from the first survey and the “write-in” qualities from the first survey, were given to students and evaluating administrators at the same school. Students and administrators were asked to watch an online teaching simulation video and then rank the effectiveness of the teacher in the video using this new survey.

Descriptive statistics from both student and administrator responses from the second survey were analyzed. Because the sample size of administrators was too small to provide statistically significant or reliable results, descriptive statistics (as opposed to a correlation analysis) from the second survey were analyzed.

CHAPTER IV

FINDINGS

Before delving into the data gleaned from the two surveys administered at the participating high school, it is necessary to be reminded of the three Research Questions that prompted their administration of the surveys. The following three Research Questions were posed at the beginning of this dissertation:

- What teacher characteristics do students identify as being necessary for effective teaching?
- To what extent do these student-identified characteristics align with characteristics outlined on the TLE?
- Are there similarities in student and administrator sentiment about effective teaching practices?

The purpose of this study was to gain a better understanding of student sentiment about characteristics of effective teachers for the purposes of informing a survey instrument that captures student perspectives. Additionally, this study compared student identified characteristics with indicators of teacher effectiveness identified in the TLE instrument. Finally, student sentiment about effective teaching practices was considered in tandem with administrator sentiment about the same practices.

Data Supporting Research Questions 1 and 2

Results from the first survey administered in this study pertain to Research Questions One and Two.

- What teacher characteristics do students identify as being necessary for effective teaching?
- To what extent do these student-identified characteristics align with characteristics outlined on the TLE?

Ninety students participated in the survey ($N = 90$), and the possible range of responses ranged from 1 (not at all important) to 5 (extremely important). The survey contained 18 items, and each item was aligned with characteristics established on the TLE as important for teaching effectiveness. Means were calculated for each item, and responses are reported in Table 3.

Table 3

Student Response Means for Survey #1

Survey Question	Student Response (M)
Question 1	4.16
Question 2	3.86
Question 3	3.80
Question 4	3.70
Question 5	4.60
Question 6	3.34
Question 7	2.43
Question 8	3.76
Question 9	4.11
Question 10	4.36
Question 11	4.80
Question 12	4.38
Question 13	3.92
Question 14	4.42
Question 15	3.50

Question 16	3.91
Question 17	3.42
Question 18	3.06

The first survey contained 18 Likert-type questions that permitted participants to indicate the degree to which they agreed that a teaching trait described in the question was important to teaching in general. Research participants selected a single value on a 5-point Likert-type response set to indicate how strongly they felt the stated characteristic was essential to effective teaching. Results from the first survey focused on the mean student responses for each of the 18 questions. Mean scores for all questions are reported in Table 3.

Of the 18 questions posed to students, four items generated means lower than 3.5 (highlighted in Table 3). These questions were:

6. A teacher should respectfully express high expectations. ($M = 3.34$)
7. A teacher should emphasize reading and writing in every lesson, even if the lesson is not about reading and writing. ($M = 2.43$)
17. A teacher should communicate with me, my family, and other educators. ($M = 3.42$)
18. A teacher should volunteer to work at school or district events and should help other teachers grow in their professional skills. ($M = 3.06$)

Because of lower mean scores, which indicated limited value by student participants, these four questions were not included in the second survey. However, four “write-in” questions were added to the second survey. These questions were developed after analysis of student open ended responses to the first survey where students were asked to indicate additional characteristics, not listed in the initial 18 survey questions, that they perceived were important for teaching and

learning. These responses were adapted contextually to reference the teaching video as explained later in this chapter. Those questions were:

15. *The teacher in the video presented herself professionally both in dress and in what she said.* One student out of the total sample of 90 indicated that this characteristic is important for successful teaching, representing a 1% write-in response rate.

16. *The teacher in the video showed that she had mastered the material she was teaching the students.* One student out of the total sample of 90 indicated that this characteristic is important for successful teaching, representing a 1% write-in response rate.

17. *The teacher in the video translated classroom lessons into practical, real-life lessons when possible.* Two students out of the total sample of 90 indicated that this characteristic is important for successful teaching, representing a 2% write in response rate.

18. *The teacher in the video expressed individual empathy toward and an individual understanding of her students.* Nine students out of the total sample of 90 indicated that this characteristic is important for successful teaching, representing a 10% write in response rate.

Data Supporting Research Question 3

Results from the second survey administered in this study pertain to Research Question 3:

- Are there similarities in student and administrator sentiment about effective teaching practices?

The second survey that was administered to both student and administrator participants followed the viewing of a 20-minute online video clip of a lesson being taught by an anonymous

teacher. Neither students nor administrators knew the teacher in the video clip, nor was the name of the district or school in which the video was produced disclosed. To increase student participant familiarity with the content of the video lesson, a video was selected that included a lesson about Shakespeare's play, *Romeo and Juliet*, as this work had been taught to the students at the school earlier in the semester.

The second survey, which was administered to both students and participating administrators, contained 18 Likert-type questions that permitted survey-takers to indicate the degree to which they felt the value described in the question was reflected by the teacher in the video lesson. Research participants selected a single value on a five-point Likert-type response set to indicate how strongly they felt the teacher in the recording embodied the principle described in each question or how vividly participants observed the principle being presented during the recorded lesson. Ninety-seven students participated in the second survey ($N = 97$), and three administrators participated in the second survey ($N = 3$). The possible range of responses ranged from 1 (not at all important) to 5 (extremely important). Each of the 18 questions on the second survey evolved from responses to the first survey in that:

1. The questions appeared in the initial survey and received a score of 3.5 or higher by student participants; or,
2. Student participants wrote in a new characteristic at the bottom of the first survey that was unique and not already articulated in the initial survey.

As evidenced by student responses on the first survey, the underlying principles in each of the 18 questions in the second survey were those that student participants felt most strongly indicated effective teaching.

Mean scores for all questions for both students and administrators are shown in Table 4. Additionally, Table 4 indicates the difference in mean scores (expressed as an absolute value)

between students and administrator responses. The standard deviation among student responses and among administrator responses is also provided. Questions adapted from the “write-in” responses from the first survey are highlighted.

Table 4

Student and Administrator Descriptive Statistics for Survey #2

Survey Question	Student Response (<i>M</i>)	Administrator Response (<i>M</i>)	Difference (<i>M</i>)	Student Responses (<i>SD</i>)	Administrator Responses (<i>SD</i>)
Question 1	4.03	5.00	0.97	1.07	0.00
Question 2	4.31	4.67	0.35	0.83	0.58
Question 3	3.67	4.33	0.66	1.09	0.58
Question 4	4.11	4.33	0.23	1.00	0.58
Question 5	4.20	4.33	0.14	1.01	0.58
Question 6	3.68	4.33	0.65	1.12	0.58
Question 7	4.41	5.00	0.59	0.96	0.00
Question 8	4.26	4.67	0.41	0.93	0.58
Question 9	4.04	4.67	0.63	1.03	0.58
Question 10	3.96	4.00	0.04	1.00	0.00
Question 11	4.52	5.00	0.48	0.82	0.00
Question 12	3.54	4.33	0.80	1.11	0.58
Question 13	3.41	4.00	0.59	1.09	1.00
Question 14	3.50	3.00	0.50	1.20	0.00
Question 15	3.94	5.00	1.06	0.98	0.00
Question 16	4.08	4.67	0.58	0.85	0.58
Question 17	3.52	4.67	1.15	1.29	0.58
Question 18	3.87	4.67	0.80	1.13	0.58

Examination of the Data

Below is a detailed examination of the descriptive statistics for each of the 18 questions posed to both students and administrators following their viewing of the 30-minute lesson about *Romeo and Juliet*. To provide some context ahead of the reporting of individual question means, it is noted that the average student mean of all questions combined was 3.95, and the average

administrator mean across all questions was 4.48. The cumulative difference in means between students and administrators was 0.53

Question 1: The teacher in the video was prepared for class.

The mean score for students for this question was 4.03 out of 5. The mean score for administrators for this question was 5. Administrators were unanimous in sentiment about the preparedness of the teacher in the video, each of them indicating that the teacher seemed as prepared as she could be.

Question 2: The teacher in the video monitored and managed the behavior of students.

The mean score for students for this question was 4.31 out of 5. The mean score for administrators for this question was 4.67. Again, both students and teachers seemed to feel strongly that the teacher in the video exemplified the quality in question. Mean scores indicate that both groups were in close agreement that the teacher monitored and managed the behavior of students appropriately.

Question 3: The teacher in the video promoted safety, health, and orderliness.

