The Relationship Between Teacher Evaluation Scores and Teachers' Years of Service and Level of Education

Michelle Y. Edwards
BA, State Texas A&M University, 1993
MA, Baker University, 2002
EdS, University of Central Missouri, 2013

Submitted to the Graduate Department and Faculty of the School of Education of Baker University in partial fulfillment of the requirements for the degree of Doctor of Education in Educational Leadership

James Robins
James Robins EdD
Baker University Major Advisor

Verneda Edwards
Verneda Edwards, EdD
Baker University Committee member

Kari Harrison
Kari Harrison, EdD
External Committee Member

Date Defended: August 30, 2023

Abstract

Hattie's (2011) research identified student engagement and formative assessment as teaching strategies with high effect size as it relates to improving student achievement. Using Hattie's research as a guide, District A uses student engagement and formative assessment as two required performance indicators in their teacher evaluation process. The purpose of this study was threefold. First, the researcher wanted to determine if the teachers in District A scored higher on the two NEE indicators than the threshold of 5, which is considered a moderate score. The second purpose of the study was to determine if the NEE scores for the two indicators were different based on the teachers' years of experience. The third purpose was to determine if the NEE scores on the two indicators were different based on the teachers' level of education. In Missouri, school districts have focused on hiring and maintaining highly qualified teachers since the early 2000's (DESE 2020). This includes hiring teachers with degrees in the subject areas taught, requiring professional development, valuing teachers with experience, and offering higher teacher salaries for advanced degrees. However, the previous research on this topic has produced mixed results on how impactful teacher experience and level of education are in increasing student achievement. The literature from this research detailed several studies that are supported by the outcomes of the research from this current study. Many research studies review the importance of years of experience and level of education in separate studies. However, this research studied both factors and connected it to two teaching strategies associated with improving student achievement. The research findings

provided evidence that there was no significant difference in NEE evaluation scores for teachers based on years of experience and level of education on the two NEE indicators: student engagement and conducting formative assessment. The research findings from this study may be surprising for educators who believe years of experience and level of education are key factors in teacher effectiveness. More insight might be gained with this same study in an urban or rural school district and with an increase in study participants.

Dedication

In middle school, I often studied my family members to determine how they were able to manage life, overcome obstacles, and attain varying levels of success. From their life experiences, I concluded two areas would contribute to my success in life: faith in God and the pursuit of education. Neither have failed me. I am grateful for the strong foundation established by the special people in my life.

This work is dedicated to my mother. She taught me the importance of hard work and perseverance. She believed that focusing on your goals is more important than focusing on the obstacles in your way. Her perspective on life has enabled me to develop a growth mindset focused on achievement. It is also dedicated to all of my aunts who inspired me throughout my life. Each of them heavily influenced my upbringing, and I value the time and care they exhibited towards me.

Acknowledgements

Completing my dissertation was an arduous process. I was initially on the fast track to finish ahead of my cohort group. However, a terrible car accident altered the course of my life and caused some long-term disabilities. I am grateful to Baker University for extending my deadlines amid my crisis. I am especially grateful to Dr. James Robins, my academic advisor, and Dr. Peg Waterman, my research analyst, for their continued patience and support as I fought through many battles to complete my dissertation. Their gentle nudges and encouragement helped me complete the journey. I would also like to acknowledge Dr. Verneda Edwards, one of my committee members. Even though her retirement was quickly approaching, she remained one of my committee members because she believed in my ability to finish the work. I would also like to acknowledge Dr. Kari Harrison, my external committee member. She was kind enough to assist me even though the review of the dissertation was at an extremely busy time in her professional and personal life.

I am grateful for my husband's inspiration in the art of writing. Although he has a business degree, his gift of expression is a true gift. I am thankful for his love, friendship, and support. The family and friends who supported me through this process should be commended as well. They endured my absence from events and tedious conversations about the process. I am especially grateful for my good friend, Lisa. She was the only person whom I truly allowed to see my vulnerability in this process and whom I asked for help.

Table of Contents

Abstract ii

Dedication iii

Acknowledgments iv

Table of Contents v

List of Tables viii

List of Figures ix

Chapter 1: Introduction 1

- Background 2
- Statement of the Problem 6
- Purpose of the Study 7
- Significance of the Study 8
- Delimitations 9
- Assumptions 10
- Research Questions 10
- Definition of Terms 11
- Organization of the Study 12

Chapter 2: Review of the Literature 14

- History of Teaching in America 14
- History of National Policies on Teacher Evaluation Systems 20
  - A Nation at Risk 20
  - No Child Left Behind (NCLB) 20
  - Every Child Succeeds Act (ESSA) 23

Race to the Top (RTTT) 24
Controversy Surrounding Teacher Evaluations 25
Highly Qualified Teachers 31
Teacher's years of experience 31
Teacher's level of education 33
History of The Network of Effective Educators (NEE) 35
Summary 42
Chapter 3: Methods 44
Research Design 44
Selection of Participants 45
Measurement 46
Data Collection Procedures 54
Data Analysis and Hypothesis Testing 55
Limitations 58
Summary 58
Chapter 4: Results 59
Descriptive Statistics 59
Hypothesis Testing 61
Summary 64
Chapter 5: Interpretation and Recommendations 67
Study Summary 67
Overview of the problem 68
Purpose statement and research questions 68

Review of the methodology 70

Major findings 71

Findings Related to the Literature 72

Conclusions 73

Implications for action 74

Recommendations for future research 74

Concluding remarks 75

References 76

Appendices 86

Appendix A. Baker University IRB 86

Appendix B. District 1 Approval Letter 87

Appendix C. NEE Data Sharing and User Agreement 88

List of Tables

Table 1. Descriptive Analysis - Years of Experience 60

Table 2. Descriptive Analysis - Level of Education 61

Table 3. Descriptive Statistics for the Results of the Test for H3…….63

Table 4. Descriptive Statistics for the Results of the Test for H4…….64

List of Figures

Figure 1. Indicator 1.2: Cognitively engages students in subject 48

Figure 2. Indicator 7.4 Monitors effect of instruction on individual and class learning 49

Chapter 1

Introduction

The teacher evaluation process has been a long-standing measuring tool to evaluate teacher quality in schools. However, school districts across the nation are in a constant struggle to determine what factors contribute to higher performance in the teacher evaluation process including a teachers' level of education or a teachers' years of experience. Ladd (2013) and Kini and Podolsky (2016) noted positive correlations between experienced teachers and student achievement. To the contrary, Darling-Hammond (2000) noted numerous research studies that showed a positive relationship between the teachers' educational level and their ability to impact student performance. The findings from Darling-Hammond's research suggest that pursuing advanced degrees could strengthen teacher knowledge in the content area and thus be impactful in improving student achievement. Since the research on this matter offers conflicting opinions on whether a teacher's level of education or a teachers' years of experience is more significant, a teacher evaluation tool that captures data on a teacher's education degree level and level of experience could offer answers on which is most important in the teacher evaluation process.

Marzano (2012) found that the most effective teacher evaluation models are those that have a twofold purpose. First, the evaluation tool must contain a measurable method to distinguish effective teachers from ineffective teachers. Second, the evaluation system must go beyond simply scoring teacher performance. Overall, the tool must be able to provide data that will improve

teacher quality. The evaluation instrument must be used to evaluate, measure, and identify areas of strengths and weaknesses to increase the number of highly qualified teachers in each classroom.

Hiring more experienced teachers has been noted as a valued method to improve student achievement. Darling-Hammond (2000) studied the relationship between teacher experience and student achievement. The results of her research were mixed. Novice teachers with less than three years of experience showed evidence of being less effective in improving student achievement. Surprisingly, her research revealed only marginally higher levels of effectiveness with more experienced teachers. There were several explanations for the marginal levels of effectiveness, which included experienced teachers seeking limited opportunities for professional development and less experienced teachers beginning their careers with a master's degree. Tran, H., & Buckman, D. G. (2020) also noted that a teacher's years of experience are only important within the first three to five years of teaching. After that, experience is less important because of the gains in content knowledge and classroom management techniques that teachers can quickly capture. Regardless of the reasons for limited growth, educational leaders must find a solution to ensure both novice and experienced teachers deliver quality instruction in the classroom.

Background

According to the National Archives and Records Administration (2016) publication, by 2013, most state leaders realized the school districts within their state would fail to meet many of the standards specified in President George

Bush's educational reform act, No Child Left Behind (NCLB), including the mandate that all teachers be deemed highly qualified by the 2014-2015 school year. In addition, The National School Board Association (2005) noted the importance of teacher evaluations as a significant measure to help states meet the rigorous demands of 100% mastery by all students. Klein (2015) noted that by 2014, no state had met the standards of No Child Left Behind. This failure prompted states to petition the U.S. Department of Education for waivers for failing to meet the provisions of NCLB and time to implement individual state plans to improve education. A condition of the No Child Left Behind waiver required states to implement a new teacher evaluation model. School districts could adopt their state's model or choose another model that met the state's standards.

A new teacher evaluation system was created in Missouri as a response to the failure of NCLB. The Network for Educator Effectiveness (NEE, 2015) began in 2011 at the University of Missouri College of Education as an assessment tool to "provide services for educator evaluation" and, according to NEE (2015), by 2014 over 220 school districts in Missouri chose the NEE model to evaluate teachers. There are five components to the NEE model:

- conducting formative assessments,
- a professional development plan,
- a unit of instruction plan,
- a student survey, and

- a summative assessment tool that provides both individual and comparative data to other teachers in the district and the state.

The NEE evaluation data includes scores for over 29 indicators that allow administrators and students to evaluate teacher qualities. However, this research study is focused on two of the indicators from the NEE model, student engagement and conducting formative assessments. The study is limited to these two indicators because the NEE evaluation tool requires a minimum of two indicators for administrators to use to evaluate teachers. Additionally, Hattie's (2011) research at the time identified student engagement and conducting formative assessment with high effect size as it pertains to improving student achievement. Therefore, school District A (the district used for this research study) chose student engagement and conducting formative assessment as the required indicators for all teacher evaluations.

In 2014, educational leaders in District A, a suburban school district south of Kansas City, Missouri, decided to adopt the NEE model as the teacher evaluation instrument. According to NEE (2021), this evaluation system exceeds the requirements for the NCLB waiver in that it goes beyond the requirements for the Missouri evaluation tool. District A chose the NEE evaluation model over the Missouri model because they believed this model is a superior tool in strengthening teacher growth and establishing more teachers as highly qualified. The NEE system tool focuses on teacher growth, not just teacher evaluation. NEE's comprehensive evaluation method includes mini-observations, professional development, student surveys, observation training for

administrators, rubrics for guidance, and a data storage function to measure and monitor teacher growth. In addition to the data for the indicators, districts can use the NEE computer program to collect data for years of experience and level of education for each teacher (NEE, 2021).

In 2015, School District A comprised approximately 17,834 students within three high schools, three middle schools, 19 elementary schools, one alternative school, and one special education center. There are 1434 certified teachers in the district. (Missouri Department of Elementary and Secondary Education [DESE], 2020). The student demographic makeup is 75% White, 12.2% Black, 3.2% Hispanic, 6.8% Multi-race, and 2.2% Asian. The percentage of students receiving free and reduced lunch is 19.2.

Most of the demographics for School District A are similar to the overall demographics of all school districts in Missouri. DESE (2020) reported in its annual report card on 879,661 students of 70.1% White, 15.5% Black, Hispanic 7%, Asian 2.1%, and Multi-race 4.6%. However, the state percentage of 49% of students receiving free and reduced lunch is much higher than District A. Nationally, NCES (2020) reported there were 56 million students enrolled in K-12 schools in the United States with 48% White, 15% Back, 27% Hispanic, 5% Asian, and 4% Multi-race.

DESE (2020) reported the average teacher had 16 years of experience in School District A and the percentage of teachers having earned an advanced degree was 80%. These statistics compare to Missouri's state average of 13 years of experience and 61% of all certified teachers having earned an advanced degree.

According to DESE (2020), the percentage of teachers nationally having earned advanced degrees in 2016 was 57% with an average of 14 years of teaching experience.