The mean score for students on question 3 was 3.67 – one of the lower scores in the survey results, and notably below average. The mean score for administrators was 4.33, also below the comprehensive average of administrator scores. That being said, the standard deviation for students on this question was high ($SD = 1.087$) compared to the standard deviations of other student means. Administrator standard deviation was .58. While administrators and students may have felt that the teacher did not expressly promote safety, health, and orderliness in the video in general, there does seem to be disconnect among both students and administrators on the topic of the promotion of safety, health, and orderliness.

Question 4: The teacher in the video created meaningful lesson plans.

The mean score for question 4 was 4.11 for students and 4.33 for administrators. Deviation among students was 1.001. Among administrators, it was .58.

Question 5: The teacher in the video evaluated students fairly and in a way that truly evaluated their growing knowledge and learning.

The mean score for question 5 was 4.20 for students and 4.33 for administrators. Mean scores between students and administrators were very close compared to mean scores of the rest of the questions on the survey – only one pair of means was closer. Student standard deviation was 1.007. Administrator standard deviation was .58.

Question 6: The teacher in the video taught about things that would help students meet state standards for student education.

The student mean on this question of 3.68 was notably below the comprehensive student mean of all questions on the survey ($M = 3.95$); however, that may have been expected, as students simply may not have knowledge of the state standards for student education. The standard deviation of student responses was the second highest of all student standard deviations. The standard deviation among administrators, while not particularly informative due to the small panel of administrators surveyed, rested in the middle of the range of administrator standard deviations – a range that spanned 0 to 1. The administrator mean on this question was 4.33.

Question 7: The teacher in the video taught “actively,” asking questions and encouraging students to act on what they learn.

The mean scores on this question among both students ($M = 4.41$) and administrators ($M = 5.00$) were notably high compared to mean scores on other questions. Indeed, the mean score on question 7 was the second highest of all student means on the survey, and it was the highest mean (and the highest possible mean) of all administrator responses. (Note: Administrator means

were 5.00 on four different questions.) Additionally, both student and administrator variance on this question was minimal ($SD = 0.958$ for students and $SD = 0.00$ for administrators) suggesting that students as well as administrators seemed to be in agreement within groups that the teacher taught “actively.”

Question 8: The teacher in the video used many methods to teach (e.g., lecture, group discussion, hands-on activities, etc.).

The mean score for question 8 was 4.26 for students and 4.67 for administrators. Deviation among students was .927. Among administrators, it was .58.

Question 9: The teacher in the video gave clear instructions and directions to students.

The mean score for question 8 was 4.04 for students and 4.67 for administrators. Deviation among students was 1.031. Among administrators, it was .58.

Question 10: The teacher in the video modeled or demonstrated what she wanted students to learn to make sure students got the lesson.

Mean scores for both students and administrators were relatively low, students’ being 3.96 and administrators’ being 4.00. Interestingly, student and administrator sentiment were closest on this question than any other question on the survey, with a difference in means of only .04. Administrators were unanimous in their means, resulting in a standard deviation of 0. Student standard deviation for this question was 0.999.

Question 11: The teacher in the video moved around the room and interacted with students to be sure they understood what they were supposed to do on assignments and lessons.

In contrast to Question 10, both students and administrators scored the teacher in the video high on Question 11. The student mean of 4.52 was the highest of all student means in the

survey. The administrator mean was the highest possible, with a mean of 5.00. Students and administrator means differed by only .48, one of the most minimal differences in student/administrator means in the survey results. Student sentiment was most consistent on this question with the lowest student standard deviation on this survey of 0.818. The administrator mean of 5.0 resulted in a standard deviation of 0.

Question 12: The teacher in the video adjusted instructional plans if she realized that students were not learning.

The mean score for question 12 was 3.54 for students and 4.33 for administrators. Deviation among students was 1.109. Among administrators, it was .58.

Question 13: The teacher in the video summarized what students learned in class that day.

On average, student sentiment seemed to be lowest on Question 13, as the student mean for the question was 3.41. The administrator mean of 4.00 was also below the group's average mean of 4.48. Differences in perception among the student group was relatively high with a standard deviation of 1.092. Administrators also seemed to be quite divided on this question. Their standard deviation of 1.00 was the highest of all administrator standard deviations.

Question 14: The teacher in the video modified lessons and tests for students who had a hard time learning.

Responses to Question 14 were unique in that they represented the only time on the entire survey that the student mean was higher than the administrator mean. Despite this finding, the student mean was a relatively low 3.50. Administrators scored the teacher the lowest of all questions on Question 14 with a mean of only 3.00. Administrators were unanimous on this low score with a standard deviation of 0. Not only was the score low, but the sentiment seemed to be

identical among all administrator respondents. Student standard deviation was 1.199, the second highest of all student standard deviations.

Question 15: The teacher in the video presented herself professionally both in dress and in what she said.

Students and administrators disagreed somewhat on the presence of the trait identified in Question 15. The student mean of 3.94 was slightly below average. Administrators on the other hand were unanimous ($SD = 0$) in their responses with a the highest mean possible of 5.00. The difference between the student mean and the administrator mean on this question was 1.06, the second-highest difference in student/administrator means in the survey.

Question 16: The teacher in the video showed that she had mastered the material she was teaching the students.

The mean score for question 16 was 4.08 for students and 4.67 for administrators. Deviation among students was .850. Among administrators, it was .58.

Question 17: The teacher in the video translated classroom lessons into practical, real-life lessons when possible.

Responses from students and administrators were the most divided on Question 17 of all questions on the survey. The student mean was a comparatively low 3.52 – a mere 2 hundredths of a point from the lowest student mean in the second survey. In contrast, the administrator mean of 4.67 was high in comparison to the comprehensive student and administrator means. The difference between the student mean and the administrator mean on this question was a full 1.15. It is worth noting that, while students scored the teacher in the video the lowest on average on this trait, students also experienced the greatest amount of inconsistency in scoring on this question

with a standard deviation of 1.287. Standard deviation for administrators on this question was 0.58 – a standard deviation that appears in administrator results commonly in this survey.

Question 18: The teacher in the video expressed individual empathy toward and an individual understanding of her students.

The mean score for question 18 was 3.87 for students and 4.67 for administrators. Deviation among students was 1.133. Among administrators, it was .58.

Comprehensive Results

Among all questions, student means ranged from 3.50 to 4.52. Administrator means ranged from 3.00 to 5.00; however, the reader should be reminded of the very low administrator survey population of 3 participants. On no survey response was the student mean identical to the administrator mean, but on several responses, mean comparisons between student and administrator were quite close, with the slimmest difference between student mean and administrator mean being 0.04 on question #10, “The teacher in the video modeled or demonstrated what she wanted students to learn to make sure students got the lesson.” Differences in student and administrator means were as high as 1.15. Among all questions, student standard deviations ranged from 0.818-1.287. Administrator standard deviations ranged between 0.00 and 1.00.

CHAPTER V

CONCLUSIONS

The purpose of this study was to gain a better understanding of student sentiment about characteristics of effective teachers for the purposes of informing a survey instrument that captures student perspectives. Additionally, this study compared student identified characteristics with indicators of teacher effectiveness identified in the TLE instrument. Finally, student sentiment about effective teaching practices were considered in tandem with administrator sentiment about the same practices.

The results of this study provided descriptive statistics that demonstrated the strength of sentiment the student participants felt about 18 of the 20 effective teacher traits presented in the TLE instrument as well as an analysis of the observation of the strongest of student sentiments as viewed during a brief teaching video. Student survey results were compared to survey results of a small group of administrators who were given an identical survey. This portion of the study served to provide a simple, although not statistically significant, comparison between students and administrators on the 18 survey traits.

This discussion of the study results will provide an opportunity to summarize the key findings; provide reflection on potential explanation of the key findings; make suggestions for further research; and discuss the potential implications for policy, scholarly research, and school practice. While the responses to every question on both surveys are unique and impactful to varying degrees, only those results that speak most poignantly to the purposes described above will be discussed.

Summary of Key Findings

The results of the empirical analysis produced some interesting findings. First, it showed that student sentiment about four concepts assessed by the TLE was quite weak; so weak that the four items failed to receive a high enough score to be included on the second survey that was administered to both student and administrator participants. To summarize, these concepts focused on issues of a teacher's expression of high expectations of his/her students; the emphasis of reading and writing in all disciplines regardless of the lesson's focus on reading and writing; regular communication with students, their families, and other educators; and the inclination for a teacher to lead in school-sponsored events. Additionally, students advanced four unique concepts that they felt were not assessed by the TLE instrument. To summarize, these concepts included the need for professionalism in speech and appearance; mastery of material taught; practical application of classroom lessons; and empathy toward students. The second survey revealed that student participant groups and administrator participant groups differed both among and within groups as to the extent to which the teacher in the sample teaching video embodied the traits assessed by the second survey.