As a response to the failures of NCLB, school districts began revamping their teacher evaluation tool to focus more efforts on teacher quality. Improvements in teacher quality should equate to improvements in student performance. "Teacher evaluation systems are associated with teacher quality, accountability, performance observations, and support" Williams & Herbert (2020). According to the research findings from Daley and Kim (2010), Sartain et al. (2011), and Taylor and Tyler (2021) student assessment scores increased when their teacher was on a cycle for evaluation that year and when teachers were involved in a robust evaluation system. Hiring and maintaining highly qualified teachers is an expensive endeavor. Podolsky, Kini, Bishop, & Darling-Hammond (2016) reported a national cost of 8.5 billion dollars to replace nearly 50% of teachers who leave the profession within their first five years. As school districts continue to spend billions hiring and maintaining highly qualified teachers, district leaders could benefit from a method to determine if higher teacher pay and/or retaining experienced teachers positively impacts classroom instruction.

Statement of the Problem

School districts often boast about the years of experience and level of education their teachers possess. Nittler (2018) reported that 88% of all school districts pay their teachers more if they earn a master's degree. Educational leaders allocate countless resources to train and retain their best teachers to

increase their ability to improve learning. However, previous research on the topic produced mixed results on whether teacher pay is connected to student achievement. Podolsky et al. (2019) provided findings from their meta-analytic study that suggested teacher experience did have positive effects on student achievement if other factors like a healthy, supportive environment and previous years of teaching the same material in the same grade were also in place. Goldhaber and Anthony (2003) found that advanced degrees were only impactful in improving student achievement if the coursework for the advanced degree was directly related to the subject they were being taught. Despite this new ability to gather data on teacher effectiveness related to the teacher evaluation process, there has been no research to study the potential connection between teacher experience and level of education with teacher performance as measured during the teacher evaluation process. However, with the inception of the NEE model, school districts now have a newer, more efficient evaluation tool to measure teacher effectiveness and provide professional development opportunities that will help teachers improve student achievement.

Purpose of the Study

The purpose of this study was to determine if the NEE teacher evaluation scores were different than the threshold of (5) on the NEE indicators for student engagement and conducting formative assessment. NEE considers a mark of 5 or higher as adequately demonstrating the NEE indicator being evaluated. The second purpose of the study was to determine to what extent the NEE teacher evaluation scores were different on the student engagement indicator and the

conducting formative assessment indicator based on teachers' years of experience. Using data from the administrator evaluations and student surveys, the study was conducted to identify whether teachers with 0-5, 6-14, and 15 or more years of experience scored higher on the two NEE indicators: student engagement and conducting formative assessments. The third purpose of the study was to determine to what extent the NEE teacher evaluation scores were higher on the student engagement indicator and the conducting formative assessment indicator based on teachers' level of education. Teachers were divided into groups of those with advanced degrees versus those with a bachelor's degree to also determine how varying levels of education impacted scores on the two NEE indicators.

Significance of the Study

The study is significant because it could help answer the question of whether years of experience or degree matter in efforts to improve teacher quality in schools as it is measured during the teacher evaluation process. Educational leaders need clear data to determine if their teacher evaluation methods are effective. Most research studies regarding the level of teacher experience and education are related to student performance outcomes. Kini and Podolsky (2016) reviewed 30 studies that found teacher experience is related to student academic growth. Therefore, they recommended a focus on teacher stability to increase student achievement.

This study was focused on determining the effect of the teachers' years of experience and level of education as markers for increasing teacher effectiveness in the classroom. As school districts continue to spend millions hiring, training,

and maintaining highly qualified teachers, they need to know if those resources are correctly allocated. The results of the data, the teachers' summative evaluation scores, could help districts decide how best to allocate resources in their efforts to improve teacher effectiveness regarding increasing student engagement and conducting formative assessment.

Educational leaders could use this data to guide their decisions about adding teacher incentives for tuition reimbursement, salary increases for continued education, incentives for early retirement, and negotiation strategies to hire highly qualified teachers. As well, school districts have long offered additional pay for teachers with advanced degrees or continuing education hours in their related fields. Districts pay on average an extra $2760 more for teachers with master's degrees than those with just a bachelor's degree (Drake, 2008). This study can help school districts make important financial decisions in allocating resources.

Delimitations

The delimitations established by the researcher were to determine how varying levels of experience and education impacted teacher evaluation scores. Lunenburg and Irby (2008) noted, "Delimitations are self-imposed boundaries set by the researcher on the purpose and scope of the study" (p. 134). The first delimitation used by the researcher was to limit the study to the three high schools in School District A. The second delimitation was to review the NEE Teacher Evaluation data for only those teachers who had six or more observations during their evaluation year between 2015-2018. Data for teachers who were not on a

cycle for an evaluation or who had fewer than six evaluations were not included. The third delimitation focused only on two of the 29 NEE indicators used to measure teacher effectiveness. The two indicators used in this study were Indicator 2.1 for student engagement and Indicator 7.4 for conducting a formative assessment. The final delimitation centers on the categorization of the teachers' years of experience and level of education. The teachers' years of experience were categorized in increments of 0 – 5 years, 6 – 14 years, and 15 years or more. The teachers' educational levels were categorized into groups of bachelor's degree and master's degree.

Assumptions

Lunenburg and Irby (2008) defined assumptions as "postulates, premises, and propositions that are accepted as operational for purposes of the research" (p. 135). The assumptions for this study included (a) the administrators were effectively trained to score the teachers accurately on the NEE indicators and avoided bias when assigning scores, and (b) the district data on years of experience and level of education were accurate.

Research Questions

The following research questions were used to conduct the research.

RQ1. To what extent are NEE scores on the indicator for student engagement different from the threshold value of 5?

RQ2. To what extent are NEE scores on the indicator for conducting formative assessment different from the threshold value of 5?

RQ3. To what extent is there a difference in scores on the NEE indicator for student engagement based on the teachers' years of experience?

RQ4. To what extent is there a difference in scores for the NEE indicator for conducting formative assessment based on the teachers' years of experience?

RQ5. To what extent is there a difference in scores for the NEE indicator for student engagement based on the teachers' level of education?

RQ6. To what extent is there a difference in scores for the NEE indicator for conducting formative assessment based on the teachers' level of education?

Definition of Terms

Lunenburg and Irby (2008) explained the importance of defining all key terms related to the research study. The key terms allow the readers to have a collective understanding of the essential components needed to understand the study.

AYP. Adequate yearly progress (AYP) require schools to meet their student performance achievement goals each year to qualify for Title I funds under the No Child Left Behind Law in 2002 Klein, 2015).

Classroom Observation. "A classroom observation is a formal or informal observation of teaching while it is taking place in a classroom or other learning environment" (Great Schools Partnership, 2014a, para. 1).

Evaluation. "Evaluation is typically based on one or more key questions the evaluator needs to pursue, and the collection and analysis of relevant data to help answer those questions. Results derived from the data then are used to inform decisions" (Robinson, 2018, para. 11).

Formative assessment. Formative assessment refers to a wide variety of methods that teachers use to conduct in-process evaluations of student comprehension, learning needs, and academic progress during a lesson, unit, or course. (Great Schools Partnership, 2014b).

Feedback. Feedback in the classroom can be defined as "information allowing a learner to reduce the gap between what is evident currently and what could or should be the case" (Hattie & Yates, 2013, para 2).

Summative Evaluation. "The summative conference that occurs between the school leader and the teacher. The actual feedback, planning, and support that occur during the summative conference set the tone for a teacher's continued professional growth journey" (NEE, 2021 p. 1).

Student engagement. In education, student engagement refers to the degree of attention, curiosity, interest, optimism, and passion that students show when they are learning or being taught, which extends to the level of motivation they have to learn and progress in their education (Great Schools Partnership, 2018).

Organization of the Study

This research study is organized into five chapters. Chapter 1 included the introduction, background, statement of the problem, the purpose of the study, delimitations, assumptions, research questions, the definition of terms, and the overview of the methodology. Chapter 2 focuses on the review of the literature. The elements of Chapter 3 include the research design, population and sample, sampling procedures, data analysis, hypothesis testing, and the limitations of the

study. Provided in Chapter 4 are the results of the research study. Chapter 5 details a summary of the study, the findings related to the literature, and the conclusion.

Chapter 2

Review of the Literature

School leaders are on a constant search to improve student achievement in America. However, they disagree on which teacher credentials are most important to improve student outcomes. Numerous studies have identified teacher experience as the most important factor (Clotfelter, Ladd, & Vigdor, 2007). Others consider advanced levels of education as a crucial element in improving student achievement (Darling-Hammond, 2000; Hanushek, O'Brien, Kain & Rivkin, 2005). The findings of these studies have produced mixed results.

History of Teacher Evaluation in America

The concept of teacher evaluation systems is not a new phenomenon in American education. Marzano, Frontier & Livingston (2011) provided a comprehensive summary of the history of teacher evaluations. In America, teacher evaluation can be traced back to the 1700s when clergy members were charged with hiring, firing, and evaluating teachers. Clergy members had unlimited authority to evaluate teachers using their own set of criteria. Furthermore, Tracy (1995) described how any community leader (merchants, those holding professional degrees in their field) could also be in charge of decision making in local schools. Their leadership over schools included hiring/firing teachers, dictating curriculum, and discipline decisions for students. Classroom observations could be performed by any number of community members to evaluate student progress, teacher methodology, and the structure and organization of the classroom (Tracy, 1995). Marzano et al. 2011 reported that by

the 1800s, larger urban schools became more prevalent and school systems looked for teachers with specialized skills. Within each school, a "principal" teacher governed the other teachers in the building. During this era, teacher supervisors were expected to be trained in core subject areas. As the century progressed, school leaders used teacher evaluation as a tool to improve teacher instruction.

Marzano et al. (2011) noted a sharp difference in teacher performance, evaluation and student achievement in the early 1900s. Schools took on a more scientific approach in that there should be a connection between learning in school and productivity as an adult. A greater importance on citizenship, democracy, factory skills, and problem-solving skills infiltrated the school system. Cubberley (1916), a noted educator of the time, believed students in school could be managed like factory workers by collecting data on an individual's performance, and then analyzing the data to determine a plan of improvement. Building upon that idea, Wetzel (1929) believed student learning should be a primary factor in measuring teacher effectiveness. He recommended three components as the basis for scientific supervision: the use of aptitude tests to determine the ability level of each child; the establishment of clear, measurable objectives for each course; and the use of reliable measures of student learning" (n.p.). These standards called more attention to the student's ability to perform well on standard tests. As the first half of the century concluded, there was much division over how to evaluate a teacher effectively. The opposing philosophies of citizenship vs. democracy vs. data on test results led to a contentious period of developing the best criterion to evaluate teachers.

During the latter half of the 20th century, clinical supervision became the most dominant method for evaluating teachers. Goldhammer (1969) stated, “The process, symbiotic relationship between practitioner and resident, where observation and discussion drove both parties to higher levels of growth and effectiveness” (p 54). The goals of clinical supervision were for teachers to constantly pursue professional development to increase their impact on student learning. Supervisors would then use the five phases of the evaluation tool to score the teacher’s effectiveness on student learning. Goldhammer (1969) described this evaluation tool as a checklist of steps the teacher completed.

- Phase 1 - Pre-observation Conference: This phase was designed to provide a conceptual framework for the observation. During this phase, the teacher and supervisor planned the specifics of the observation.
- Phase 2 - Classroom Observation: During this phase, the supervisor observed the teacher using the framework articulated in Phase 1.
- Phase 3 - Analysis: Data from the observation was organized by the supervisor with the intent of helping teachers participate "in developing evaluations of their own teaching" (p. 63).
- Phase 4 - A Supervision Conference: The teacher and supervisor engaged in a dialogue about the data. The teacher was asked to reflect upon and explain his or her professional practice. This

stage also could include providing "didactic assistance" (p. 70) to the teacher.

- Phase 5 - Analysis of the Analysis: The supervisor's "practice was examined with all of the rigor and for basically the same purposes that Teacher's professional behavior was analyzed therefore" (p. 71).

In summary, the clinical supervision model merged the technical skills of gathering and analyzing data with the soft skills of relationship building between the teacher and administrator.

During the 1970s, another shift in public education emerged as American educators looked for a better model to produce highly qualified teachers. Elam (1971) described how Performance-Based Teacher Education (PBTE) started as a paradigm shift from experience-based to performance-based. The author noted how the model of experience based simply required teacher candidates to experience a certain set of classes and earn satisfactory grades. However, the new shift to performance based focused more on teacher candidates being able to demonstrate their ability to impact student learning and actualization of skills at an acceptable competency level. Evidence of a competent teacher moved from passing required teacher preparation classes to being able to produce quality work in students.