Findings

Key findings of this exploratory investigation are organized by research question. Much of the explanatory narrative of this chapter represents plausible speculation.

RQ1 and RQ2: *What teacher characteristics do students identify as being necessary for effective teaching? To what extent do these student-identified characteristics align with characteristics outlined on the TLE?*

Data from Survey 1 suggest the strongest student sentiment about teacher provision of “clear instructions and directions to students” ($M = 4.80$), a sentiment that was assessed in Question 11 of the first survey. The mean of responses to this question was notably higher than

the average of means for Survey 1 ($M = 3.82$). The straightforward concept of a desire for clear instruction indicates the importance to students of solid pedagogy at the foundational stage of lesson execution. This finding suggests that students realize that, without clear instruction, the remainder of the lesson has the potential to go awry.

The second-highest mean on Survey 1 was in response to Question 5: “A teacher should grade things fairly and in a way that truly evaluates a student’s growing knowledge and learning” ($M = 4.60$). This sentiment corresponds with literature pertaining to the value of fairness and justice in secondary education settings (Berti, et al., 2016; Wendor and Alexander, 2005). Interestingly, the two strongest sentiments expressed by survey participants bookend the teaching process, the strongest sentiment occurring at the beginning of the lesson (instruction) and the next strongest occurring at the end of the lesson (assessment). Responses to Question 5 exemplify the importance of fairness in education that students seek as well as the premium placed on the accuracy and legitimacy of assessment with regard to a lesson’s intended purpose. This finding suggests that, if an evaluation does not clearly assess the articulated purpose of the lesson, teachers may see frustration in their students.

On the other end of the range of Survey 1 means are the four questions whose means disqualified them from appearing on Survey 2:

6. A teacher should respectfully express high expectations. ($M = 3.34$)
7. A teacher should emphasize reading and writing in every lesson, even if the lesson is not about reading and writing. ($M = 2.43$)
17. A teacher should communicate with me, my family, and other educators. ($M = 3.42$)

18. A teacher should volunteer to work at school or district events and should help other teachers grow in their professional skills. ($M = 3.06$)

Questions 17 and 18 may become more important to students in later years, and follow-up with research participants in 5 or 10 years after this initial survey may prove to be an interesting contribution to the literature. Indeed, Question 17 may result in consternation for struggling students, students who may not fully apply themselves educationally, or students who experience high academic expectations from family members. Students simply may not see, understand, and appreciate the value of Question 18.

It is unclear whether the relatively low mean for Question 6 is due to weaker sentiment among students about the level of respect extended to students from teachers, the intensity of academic expectations teachers have of their students, or some combination thereof. Additional research that polls students using the same question but excluding the modifier, “respectfully” may show different results.

The question with the lowest student mean on Survey 1 was Question 7: “A teacher should emphasize reading and writing in every lesson, even if the lesson is not about reading and writing” ($M = 2.43$). This response may make one wonder if students see a disconnect between the disciplines of reading and writing and other school subjects.

Just as significant as student sentiment about items that are addressed by the TLE is student sentiment about concepts that are *not* explicitly assessed by the TLE. Four concepts that were presented by students in the write-in portion of the first survey were included in the second survey. These concepts pertained to professional dress and presentation, teacher mastery of material, practical application of lessons, and teacher empathy.

One behavior mentioned in the write-in portion of Survey 1 was that of “maintain[ing] professionalism in dress and speech.” To look ahead to RQ3, students seemed to feel that the

teacher in the sample video did not necessarily maintain this trait, while administrators felt strongly that the teacher in the sample video did adhere to this virtue. That being said, there was notable disagreement among the student group ($SD = 1.199$) as to the presence of this particular trait in the teaching video.

One surprising realization in this study came not from student or administrator survey responses, but simply from the TLE instrument itself. That realization was that the TLE instrument does not directly assess *mastery of subject material* as presented in the classroom. One student articulated his/her concern as such:

“A teacher should always check the work for the students before assigning it. *A teacher should only assign work that they have answers to*” (italics added).

It would seem natural to assume that a teacher had, indeed, mastered to a notable extent the material about which he or she was teaching. However, it does seem to be a concern among students. Future iterations of the TLE or alternative evaluation tools may consider adding a question that assesses this trait.

Student desire to learn disciplines that apply to “real life” was presented in the write-in answers from Survey 1. One student wrote, “We should be learning how to do taxes and property; not SSS postulate.” Another wrote, “Teachers should give examples of real-world situations as to give a lesson more meaning, and to engage students.” Student frustration about a perceived lack of practical application of lessons taught in school was apparent.

The most common sentiment underlying write-in answers from Survey 1 had to do with *teacher empathy*. This sentiment was presented in a variety of student statements:

- “Being able to communicate with students on a personal level.”

- “Teachers should sometimes have more lenient policies. For example, a family emergency comes un [sic] and you still have work due and assigned.”
- “Teachers should consider their students work load [sic] to ensure that the student isn't stressed.”
- “Teachers need to be more understanding when things don't get done because most don't realize that we have lives and are busy A LOT!”
- “A teacher should take into consideration students [sic] home life before disciplinary action.”
- “I feel that is very important for teachers to give us plenty of time for homework because we have other classes, sports, and an outside life.”
- “I think teachers should take real life into account and also the stress their homework gives their students and their students mental health and state.”
- “For the teachers to be understanding. Like to understand that we have things out of school.”

In discipline, pedagogy, homework, and classroom policies, survey participants expressed a strong desire for empathy – one that could perhaps be included in the teacher evaluation process in the future. Indeed, the literature affirms the sentiment of these participants with regard to the value of teacher empathy (Berkovich, 2020; Makoelle, 2019; Dereli, & Aypay, 2012).

RQ3 *Are there similarities in student and administrator sentiment about effective teaching practices?*

Results from Survey 2 indicate that, while administrators and students may have felt the teacher in the sample teaching video did not expressly promote safety, health, and orderliness, there does seem to be dissention among both the student group ($SD = 1.087$) and administrator group ($SD = 0.58$) on Q3: “The teacher in the video promoted safety, health, and orderliness.”

This dissention may be due to the nature of the observation. A brief clip of a moment-in-time teaching video simply may not have afforded sufficient opportunity for students and administrators to make a sound judgement on this issue.

The student mean ($M = 3.68$) in Q6, “The teacher in the video taught about things that would help students meet state standards for student education,” was notably below the comprehensive student mean of all questions on the survey ($M = 3.95$); however, that may be expected, as students simply may not have knowledge of the state standards for student education. On this point, the administrator mean ($M = 4.33$) is much higher than the student mean. Theoretically, administrators would indeed be more familiar with state standards for student education. The standard deviation of student responses for Q6 ($SD = 1.121$) was the second highest of all student standard deviations, which may attest to a general lack of understanding among the student group of what the state standards are (and a teacher’s satisfaction of those standards). Perhaps a different or new evaluation instrument could be more specific about the “state standards” in question at the time of evaluation and would articulate those traits in a way that is more accessible to students. Additionally, increased student involvement in the formal teacher evaluation process may also improve student understanding of what is (and perhaps needs to be) assessed during formal teacher evaluations.

Q7, “The teacher in the video taught ‘actively,’ asking questions and encouraging students to act on what they learn,” assesses the teacher’s incorporation of active learning in the classroom, particularly through inquiry. The value of this type of active learning by inquiry includes “[encouraging] student curiosity, [making] inquiry visible, [emphasizing] the importance of topics and questions, [facilitating] the process of gathering and presenting information, and [integrating] technology” (Cattaneo, 2020). The student mean ($M = 4.41$) and the administrator mean ($M = 5.00$) for Q7 were quite close, with a difference in means of only 0.59. Additionally, the standard deviation among student participants ($SD = 0.958$) and among administrator

participants ($SD = 0.00$) was particularly low relative to those of other questions in the survey. The combination of the low variance among students and administrators with the relatively high means of each may suggest that the assessed trait as exhibited by the teacher in the video was clear to both students and administrators and that they were in agreement within groups about the presence of this trait in the teaching video.

Similarly, the variance among students ($SD = 0.999$) as well as among administrators ($SD = 0.00$) on Q10, “The teacher in the video modeled or demonstrated what she wanted students to learn to make sure students got the lesson,” was relatively minimal. Additionally, the student mean ($M = 3.96$) and the administrator mean ($M = 4.00$) were low relative to the means of the other 18 survey questions, suggesting that both groups felt similarly that the teacher in the video may clearly have fallen short of exhibiting the behavior described in the question.