The Hunter model, created by Madeline Hunter, was a prominent evaluation system of the 1980's (Hunter, 1982). This system was focused on a seven-step model for teachers to mimic for each lesson. They included the

anticipatory set, objective and purpose, input, modeling, checking for understanding, guided practice, and independent practice. The anticipatory set focused on activities to get the students thinking about what they were going to learn. Objective and purpose centered on making sure the students knew the learning target for the lesson so they knew what they were learning and why. Input creates the opportunity for the teacher and student to anticipate the learning and final objectives of the learning. Modeling focused on the teacher demonstrating several ways to acquire the knowledge for the objective. Checking for understanding activities helps the teacher determine if the students are learning the objectives. Guided practice involves the teacher providing feedback as the students are learning to ensure they are using the correct skills. Finally, independent practice activities provided students with the opportunity to demonstrate their knowledge on their own. During the teacher evaluation process, the supervisor would look for evidence of these 7 steps, then hold a post - conference with the teacher to discuss the results (Hunter 1972; Marzano et al. 2011).

Despite the popularity of the Hunter model in the late 20th century, the results were mixed in terms of the impact of the Hunter model on student achievement. Stallings and Krasavage (1986) reported an increase in student achievement in reading and math scores from 1982 to 1984 on the Instructional Skills Observation Instrument when teachers used the Hunter model. However, scores in those same areas dropped in the subsequent years of 1985 to 1985. Scores for English Language learners increased in the areas of student

engagement and achievement rates. However, there was no long-term evidence that showed teachers who implemented the Hunter model produced students with higher achievement levels. Cuban (2019) noted three main criticisms of Madeline Hunter's methodology. First, there was too much focus on direct instruction and not enough emphasis on student ownership of their education. Second, this model eliminated the concept of student choice. Instead, student learning was more robotic and devoid of student choice. Third, the teacher-centered focus produced students who mimicked their teacher's philosophy instead of encouraging students to pursue their own self-interest.

As the 20th century was coming to an end, another shift occurred in teacher evaluation systems. Danielson (2007) created an evaluation model that merged the concepts of teacher accountability and professional development. What started in 1996 as a program to develop first-year teachers, the Danielson model grew in popularity and "the Framework quickly found wide acceptance by teachers, administrators, policymakers, and academics as a comprehensive description of good teaching, including levels of performance—unsatisfactory, basic, proficient, and distinguished—for each of its 22 components" (Danielson, 2013, p. iv). The Danielson model consisted of 4 domains: planning and preparation, classroom environment, instruction, and professional responsibilities. Danielson recognized the complex components needed for effective teaching and learning; thus, he created a model to measure all of those components of student achievement.

History of National Policies on Teacher Evaluation Systems

During the late 20th century and early 21st century, the role of the national government began to heavily influence teacher evaluation systems and the correlation to student achievement. Three policies that defined this area were A Nation at Risk, No Child Left Behind, and Race to the Top.

A Nation At Risk. In 1983, A Nation at Risk revealed the need for improvement in American schools. Ruff (2019) studied teacher evaluation systems that stemmed from A Nation at Risk and the transition to the era of No Child Left Behind. The purpose of Ruff's study was to conduct a comparative analysis of the accountability policy of two state systems, the Virginia Standards of Learning and the Nebraska School-based Teacher-led Assessment and Reporting System. The research was an analysis of the cultural, historical, and political influence in developing a tool for measuring student achievement. Data from the Nebraskan School-based, Teacher-led Assessment & Reporting System (STARS), a low-stakes test, was compared to data from Virginia's Standard of Learning (SOL) high-stakes tests.

No Child Left Behind. Klein (2015) noted the impact of No Child Left Behind. Created in 2001 as part of House Resolution 1 under the presidency of George W. Bush, No Child Left Behind was designed to provide educational opportunities that would allow all students to succeed regardless of race or socioeconomic status. NCLB mandated that all students must be proficient according to the educational standards of that state by the 2013/2014 school year. Klein (2015) described how states mandated testing math and reading in grades

3rd and 8th, and at least once in high school. Schools had to demonstrate Adequate Yearly Progress (AYP) by submitting performance data to show their students' progress in meeting the annual achievement targets for all student populations. The spotlight was on subgroup populations in the areas of racial minority groups, students in English Language Learners and special education programs, and students of poverty in all racial groups.

Klein (2015), Chen (2019), and DeAngelis, White, and Presley (2010) detailed the numerous implications if schools did not meet AYP. Students who failed to meet proficiency could be eligible to move to another school within the district or receive free tutoring. Failing schools would also have to reserve 10% of their Title I budget for school choice vouchers. In addition, the school district could receive disciplinary action or state takeover and lose federal Title I funds. However, if schools met the proficiency standards, they would be eligible for additional funding.

Another mandate of NCLB, enacted in 2002-2003, required school districts to only employ highly qualified teachers by the end of the 2005-2006 school year. Chen (2019) and DeAngelis et al. (2010) defined highly qualified teachers as those with a minimum bachelor's degree and certification to teach. In addition, classified staff who served as paraprofessionals were required to have completed at least two years of college, an associate's degree, or pass a proficiency test to demonstrate their competency to work with students. Chen (2015) further noted that if there was an achievement disparity between social-

economic levels of the schools within a school district, the highly qualified teachers had to be equitably distributed between wealthy and poor schools.

DeAngelis et al. (2010) conducted research to study multiple factors that constitute qualified teachers. They further defined NCLB's definition of a highly qualified teacher. In particular, they evaluated the changes the state of Illinois had made in creating highly qualified teachers over the six years between 2001 and 2006. Their research focused on three questions: at what level did teacher qualifications change, what changes occurred during the distribution of teacher qualifications across the varying diversity levels of schools within the district, and what defines the changes in teacher qualifications in Illinois? Their data sources included files from the Teacher Service Record that gathered information on teachers' years of experience, name of school, position within the school, and hours employed. They also used data from the Illinois State Board of Education to collect information on teachers' certifications and levels of education. The results of the study by DeAngelis et al. (2010) indicated that highly qualified teachers made the most significant improvements in schools with high minority populations and high levels of poverty. Increased recruitment efforts for high minority/high poverty schools and the use of alternative route programs were key factors in increasing the percentage of highly qualified teachers. In addition, the data showed that inexperienced teachers were often more qualified than experienced teachers so DeAngelis et al. concluded more focus should be placed on them.

Every Child Succeeds Act. As the second decade of the 21st century began, it became evident that states would not meet the mandate that 100% of the students were proficient in state standards. In addition, there were no solutions on how to solve the inadequacies of NCLB. As a remedy, President Barack Obama introduced a plan to restore state control in improving schools but implemented new mandates for students to perform at high levels as well as creating an improved teacher evaluation system tied to student outcomes. President Obama's work would eventually become the Every Student Succeeds Act on December 15, 2015 (U.S. Department of Education, n.d.).

According to the U.S. Department of Education (2013), 34 states and Washington D.C. applied for waivers to gain additional time to reform their plan to increase student achievement. Klein (2015) reported an increase to 42 states plus Puerto Rico as more states failed to meet the NCLB mandates. President Obama would eventually introduce waivers from NCLB requirements. The U.S. Department of Education noted, "the Administration would provide State Education Agencies (SEAs) with flexibility regarding specific requirements of NCLB in exchange for college- and career-ready expectations for all students; differentiated accountability, including targeting the lowest-performing schools, schools with the largest achievement gaps, and other schools with performance challenges for subgroups; and teacher and principal evaluation and support systems that take into account student growth and are used to help teachers and principals improve their practices." ESSA (2015).

Aragon (2018) summarized the new flexibility of ESSA laws, stating federal incentives to create prescribed evaluation systems have been removed and states now have full discretion over whether and how to evaluate teachers. However, ESSA still requires states to demonstrate that they are working to provide disadvantaged students equal access to effective teachers. The guidance for new teacher evaluation systems required states to adopt education redesign in the following areas:

1. develop evaluation systems with continuing educator input;
2. provide clear, timely, and useful feedback;
3. improve instruction; use multiple measures, including student growth;
4. differentiate performance; and
5. inform personnel decisions. (Aragon, 2018, para 3)

Race to the Top. The Reform Support Network (2012) noted a shift in education. Beginning in 2009, Race to the Top (RTTT), another federal education program of the Obama administration, offered states the opportunity to compete for federal grant money if they revised four critical areas in education: a focus on college and career readiness, a focus on collecting student data to measure growth and make improvements, a focus on establishing and maintaining highly qualified teachers, and a focus on improving the lowest performing schools. (National Archives and Records Administration, 2016). Education First (2023) summarized how the state of Georgia called upon The Reform Support Network to collect data on states that had revamped their teacher evaluation systems. The Reform Support Network conducted a study based on the state of

Georgia's quest to analyze the classroom observation policies from various states after the initiation of Race to the Top newly founded policies for classroom observations. The findings from the study included commonalities between the state evaluation systems such as conducting multiple formative teacher evaluations during the school year, adjusting the number of evaluations based on the teacher's experience level, and requiring the evaluator to share feedback with the teacher after the evaluation. According to the outcomes of the study, Education First (2023) wrote, "States are poised to better implement higher quality evaluation systems to improve teacher development, instruction and, ultimately, student learning" (np).

The research of Close, Amrein-Beardsley, and Collins (2019) involved the study of the changes that states made to their teacher evaluation systems after the federal government adopted the ESSA in 2016. Regarding the highly controversial use of VAMs (Value Added Measure), 15 states reported continued usage of this tool while 23 states reported they had discontinued use of this VAMs. Likewise, other handful of states informal used VAMS manner to collect data and provide teacher feedback for personal growth. According to the researchers, about 55% of statements are now using some form of Student Learning Objectives (SLOs). SLOs are objectives that teachers create to measure student growth for the school year. This step gives teachers more ability to measure the growth of students while the students are being taught by that particular teacher. Another theme after the passage of ESSA was the continued use of teacher observations as part of the teacher evaluation system. Over 70% of

states reported teacher observations as a dominant component of the teacher's summative evaluation.

Controversy Surrounding Teacher Evaluations

Over the years, both administrators and teachers have argued about the effectiveness of teacher evaluation systems. Some believe they are viable tools that appropriately measure the teacher's performance in the classroom while others view them as biased and not a true reflection of the teacher's impact on student learning. Callahan and Sadeghi (2015) studied the perceptions of teacher evaluations in New Jersey. The researchers believed teacher evaluation systems should measure the teacher's strengths and weaknesses fairly and consistently. Afterwards, professional development becomes the improvement tool used to both strengthen and/or correct any areas of concern. They noted that evaluation systems have too many flaws including a dearth of activities to improve poor performance, a lack of timely feedback, and no connections to professional development and standards in the evaluation tool. Callahan and Sadeghi (2015) ultimately felt teacher evaluations fail in their attempts to improve teacher quality because they are too short, too infrequent, and do not differentiate between the varying levels of teacher experience. Furthermore, rubrics were deemed ineffective because they focused on a checklist of what was observed instead of practical skills to help the teachers.

Callahan and Sadeghi (2015) also found that teachers felt more stress and less satisfaction in the profession. Instead of focusing on a checklist of standards, the researchers believed there should be a greater emphasis on the "significant

relationship between teacher effectiveness and student learning" (Callahan and Sadeghi (2015). Most teacher evaluation systems ignore the teacher-student relationship factor as it relates to student performance. Not only is there little effort to measure good relationships, but there is also no correlation between high scores on teacher evaluations and low scores on student performance data. For example, Brill (2009) found that 97% of teachers scored in the top two levels on their evaluation summary, but 25% of students do not graduate on time and 33% of the graders are below grade level in reading. Clearly, there is a discrepancy between the high marks teachers receive on performance evaluations and how their students perform academically.

Warring (2015) noted a host of negative perceptions about the teacher evaluation system. First, most evaluation tools do not have a correlation between teacher evaluation scores and student achievement. In other words, the teacher may score high on all indicators on the evaluation, but their students may have failing grades or low assessment scores. Closely related to this point is the practice of students' test scores being the main factor in a teacher's evaluation score. Numerous school districts including Dallas, Houston, New York, and Washington D.C. are short-sighted in this type of evaluation method because not all grade levels and subjects require state testing. Therefore, some teachers receive the benefits or detriment of this practice while others do not. This method also led to teachers being evaluated on students whom they have not taught. In addition, teachers who taught testing subjects received even more criticism when parents accused them of teaching too much to the test and ignoring other

important curriculum materials. Overall, in many cases, too much of the teacher's evaluation score was based on student performance results, which does not reflect the full scope of the teacher's teaching ability.