Q14, “The teacher in the video modified lessons and tests for students who had a hard time learning,” represented somewhat of an anomaly. The student mean ($M = 3.50$) was the lowest of all student means in the second survey, and the administrator mean ($M = 3.00$) was even lower. The difference in means between student and administrator on this question is the only one in which the administrator mean was lower than the student mean. That being said, variance among students ($SD = 1.199$) was second highest among student variances on Survey #2, while the administrator mean ($SD = 0.00$) was the lowest among administrator variance in Survey #2. Administrators had a united front when it came to assessing the teacher’s ability to modify lessons and tests for students who had a hard time learning – and they all seemed to feel the teacher performed poorly in this area. Students also seemed to feel that the teacher did not perform well in this area; however, there was dissent in perceptions among the group.

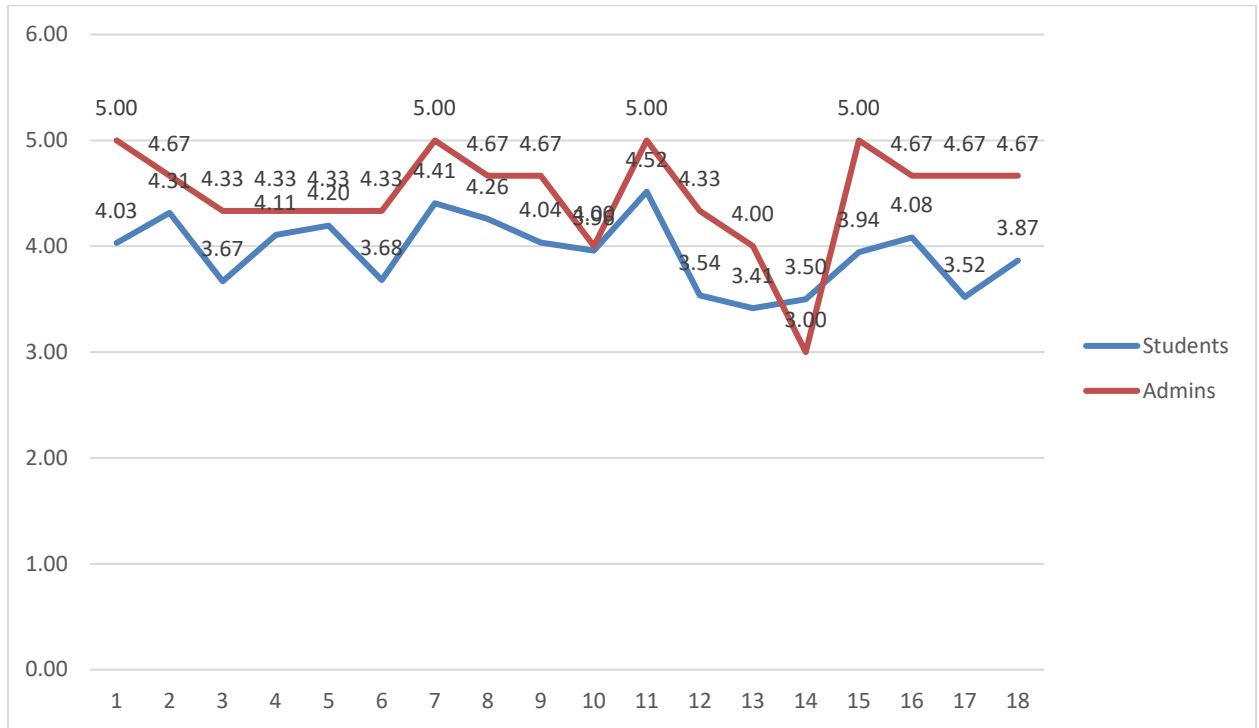
Findings Overview

To generalize, student sentiment pertaining to essential qualities of effective teaching do align with the TLE for the most part. In observing results from the first survey designed to assess student sentiment about 18 of the 20 dimensions evaluated using the TLE instrument, and utilizing a mean threshold of 3.5, students expressed that they felt that 77.8%, or 14 of 18, of the dimensions evaluated using the TLE instrument were indicators of effective teaching. Additionally, appending the four characteristics that students wrote in on Survey #1 to the TLE instrument may help inform teacher evaluation while at the same time incorporating student sentiment into this evaluation tool – a solid step in involving students in the formal teacher evaluation process.

When it comes to observing responses to Survey #2 in which students and administrators were asked to view an excerpt of a videotaped lesson and then “evaluate” the teacher in the video based on qualities present in the modified TLE, findings are more difficult to interpret. In reviewing Chart 1 below and in looking back at the *Difference in Means* listed in Table 3 above, one notices that the difference in mean responses between students and administrators ranges from .04 to 1.15. Anecdotally, one may notice that student and administrator means do follow a similar pattern in Chart 1; however, administrator means are consistently higher than student means with the exception only of Question 14. This phenomenon was addressed previously, but it would not belabor the point to acknowledge (again, anecdotally) the pattern of student vs. administrator means, which does have many similarities.

Figure 2

Survey #2 – Student Means vs. Administrator Means



The phenomenon of the consistently higher administrator means may be attributed to the very small population of administrators being surveyed. With a population of only three administrator participants, means identical to student response means (which evolved from a much larger population) simply were not mathematically possible. Had the administrator population size been closer to the size of the student population, means may have aligned even more closely. Such theorizing aside, similarities in student and administrator means on Survey #2 may indicate that students' ability to evaluate teachers somewhat rival those of administrators on many, although obviously not all, points. This is important in that it speaks to a potential concern that students simply lack the knowledge, ability, and maturity to evaluate teachers properly. Perhaps, if administrator evaluation instruments were adapted universally (as they were in this study), students may be seen as very adept in evaluating teachers. To extrapolate further, perhaps

evaluation instruments could be adapted for use by even younger student populations such as those in primary grades. Further research in this area might be valuable to the teacher evaluation literature.

Suggestions for Future Research

The purpose of this study was to gain a better understanding of student sentiment about characteristics of effective teachers, ultimately with the hope of informing the development a survey instrument that better captures student perspectives. While this study looked closely at the TLE evaluation instrument, little research focused on student sentiment about teaching traits assessed using other teacher competency-based evaluation instruments (e.g., the Framework model and the Marzano Focused Teacher model, to name a couple). In addition to surveying student sentiment about traits examined using specific evaluation *instruments*, one might also conduct a study in which other *methods* of teacher evaluation (e.g., the Value-Added Model of teacher evaluation) are examined. A study examining a combination of instruments and methods may also be valuable to the prevailing literature on student involvement in the formal teacher evaluation process.

One interesting discovery that was not part of the original purpose of this study has to do with the observation of variance among the student participant group as well as the administrator participant group when asked about 18 teaching qualities. This particular metric was discovered simply by calculating the standard deviation of survey responses across all participants within each respective group. While it was not the focus of this particular study, further study about variances among surveyed groups may prove to be valuable for future research in the context of assessing sentiment across participating groups. More specifically, a study that observes congruence or dissidence among various populations with regard to sentiment about teacher effectiveness would constitute a salient contribution to teacher evaluation literature.

Future research should also attempt to explore other theoretical frameworks through which teacher effectiveness is analyzed. The foundation of the TLE evaluation instrument, and consequently, the foundation of this study, rests on the contextual and instructional attributes espoused and propagated by Kathleen Cotton and the Northwest Regional Educational Lab. Research employing other theoretical frameworks could prove useful in informing teacher effectiveness research.

Implications for Future Research, Policy, and Practice

This exploratory study about student sentiment about effective teaching practices as assessed using the TLE evaluation instrument could have implications for research, policy, and practice.

Research. While scores of studies have been conducted that examine teacher evaluation methods and even more examine the relationship between student and teacher, very few examine the important role students can play in the formal teacher evaluation process. Should the inclusion of student sentiment in the teacher evaluation process become a serious consideration on a broad scale at the secondary level, opportunities for research would abound. This study bolsters the argument that student voice should be a required consideration in the teacher evaluation, particularly in secondary schools. This study, along with future investigations, could begin to provide scholars and school leaders with evidence useful in developing new methods of student-inclusive evaluation methodologies.

Previous literature that examines the student's role in teacher evaluation is often limited to instructor evaluations in university settings. To this researcher's knowledge, no previous study examining student voice in the context of the utilization of the TLE in formal teacher evaluation has been conducted. This deficiency in the literature need not be limited to the TLE only but

could be easily expanded to include a variety of teacher evaluation processes and instruments in the secondary school context.

As there is currently no widespread requirement to include student voice in the formal teacher evaluation in secondary schools, and as the evolution of a requirement would likely not come without some controversy, it would behoove the research community to produce more scholarly research about this topic.