There have also been reports of bias in regard to teacher evaluations. Bias may be apparent in several forms. Warring (2015) found that teachers with higher percentages of lower functioning students could automatically have lower ratings on student performance-based evaluations. In addition, their research found that high-performing students typically rate teachers higher on student surveys than lower-performing students. To counter bias ratings connected to demographics, some systems allow adjustments in the evaluation scores. This attempt to counter the demographic bias could then create an unfair score compared to teachers who had no adjustments in their evaluation scores.

Punyanunt-Carter and Carter (2015) also noted several types of bias in teacher evaluations when there is a student survey component. Some evaluation systems have a student survey component that allows students to score their teachers on the same standards that supervisors use to score teachers. Several factors impact those student scores. The research showed that class size is one such factor. Students in smaller classes have greater opportunities to receive more personalized instruction. This could positively affect the teacher's evaluation. Unfortunately, the opposite could also be true. Students in larger classes, who receive less teacher support, could score the teacher lower on the student surveys.

In the 21st century, there has been a steady flow of criticism of teacher evaluations. Toch and Rothman (2008) heavily criticized the teacher evaluation system. Instead of focusing on instructional effectiveness and student achievement, most evaluation tools focus on a checklist of steps to include in a lesson. In addition, they ignore measures to determine student learning and growth. Their research also found only 14 states required school districts to perform yearly teacher evaluations. In another study, Weisber, Sexton, Mulhern, & Keeling (2009) described the failure of teacher evaluation practices in a term called the widget effect. This is the practice of measuring teacher effectiveness using the same standards for each teacher regardless of years of experience, level of education, strengths, weaknesses, and previous instructional growth. The researchers believed this method of conducting teacher evaluations removed the individualization of teachers. Therefore, teachers seldom received individualized feedback that would help improve or address any instructional concerns. The researchers noted additional flaws in teacher evaluations such as being too short to measure true teacher effectiveness and too infrequent to capture the full realm of the teacher's teaching style and ability. In addition, they felt school leaders were under trained in this area and could be susceptible to school climates of who was a good teacher or not.

Researchers have noted other flaws with teacher evaluation systems in terms of the limitations of what the tool actually measures. Too often the evaluation tool focuses on the teacher following a set of steps to teach a lesson. Cohen and Goldhaber (2016) found that most evaluation tools ignore various

ancillary components related to student growth. For example, school leaders have charged teachers to include more lessons on social-emotional learning. However, there are little to no indicators of teacher evaluation tools that measure a teacher's effectiveness to improve this concern for students to ultimately improve student performance. In addition, teachers have been charged with identifying cultural bias and improving relationships with students of diverse cultural backgrounds. Likewise, teacher evaluation instruments do not measure a teacher's growth in this area. Finally, as more schools emphasize the importance of developing a positive climate in the classroom to help students foster a greater sense of belonging, this is another ignored area that can impact student learning.

Traditionally speaking, numerous factors contribute to a student's academic performance including the student's innate ability, parents, community, access to resources, previous teachers, etc. Cohen and Goldhaber (2016), Hanushek et al. (2005) noticed the absence of value-added measures (VAM) in teacher evaluations. VAM are "growth measures, are used to estimate or quantify how much of a positive (or negative) effect individual teachers have on student learning during the course of a given school year" (para 1). Proponents of VAM have devised a method to extrapolate an individual teacher's impact on student growth using a sophisticated algorithm to single out a single teacher's impact on a student from other factors that may contribute to student learning. Cocoran (2010) defined value added as "the unique contribution she (teacher) makes to her students' academic progress. That is, it is the portion of her students' success (or

lack thereof) that cannot be attributed to any other current or past student, school, family, or community influence." (p 4).

Highly Qualified Teachers

The Department of Education (2006) published the Missouri Revised State Plan for Highly Qualified Teachers for 2006-2007. The document detailed the following requirements to define a highly qualified teacher in the state of Missouri:

- The teacher held a bachelor's degree in the subject taught
- The teacher has full Missouri certification to teach in the state
- The teacher has passed the subject matter competency test for the academic subject taught
- The teacher's certification has not been waived or altered with lower expectations in any manner

The document also outlined how schools had to identify areas where they were deficient in the areas of highly qualified teachers (NQT). This included detailing the number of teachers who were NQT, the steps in place to rectify the problem, and technical assistance programs to assist the school district in meeting the mandates. Missouri's requirements for highly qualified teachers also noted the value of experienced teachers (those with five or more years of teaching) in each school district, especially school districts with high poverty levels. According to DESE (2023), in addition to earning the initial teaching certification in Missouri,

teachers must renew their licenses after four years. The renewal process includes a minimum of 30 hours of professional development or earning a master's degree.

Teacher's years of experience. Researchers have found a correlation between a teacher's years of experience and student achievement. In a meta-analysis study, Kini and Podolsky (2016) found that a teacher's years of experience had a positive relationship with student achievement. In a review of 30 studies, the results of 28 determined the longer a teacher taught, the higher their students scored on standardized tests. In addition to academic achievement, students also showed improvement in school attendance and classroom behavior. Kini and Podolsky reported, "the North Carolina study found that one year of experience allowed an ELA teacher to reduce the proportion of students with high absenteeism by 2 percentage points", and that a teacher "who obtains over 21 years of experience on average reduces the incidence of high student absenteeism by 14.5 percentage points." Moreover, the researchers also found that the teacher's effectiveness improved even faster if the teacher worked with other experienced teachers and if they were in a positive and supportive school culture.

Clotfelter et al. 2007) also noted the importance of teacher experience in student achievement. However, their research also included certification test scores and teacher licensure as significant credentials that influence student achievement. The longitudinal study used math and reading achievement scores from Grade 3 to Grade 5 students in North Carolina from 1995-2004. The authors found that teachers' years of experience were a factor in achievement returns, but only within the first few years. They also found that board-certified teachers

disproportionately taught in affluent school districts with predominately-white students, which ultimately led to those students making greater gains in student achievement. Finally, the researchers found that a teacher's level of education positively influenced math scores, yet only marginally influenced reading scores.