Policy. The exclusion of student sentiment in the formal teacher evaluation process may be seen as a deficiency in the teacher evaluation process in general, and one that, if rectified, could compel policymakers to enact changes in teacher evaluation requirements and methods. At minimum, formal evaluation instruments may require a refresh to include the voice of students. Longer term, policymakers could standardize the formal teacher evaluation process across public schools in the United States and include in the new standard for secondary teacher evaluation a requirement for obtaining student input.

Practice. While this study focuses on theory, the underlying intent of this study is to inform improvements in teaching practice – practice that is both required and evaluated. Student *feedback* (that is, information gleaned from an informal process outside of the formal evaluation process) has been a target of study, and while some argue that “reliability and quality of student feedback may be questionable” (Hattie 2009), research has found student “perspectives insightful [in] providing information that an outside observer or teacher may have missed” (Mandouit, 2017; Gün, 2011). However, feedback alone may fail to motivate change (Kember, Leung, and Kwan 2002). Incorporating student sentiment into the *formal* evaluation process may motivate teachers on a career level, as the *formal evaluation* is less of a suggestion than is *feedback*. On the contrary, formal evaluation is more of a professional requirement. Again, more research would be required in this area to inform what would surely be a controversial measure. Ultimately, teachers

must perceive the process as an opportunity for growth and development (Mandouit, 2017), serving the good of the students to whom they have devoted themselves professionally. Failure to do so may “become a mere ritual completed at the end of every subject, with little meaning or purpose” (Abrami, d’Apollonia, & Rosenfield, 2007) and have no impact on sincere practice whatsoever.

The evidence from this study shows that students seem to be aligned with the creators of the TLE with regard to many traits seen as most valuable to teaching effectiveness. Of the 18 traits that were scored by students in the first survey of this study – traits that were adapted directly from the TLE instrument – only four fell below the minimum score to be included on the second survey, and only four traits not currently measured by the TLE were added to Survey #2. Despite this relative alignment between students and the accepted evaluation instrument used in many schools, the evolution from study to practice would require time and care, as “some teachers may worry that the practice of giving students a voice regarding teaching practices undermines the teacher’s authority,” (Flutter, 2007; Mandouit, 2017) while others may simply disregard results, arguing that secondary students lack both the academic or professional knowledge and experience necessary to evaluate a teacher’s performance (Mandouit, 2017). This study argues that students are a teacher’s primary “client,” and, no other stakeholder has as much face time and exposure to real teaching behaviors in the classroom as students. In light of this, students may actually be the most qualified to evaluate teaching performance. Perhaps the evidence gathered from this exploratory investigation will lead to decisions at the district, state, and national levels to weave secondary student sentiment into the evaluation tool used in formal teacher review processes in the future.

REFERENCES

- Abrami, P.C., d'Apollonia, S., Rosenfield, S. (2007). The Dimensionality of Student Ratings of Instruction: What We Know and What We Do Not* . In: Perry, R.P., Smart, J.C. (eds) *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*. Springer, Dordrecht. https://doi.org/10.1007/1-4020-5742-3_10.
- Alderman, G. (2009) Defining and measuring academic standards: a British perspective. *Higher Education Management and Policy*, 21(3), 11–22.
- Aleamoni, L. M. (1981). Student ratings of instruction. *Handbook of teacher evaluation, Beverly Hills, CA: Sage*, 110-145.
- Aleamoni, L.M. (1987). Student rating myths versus research facts. *Journal of Personnel Evaluation in Education*, 1, 111-119.
- Aleamoni, L.M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education*, 13, 153-166.
- Arubayi, E.A. (1987). Improvement of instruction and teacher effectiveness: Are student ratings reliable and valid? *Higher Education*, 16(3), 26-278.
- Au, W., & Gourd, K. (2013). Asinine Assessment: Why High-Stakes Testing Is Bad for Everyone, Including English Teachers. *The English Journal*, 103(1), 14-19. Retrieved from <http://www.jstor.org.ezproxy.osu-tulsa.okstate.edu/stable/24484054>.

- Baker, Eva L., Barton, Paul E., Darling-Hammond, Linda, Haertel, Edward, Ladd, Helen F., Linn, Robert L., Ravitch, Diane, Rothstein, Richard, Shavelson, Richard J., & Shepard, Lorrie A. (2010). Problems with the Use of Student Test Scores to Evaluate Teachers. *Economic Policy Institute Briefing Paper #278*. Retrieved from <http://www.epi.org/publication/bp278/>.
- Bales, B. L. (2006). Teacher education policies in the United States: The accountability shift since 1980. *Teaching and Teacher Education, 22*(4), 395–407.
- Berkovich, I. (2020). Conceptualizations of empathy in K-12 teaching: a review of empirical research. *Educational Review, 72*(5), 547–566.
<https://doi.org/10.1080/00131911.2018.1530196>.
- Berti, Mameli, C., Speltini, G., & Molinari, L. (2016). Teacher justice and parent support as predictors of learning motivation and visions of a just world. *Issues in Educational Research, 26*(4), 543–560.
- Brookheart, S. M. (2013). The public understanding of assessment in educational reform in the United States. *Oxford Review of Education, 39*(1), 52-71.
- Brookover, W.B., & Lezotte, L.W. (1979). *Changes in school characteristics coincident with changes in student achievement* (Occasional Paper No. 17). East Lansing, MI: Michigan State University, Institute for Research on Teaching. (ERIC Document Reproduction Service No. ED 181 005).
- Brophy, J.E. (1979). *Teacher behavior and its effects*. (Occasional Paper No. 25). East Lansing, MI: Michigan State University, Institute for Research on Teaching. (ERIC Document Reproduction Service No. ED181 014).

- Brophy, J.E. (1987). Synthesis of research on strategies for motivating students to learn. *Educational Leadership*, 45(2), 40-48.
- Brott, T. (2016). *A Model for Stabilized Creative Education: Using Creativity in the Arts to Promote Student Success*. Unpublished master's project, The University of Minnesota, Minneapolis, Minnesota.
- Campbell, J., Kyriakides, L., Juijs, D., & Robinson, W. (2004). *Assessing Teacher Effectiveness*. London: RoutledgeFalmer.
- Cantrell, S. and Kane, T. J. (2013). *Ensuring Fair and Reliable Measures of Effective Teaching*. Retrieved from <http://www.edweek.org/media/17teach-metl.pdf>.
- Casellas, J., & Shelly, B. (2012). No Latino Left Behind? Determinants of Support for Education Reform in the U.S. Congress. *Journal of Latinos & Education*, 11(4), 260-270. doi:10.1080/15348431.2012.715505.
- Chen, Y. and Hoshower, L. (2003). Student Evaluation of Teaching Effectiveness: An assessment of student perception and motivation. *Assessment & Evaluation in Higher Education*, 28(1), 71-88.
- Clark, D. L., & Astuto, T. A. (1994). Redirecting reform: Challenges to popular assumptions about teachers and students. *Phi Delta Kappan*, 75, 513-520.
- Cohen, S.A. (1987). Instructional alignment: Searching for a magic bullet. *Educational Researcher*, 16(8), 16-20.
- Collins, J., & O'Brien, N. (2003). *The Greenwood dictionary of education*. Westport, Conn.: Greenwood Press.

- Cook-Sather, A. (2002). Authorizing Students' Perspectives: Toward Trust, Dialogue, and Change in Education. *Educational Researcher*, 31(4) 3-14.
- Cook-Sather, A. (2010). Students as Learners and Teachers: Taking Responsibility, Transforming Education, and Redefining Accountability. *Curriculum Inquiry*, 40(4), 555-575.
doi:10.1111/j.1467-873X.2010.00501.x.
- Costin, F., Greenough, W.T., & Menges, R.J. (1971). Student ratings of college teaching: Reliability, validity, and usefulness. *Review of Educational Research*, 41.511-535.
- Cotton, Kathleen. (2000). *The Schooling Practices that Matter Most*. Alexandria, VA: Northwest Regional Educational Laboratory. Retrieved from
<http://files.eric.ed.gov/fulltext/ED469234.pdf>.
- Daggett, B. (2014). *Addressing Current and Future Challenges in Education*. 22nd Annual Model Schools Conference, Lake Buena Vista, FL, June 2014. Rexford, NY: International Center for Leadership in Education.
- Darling-Hammond, Linda. (2000). Solving the Dilemmas of Teacher Supply, Demand, and Standards: How We Can Ensure a Competent, Caring, and Qualified Teacher for Every Child. National Commission on Teaching & America's Future,
<https://files.eric.ed.gov/fulltext/ED463337.pdf>.
- Dereli, E. and Ayşe Aypay. (2012). The Prediction of Empathetic Tendency and Characteristic Trait of Collaboration on Humane Values in Secondary Education Students and the Examining to Those Characteristics. *Educational Sciences : Theory & Practice*, 12(2), 1262–1270.
- Doherty, Kathryn M. & Jacobs, Sandi. (2013). Connect the Dots: Using evaluations of teacher effectiveness to inform policy and practice. National Council on Teacher Quality.

- Edmonds, R.R. (1979). Effective schools for the urban poor. *Educational Leadership*, 37(1), 15-18, 20-24.
- Ellett, Chad D. & Teddlie, Charles. (2003). Teacher Evaluation, Teacher Effectiveness and School Effectiveness: Perspectives from the USA. *Journal of Personnel Evaluation in Education: 17*(1), 101-128.
- Ellis, E.S., & Worthington, L.A. (1994). *Research synthesis on effective teaching principles and the design of quality tools for educators* (Tech. Rep. No. 5). Eugene, OR: University of Oregon, National Center to Improve the Tools of Educators. (ERIC Document Reproduction Service No. ED 386 853).
- Every Student Succeeds Act (ESSA)*. (2015). Retrieved from <https://www.ed.gov/essa>.
- Fennell, M. (2016). What Educators Need to Know about ESSA. *Educational Leadership*, 73(6), 62-65. Retrieved from <http://www.ascd.org/publications/educational-leadership/jun16/vol73/num09/What-Educators-Need-to-Know-about-ESSA.aspx>.
- Flutter, J. (2007). "Teacher Development and Pupil Voice." *Curriculum Journal*, 18(3): 343–354.
- Furrer, C., Skinner, E. & Pitzer, J. (2014). *The Influence of Teacher and Peer Relationships on Students' Classroom Engagement and Everyday Motivational Resilience*. *National Society for the Study of Education*, 113(1), 101-123.
- Glickman, C. D., Gordon, S.P., & Ross-Gordon, J.M. (2018). *Instructional supervision: A developmental approach (10th ed.)*. New York: Pearson.
- Goodboy, A. K. (2011). Making Sense of Students' Complaints, Criticisms, and Protests. *Communication Currents*, 6.

- Gün, Bahar. (2011). Quality self-reflection through reflection training, *ELT Journal*, 65,(2), 126–135.
- Guskey, T.R. (1994). Making the grade: What benefits students? *Educational Leadership*, 52(2), 14-20.
- Hamilton, Laura & Stecher, Brian & Yuan, Kun. (2008). Standards-Based Reform in the United States: History, Research, and Future Directions. 78.
- Harris, D. N., Ingle, W. K., & Rutledge, S. A. (2014). How Teacher Evaluation Methods Matter for Accountability: A Comparative Analysis of Teacher Effectiveness Ratings by Principals and Teacher Value-Added Measures. *American Educational Research Journal*, 2014 51: 73-75. doi:10.3102/0002831213517130.
- Harvey, L. (2002) Evaluation for what?, *Teaching in Higher Education*, 7(3), 246–263.
- Herman, J.L. (1992). What research tells us about good assessment. *Educational Leadership*, 49(8), 74-78.
- Hill, D. (2000). He’s Got Your Number. *Education Week Teacher*. 11(8), 42-47.
- Hood, Cattaneo, K. (2017). Telling Active Learning Pedagogies Apart: from theory to practice. *Journal of New Approaches in Educational Research (NAER Journal)*, 6(2), 144-152.
- “Improving Basic Programs Operated by Local Educational Agencies (Title I, Part A).” (2015). U.S. Department of Education. Retrieved from <https://www2.ed.gov/programs/titleiparta/index.html>.

- Ingle, W. K., Rutledge, S. A., & Bishop, J. L. (2011). Context matters: Principals' sensemaking of teacher hiring and on-the-job performance. *Journal of Educational Administration*, 49, 579–610.
- Jennings, Patricia A., & Greenberg, Mark T. (2009). The Prosocial Classroom: Teacher Social and Emotional Competence in Relation to Student and Classroom Outcomes. *Review of Educational Research*, 79(1), 491-525.
- Jewell, J. W. (2017). From Inspection, Supervision, And Observation to Value-Added Evaluation: A Brief History of U.S. Teacher Performance Evaluations. *Drake Law Review*, 65(2), 363-419.
- Johns, R. (2005). One size doesn't fit all: Selecting response scales for attitude items. *Journal of Elections, Public Opinion & Parties*, 15, 237-264. doi: 10.1080/13689880500178849.
- Kember, D., D.Y. Leung, and K. Kwan. (2002). "Does the Use of Student Feedback Questionnaires Improve the Overall Quality of Teaching?" *Assessment & Evaluation in Higher Education*, 27(5): 411–425.
- Klein, A. (2015, April 10). No Child Left Behind: An Overview. *Education Week*. Retrieved from <https://www.edweek.org/ew/section/multimedia/no-child-left-behind-overview-definition-summary.html>.
- Klem, A. M., & Connell, J. P. (2004). Relationships matter: Linking teacher support to student engagement and achievement. *Journal of School Health*, 74(7), 262-273.
- Kline, R. B. (2011). *Principles and Practice of Structural Equation Modeling* (3rd edition). New York: Guilford Press.

- Klock, J. (2010). Building Novel Connections in an Increasingly Standardized World. *Teacher Librarian*, 38(2), 15-18.
- Koretz, D., & Debert, E. (1995). Setting Standards and Interpreting Achievement: A Cautionary Tale from the National Assessment of Educational Progress. *Educational Assessment*, 3(1), 53.
- Krug, S.E. (1992). *Instructional leadership, school instructional climate, and student learning outcomes* [Project report]. Urbana, IL: National Center for School Leadership. (ERIC Document Reproduction Service No. ED 359 668).
- Lareau, A. (2011). *Unequal Childhoods: Class, Race, and Family Life*. Los Angeles, CA: University of California Press.
- Lawrence, J. J., & McCollough, M. A. (2004). Implementing Total Quality Management in the Classroom by Means of Student Satisfaction Guarantees. *Total Quality Management & Business Excellence*, 15(2), 235-254.
- Levine, D.U., & Lezotte, L.W. (1995). Effective schools research. In J.A. Banks & C.A.M. Banks, *Handbook of research on multicultural education*. New York: Macmillan.
- Little, O. (2009). *Teacher evaluation systems: The window for opportunity and reform*. Washington, DC: National Education Association.
- Lohman, J. (2010). Comparing No Child Left Behind and Race to the Top. Connecticut General Assembly. Retrieved from <https://www.cga.ct.gov/2010/rpt/2010-r-0235.htm>.
- Louis, K., Murphy, J., & Smylie, M. (2016). Caring Leadership in Schools. *Educational Administration Quarterly*, 52(2), 310-348.

- Mahatmya, D., Brown, R. C., & Johnson, A. D. (2014). Student-as-Client. *Phi Delta Kappan*, 95(6), 30-34.
- Makoelle, Tsediso. (2019). Teacher empathy, A Prerequisite for an Inclusive Classroom. In M. A. Peters (Ed.), *Encyclopedia of Teacher Education*. Retrieved from SpringerLink database.
- Mandouit , Luke (2018) Using student feedback to improve teaching, *Educational Action Research*, 26(5), 755-769.
- Marsh, H.W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*. 76.707—754.
- Marsh, H.W. and Roche, L. A. Making Students' Evaluations of Teaching Effectiveness Effective. *American Psychologist*. 52(11), 1187-1197.
- Marzano, R. J., Frontier, T., and Livingston, D. (2011). *Effective Supervision: Supporting the Art and Science of Teaching*. Alexandria, VA: ASCD.
- McGee, Laura. (2013). *TLE Overview* [PDF file]. Retrieved from http://sde.ok.gov/sde/sites/ok.gov.sde/files/documents/files/TLE_White_Paper_01_13.pdf.
- McReynolds, K. (2006). The No Child Left Behind Act Raises Growing Concerns. *Encounter*, 19(2), 33-36.
- 'MET' Made Simple: Building Research-Based Teacher Evaluations. (2012). Retrieved from <http://files.eric.ed.gov/fulltext/ED544428.pdf>.
- Millman, J., & Darling-Hammond, L. (Eds.) (1990). *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers*. Newbury Park, CA: Sage.

- Moir, E. (2014, May). Keynote. In E. Harris (Administrator), *Brock Approval*. Symposium conducted at the meeting of the 2014 Brock Prize Symposium, Norman, OK.
- Morgan, W. & Wyatt-Smith, C.M. (2000) Im/proper accountability: towards as theory of critical literacy and assessment, *Assessment in Education*, 7(1), 123–142.
- Muldoon, L., ed. (2009). "Race to the Top: Accelerating College and Career Readiness in States - Teacher Effectiveness". Retrieved from <https://www.achieve.org/files/RTTT-TeacherEffectiveness.pdf>.
- National Commission on Excellence in Education. (1983). *A Nation at risk: The Imperative for educational reform*.
- No Child Left Behind Act of 2001, 20 U.S.C. (2008).
- No Child Left Behind Act: Volume IX—Accountability Under NCLB: Final Report* [PDF]. (2010). US Department of Education. Retrieved from <https://files.eric.ed.gov/fulltext/ED508912.pdf>.
- Nowlis, S. M., Kahn, B. E., & Dhar, R. (2002). Coping with ambivalence: The effect of removing a neutral option on consumer attitude and preference judgments. *Journal of Consumer Research*, 29, 319-334. doi: 10.1086/344431.
- O'Conner, K. (1995). Guidelines for grading that support learning and student success. *NASSP Bulletin*, 79, 91-101.
- Oklahoma State Board of Education Adopted Policies Pursuant to 70 O.S. § 6-101.16A* [PDF file]. (2011). Retrieved from <http://sde.ok.gov/sde/sites/ok.gov.sde/files/TLE-AdpotedPolicies.pdf>.

Oklahoma Teacher and Leader Effectiveness Evaluation System (TLE) REPORT [PDF file].

(2011). Retrieved from <http://sde.ok.gov/sde/sites/ok.gov.sde/files/TLE-ReportLeg.pdf>.

Peterson, K., Wahlquist, C., and Bone, K. (2000). Student Surveys for School Teacher Evaluation. *Journal of Personnel Evaluation in Education*, 14(2), 135-153. Ravitch, Diane. (2000). *Left Back: A Century of Failed School Reforms*. New York, NY: Simon & Schuster.

Quantitative and Qualitative Research. (2009). Retrieved May 3, 2017, from Explorable.com: <https://explorable.com/quantitative-and-qualitative-research>.

Ravitch, Diane. (2000). *Left Back: A Century of Failed School Reforms*. New York, NY: Simon & Schuster.

Rivkin, S. G., Hanushek, E. A. and Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, 73: 417-458. doi:10.1111/j.1468-0262.2005.00584.x.

Rosenshine, B. (1976). Recent research on teaching behaviors and student achievement. *Journal of Teacher Education*, 27(1), 61-64.

Rosenshine, B. (1986). Synthesis of Research on Explicit Teaching. *Educational Leadership*, 43(7), 60-69.

Rosenshine, B. (1995). Advances in research on instruction. *Journal of Educational Research*, 88, 262-268.

Ryan, J. (2004). The Perverse Incentives of the No Child Left Behind Act. *New York University Law Review*. 79(3), 932-989.

Saltman, K. (2017). *Scripted Bodies*. Abingdon, Oxon: Taylor & Francis.

- Shevlin, M., Banyard, P., Davies, M., and M. Griffiths. "The Validity of Student Evaluation of Teaching in Higher Education: Love Me, Love My Lectures?" *Assessment and Evaluation in Higher Education* 25(4), 397–405.
- Shin, J.C. (2010). Impacts of performance-based accountability on institutional performance in the U.S. *Higher Education*, 60(1), 47-68. doi:10.1007/s10734-009-9285-y.
- Siegle, D., Rubenstein, L.D., & Mitchell, M.S. (2014). Honors Students' Perceptions of Their High School Experiences: The Influence of Teachers on Student Motivation. *Gifted Child Quarterly*, 58(1), 35-50.
- Slavin, R.E. (1994). Quality, appropriateness, incentive, and time: A model of instructional effectiveness. *International Journal of Educational Research*, 21, 141-157.
- Smith, J. and Philip E. Kovacs. (2011). The impact of standards-based reform on teachers: the case of 'No Child Left Behind', *Teachers and Teaching*, 17:2, 201-225, DOI: 10.1080/13540602.2011.539802.
- Stahl, S.A., & Clark, C.H. (1987). The effects of participatory expectations in classroom discussion on the learning of science vocabulary. *American Educational Research Journal*, 24, 541-555.
- Stallings, D. T. (2002). A Brief History of the United States Department of Education: 1979-2002. *Phi Delta Kappan*. 83(9), 4.
- Stallings, J.A. (1985). A study of implementation of Madeline Hunter's model and its effects on students. *Journal of Educational Research*, 78, 325-337.

- Stronge, J. H. (1995). Balancing Individual and Institutional Goals in Educational Personnel Evaluation: A Conceptual Framework. *Studies in Educational Evaluation*. 21(2), 131-151.
- TLE Commission Recommendation Regarding Other Academic Measures* [PDF file]. (2012). Retrieved from <http://sde.ok.gov/sde/sites/ok.gov.sde/files/sb-december19-9b2memo.pdf>.
- Taylor, J., O'Day, J., and Le Floch, K, 2010. *State and Local Implementation of the No Child Left Behind Act: Volume IX—Accountability Under NCLB: Final Report* [PDF]. (2010). US Department of Education. Retrieved from <https://files.eric.ed.gov/fulltext/ED508912.pdf>.
- Teacher and Leader Effectiveness (TLE)*. (2015). Retrieved from <http://sde.ok.gov/sde/tle>.
- Teacher and Leader Evaluations (TLE) Frequently Asked Questions* [PDF file]. (2012, August). Retrieved from http://sde.ok.gov/sde/sites/ok.gov.sde/files/TLE-FAQ_0.pdf.
- Teacher and Leader Qualitative Assessment Models* [PDF Download]. (2011). Retrieved from <http://sde.ok.gov/sde/sites/ok.gov.sde/files/TLE-Models-Overview.pdf>.
- Thomas, J. Y., and Brady, K. P. (2005). The Elementary and Secondary Education Act at 40: Equity, Accountability, and the Evolving Federal Role in Public Education. *Review of Research in Education*. 29(1), 56-67.
- Towles-Reeves, E., Kleinert, H., & Muhomba, M. (2009). Alternate Assessment: Have We Learned Anything New?. *Exceptional Children*, 75(2), 233-252.
- Tschannen-Moran, M. & Hoy, A. (2001). Teacher efficacy: capturing an elusive construct. *Teaching and Teacher Education*. 17(7), 783-805.

- Tschannen-Moran, M., Hoy, A., & Hoy, W. (1998). Teacher Efficacy: Its Meaning and Measure. *Review of Educational Research*, 68(2), 202-248. Retrieved from <http://www.jstor.org.ezproxy.osu-tulsa.okstate.edu/stable/1170754>.
- Tulsa Public Schools' Teacher Observation and Evaluation System: Its Research Base and Validation Studies* [PDF file]. (2011). Retrieved from https://www.tulaschools.org/4_About_District/_documents/TLE/Teacher_Eval_System_Research_Brief.pdf.
- Tulsa TLE Model Training for New Administrators*. (n.d.). Retrieved from https://www.ccosa.org/index.php?events&a=view&event_id=196.
- US Department of Education. *Every Student Succeeds Act*. Retrieved from <http://www.ed.gov/essa>.
- US Department of Education. (2004). *New No Child Left Behind Flexibility: Highly Qualified Teachers*. Retrieved 27 July 2014, from <http://www2.ed.gov/nclb/methods/teachers/hqtflexibility.html>.
- US Department of Education. (2009). *Race to the Top Fact Sheet*. Retrieved from <https://www2.ed.gov/programs/racetothetop/factsheet.pdf>.
- Wendorf, & Alexander, S. (2005). The Influence of Individual- and Class-Level Fairness-Related Perceptions on Student Satisfaction. *Contemporary Educational Psychology*, 30(2), 190–206.
- Wilkerson, D., Manatt, R., Rogers, M., & Maughan, R. (2000). Validation of Student, Principal, and Self-Ratings in 360° Feedback for Teacher Evaluation. *Journal of Personnel Evaluation in Education*: 14(2). 179-192.

Winstead, L. (2011). The Impact of NCLB and Accountability on Social Studies: Teacher Experiences and Perceptions about Teaching Social Studies. *Social Studies*, 102(5), 221-227. doi:10.1080/00377996.2011.571567.

Zhu, J. (2014). *Quantitative Models for Performance Evaluation and Benchmarking*. Ney York, NY: Springer.

APPENDICES

APPENDIX A: Survey Items

Survey #1

1. A teacher should always be prepared for class.

1 - 2 - 3 - 4 - 5
Not important Somewhat important Very important

2. A teacher should monitor and manage the behavior of students.

1 - 2 - 3 - 4 - 5
Not important Somewhat important Very important

3. A teacher should follow practices outlined by the school to promote safety, health, and orderliness.

1 - 2 - 3 - 4 - 5
Not important Somewhat important Very important

4. A teacher should create meaningful lesson plans that are based on successful teaching practices of other teachers.

1 - 2 - 3 - 4 - 5
Not important Somewhat important Very important

5. A teacher should grade things fairly and in a way that truly evaluates a student's growing knowledge and learning.

1 - 2 - 3 - 4 - 5
Not important Somewhat important Very important

6. A teacher should respectfully express high expectations of his/her students.

1 - 2 - 3 - 4 - 5
Not important Somewhat important Very important

7. A teacher should emphasize reading and writing in every lesson, even if the lesson is not about reading and writing.

1 - 2 - 3 - 4 - 5
Not important Somewhat important Very important

8. A teacher should teach about things that help students meet state standards for student education.

1 - 2 - 3 - 4 - 5
Not important Somewhat important Very important

9. A teacher should teach “actively,” asking questions and encouraging students to act on what they learn.

1 - 2 - 3 - 4 - 5
Not important Somewhat important Very important

10. A teacher should use many methods to teach (e.g., lecture, group discussion, hands-on activities, etc.).

1 - 2 - 3 - 4 - 5
Not important Somewhat important Very important

11. A teacher should give clear instructions and directions to students.

1 - 2 - 3 - 4 - 5
Not important Somewhat important Very important

12. A teacher should “model” or demonstrate what he/she wants students to learn to make sure students get the lesson.

1 - 2 - 3 - 4 - 5
Not important Somewhat important Very important

13. A teacher should move around the room and interact with students to be sure they understand what they’re supposed to do on assignments and lessons.

1 - 2 - 3 - 4 - 5
Not important Somewhat important Very important

14. A teacher should adjust instructional plans if he/she realizes that students aren't learning.

1 - 2 - 3 - 4 - 5
Not important Somewhat important Very important

15. A teacher should summarize what students have learned in class that day.

1 - 2 - 3 - 4 - 5
Not important Somewhat important Very important

16. A teacher should modify lessons and tests for students who have a hard time learning.

1 - 2 - 3 - 4 - 5
Not important Somewhat important Very important

17. A teacher should regularly communicate with students, their families, and other educators.

1 - 2 - 3 - 4 - 5
Not important Somewhat important Very important

18. A teacher should volunteer to lead at school and/or district events and help other teachers become better teachers.

1 - 2 - 3 - 4 - 5
Not important Somewhat important Very important

If you feel that are some characteristics of good teaching that weren't mentioned in the statements above, please write them in below. If you feel like the statements above covered everything, feel free to leave this section blank.

Survey #2

1. The teacher in the video was prepared for class.

1 - 2 - 3 - 4 - 5
Not effective Somewhat effective Very effective

2. The teacher in the video monitored and managed the behavior of students.

1 - 2 - 3 - 4 - 5
Not effective Somewhat effective Very effective

3. The teacher in the video promoted safety, health, and orderliness.

1 - 2 - 3 - 4 - 5
Not effective Somewhat effective Very effective

4. The teacher in the video created meaningful lesson plans.

1 - 2 - 3 - 4 - 5
Not effective Somewhat effective Very effective

5. The teacher in the video evaluated students fairly and in a way that truly evaluated their growing knowledge and learning.

1 - 2 - 3 - 4 - 5
Not effective Somewhat effective Very effective

6. The teacher in the video taught about things that would help students meet state standards for student education.

1 - 2 - 3 - 4 - 5
Not effective Somewhat effective Very effective

7. The teacher in the video taught “actively,” asking questions and encouraging students to act on what they learn.

1 - 2 - 3 - 4 - 5
Not effective Somewhat effective Very effective

8. The teacher in the video used many methods to teach (e.g., lecture, group discussion, hands-on activities, etc.).

1 - 2 - 3 - 4 - 5
Not effective Somewhat effective Very effective

9. The teacher in the video gave clear instructions and directions to students.

1 - 2 - 3 - 4 - 5
Not effective Somewhat effective Very effective

10. The teacher in the video modeled or demonstrated what she wanted students to learn to make sure students got the lesson.

1 - 2 - 3 - 4 - 5
Not effective Somewhat effective Very effective

11. The teacher in the video moved around the room and interacted with students to be sure they understood what they were supposed to do on assignments and lessons.

1 - 2 - 3 - 4 - 5
Not effective Somewhat effective Very effective

12. The teacher in the video adjusted instructional plans if she realized that students were not learning.

1 - 2 - 3 - 4 - 5
Not effective Somewhat effective Very effective

13. The teacher in the video summarized what students learned in class that day.

1 - 2 - 3 - 4 - 5
Not effective Somewhat effective Very effective

14. The teacher in the video modified lessons and tests for students who had a hard time learning.

1 - 2 - 3 - 4 - 5
Not effective Somewhat effective Very effective

15. The teacher in the video presented herself professionally both in dress and in what she said.

1 - 2 - 3 - 4 - 5
Not effective Somewhat effective Very effective

16. The teacher in the video showed that she had mastered the material she was teaching the students.

1 - 2 - 3 - 4 - 5
Not effective Somewhat effective Very effective

17. The teacher in the video translated classroom lessons into practical, real-life lessons when possible.

1 - 2 - 3 - 4 - 5
Not effective Somewhat effective Very effective

18. The teacher in the video expressed individual empathy toward and an individual understanding of her students.

1 - 2 - 3 - 4 - 5
Not effective Somewhat effective Very effective

APPENDIX B: IRB Approval



Oklahoma State University Institutional Review Board

Date: 01/11/2021
Application Number: IRB-20-181
Proposal Title: Student Sentiment about Teacher and Leader Effectiveness
Principal Investigator: Chris Posey
Co-Investigator(s):
Faculty Adviser: Kathy Curry
Project Coordinator:
Research Assistant(s):
Processed as: Expedited
Expedited Category:

Status Recommended by Reviewer(s): Approved

Approval Date: 01/08/2021

The IRB application referenced above has been approved. It is the judgment of the reviewers that the rights and welfare of individuals who may be asked to participate in this study will be respected, and that the research will be conducted in a manner consistent with the IRB requirements as outlined in section 45 CFR 46.

This study meets criteria in the Revised Common Rule, as well as, one or more of the circumstances for which continuing review is not required. As Principal Investigator of this research, you will be required to submit a status report to the IRB triennially.

The final versions of any recruitment, consent, and assent documents bearing the IRB approval stamp are available for download from IRBManager. These are the versions that must be used during the study.

As Principal Investigator, it is your responsibility to do the following:

1. Conduct this study exactly as it has been approved. Any modifications to the research protocol must be approved by the IRB. Protocol modifications requiring approval may include changes to the title, PI, adviser, other research personnel, funding status or sponsor, subject population composition or size, recruitment, inclusion/exclusion criteria, research site, research procedures and consent/assent process or forms.
2. Submit a status report to the IRB when requested
3. Promptly report to the IRB any harm experienced by a participant that is both unanticipated and related per IRB policy.
4. Maintain accurate and complete study records for evaluation by the OSU IRB and, if applicable, inspection by regulatory agencies and/or the study sponsor.
5. Notify the IRB office when your research project is complete or when you are no longer affiliated with Oklahoma State University.

If you have questions about the IRB procedures or need any assistance from the Board, please contact the IRB Office at 405-744-3377 or irb@okstate.edu.

Sincerely,

Oklahoma State University IRB

VITA

Christopher D. Posey

Candidate for the Degree of

Doctor of Philosophy

Dissertation: STUDENT SENTIMENT ABOUT TEACHER AND LEADER
EFFECTIVENESS

Major Field: Educational Leadership and Policy Studies

Education:

Completed the requirements for the Doctor of Philosophy in Educational Leadership and Policy Studies at Oklahoma State University, Stillwater, Oklahoma in July, 2022.

Completed the requirements for the Master of Business Administration at Oklahoma State University, Stillwater, Oklahoma in 2007.

Completed the requirements for the Bachelor of Arts in Church Ministries at Southeastern University, Lakeland, Florida in 1993.

Experience:

OGI Process Equipment, Sand Springs, OK
Marketing Manager, January 2020 – Present

PDI-TouchStar, Tulsa, OK
Group Marketing Manager, April 2016 – December 2019

Tulsa Community College, Southeast Campus, Tulsa, OK
Adjunct Instructor, August 2001 – May 2012; January 2018 - present

Union Public Schools, Tulsa, OK
Secondary English Teacher, August 1999 – June 2007

Professional Memberships:

Board Member, Domestic Violence Intervention Services (DVIS), Tulsa, OK, 2020-
current

Marketing Committee Member, Up with Trees, Tulsa, OK, 2017-2019