Rockoff, Jacob, Kane, & Staiger (2008) studied a broader range of nontraditional teacher characteristics as they related to student achievement. The study included first-year math teachers from the elementary and middle school levels in New York. During the 2006-2007 school year, the researchers gathered data from a survey that "collected information on a number of non-traditional predictors of effectiveness including teaching specific content knowledge, cognitive ability, personality traits, feelings of self-efficacy, and scores on a commercially available teacher selection instrument" (Rockoff et al., 2008, p 1). Teacher performance scores on the Haberman PreScreener were also included in the data collection. The findings indicated that there is value in using nontraditional criteria as a method to identify effective teachers during the hiring process.

Teacher's level of education. School teachers earning master's degrees is not a new phenomenon. The National Center for Educational Statistics (2021) reported that the number of teachers with a master's degree or higher has increased from 47% in the 1999-2000 school year to 58% in the 2017-2018 school year. Many states allow teachers to use post-bachelor's credit to meet recertification requirements. In addition, many school districts incentivize teachers to earn hours or degrees beyond their bachelor's degree. However, the

question remains whether an increase in educational levels positively impacts teacher quality and student achievement.

The findings from Horn and Jang (2017) included mixed results on the correlation between a teacher's level of education and student learning. At the middle school level, the findings were miniscule in reading and inconclusive in math. At the high school level, there were more promising findings with conditions. In the areas of math and science, there was increased student achievement if the teacher's master's degree was in the area of math or science. In other core subjects such as English and reading, there was no significant improvement if the teacher held an advanced degree. Horn and Jang concluded that more research was needed to determine how many teachers held advanced degrees in the subjects they taught and if this factor could lead to higher academic achievement for students.

Selke (2001) studied the effect of degree type on student achievement. This study was conducted using data collected from research in the late 1990's reviewing a nationwide perception of the teachers with master's degrees. The purpose of the study was to examine the three types of master's degrees and their impact on teacher effectiveness. The research evaluated the masters of teaching degree, the traditional master's degree program, and the practitioners' master's degree program. The researcher used three main questions to guide his study. Using a qualitative approach, the researcher gathered data on the teacher's perceptions of the three degrees and the reasons why the teachers chose one degree over the other. The results indicated the practitioners' master's degree

program was most beneficial for teachers choosing to remain in the classroom as opposed to those who sought building-level leadership or central office positions. This degree was better suited to help practitioners become teacher researchers constantly looking for strategies to improve student performance and self-efficacy.

Encouraging teachers to pursue advanced degrees has been another widely used method to improve teacher quality and student achievement. Selke (2001) studied teachers who pursued three types of master's degrees: masters of teaching degree, the traditional master's degree program, and the practitioners' master's degree program. He noted in his research that a master's degree was not a factor that automatically improved teacher quality and student performance. The researcher noted the most significant gains were from teachers who pursued the practitioner's master's degree and chose to remain in the classroom as career teachers. For years, many school districts have required teachers to pursue advanced degrees or other continuing education programs. However, before educational leaders issue a mandate requiring all teachers to pursue an advanced degree, they must first study the effective size of this action.

Hanushek, O'Brien, Kain & Rivkin (2005) noted the importance of teacher quality as a means to improve student achievement. Their research identified advanced degrees, teacher certification scores, and levels of experience as key concepts to that achievement. The research was conducted in a large school district in Texas in 2005 using data from the Texas Schools Microdata Panel and achievement data from the Texas Assessment of Academic Skills

(TAAS) test. The purpose of the research was to determine how much teacher quality affected student achievement. The district felt they were losing quality teachers from their urban school district to higher paying jobs in other districts. If highly qualified teachers continued to leave the district, student achievement could stall or decline. The researcher used three main questions to guide the study. Using an empirical model, the researchers gathered data to identify the impact of experienced vs. non-experienced teachers on student achievement. The results of the data revealed there is little evidence to support the claim that urban schools suffer because of a lack of teachers with experience. While there was a drop-in achievement for students taught by first-year teachers, teacher experience was not a dominant factor in student achievement.

History of The Network of Effective Educators (NEE)

The Network for Educator Effectiveness was created in 2011 by Doss, a former educator and superintendent, and Bergin, associate dean for research and innovation at the University of Missouri College of Education. Doss (2016), shared the research used to create a teacher evaluation tool that Missouri school districts could use in place of the state model. The foundational research used to guide the NEE model evolved from *A Practical Guide to Designing Comprehensive Teacher Evaluation Systems* (Goe et al., 2011). The purpose of the study was to show school districts how to create a teacher evaluation system that had the dual purpose of identifying highly qualified teachers and providing supervisor and student data, which teachers could use to create professional development plans for the next school year. Goe et al. (2011) also noted the

failures of previous teacher evaluation systems that resulted in the Widget Effect. When districts across the country failed to meet the mandates of NCLB, they had to adhere to new mandates of the Elementary and Secondary Education Act of 2011 that required more emphasis on hiring and developing highly qualified teachers. Therefore, new teacher evaluation systems under ESSA had to address these issues. According to Goe et al., states and local school districts used the following 8-step guide to begin the process.

- Component 1: Specifying Evaluation System Goals focused on helping organizations determine how to create goals that meet the needs of the district, how to measure and model the goals, and how to align the goals with the state standards.
- Component 2: Securing and Sustaining Stakeholder Investment and Cultivating a Strategic Communication Plan involved making sure organizations created an evaluation system that a broad range of stakeholders would support. That range of stakeholders goes beyond the traditional set of teachers, administrators, school board, and community members. The support would also need to come from state teacher unions and teacher preparation programs.
- Component 3: Selecting Measures focused on choosing the right measures to evaluate teacher growth.
- Component 4: Determining the Structure of the Evaluation System involved creating a system that determined how many measures and how

much weight each measure should have in terms of determining teacher proficiency.

- Component 5: Selecting and Training Evaluators focused on creating a system where administrators would have to be trained and receive certification before evaluating teachers.
- Component 6: Ensuring Data Integrity and Transparency involved developing an infrastructure that would gather, disaggregate, and analyze data with fidelity.
- Component 7: Using Teacher Evaluation Results involved determining how to use the data to create the summative report and make goals for a professional development plan.
- Component 8: Evaluating the System involved analyzing the system to determine if it actually improved its original goals. (p. 7)

Doss (2016) also noted the influence and research of Marshall (2009) as a baseline standard of protocols that must be included in classroom observations. The work of DeNisi and Sonesh (2011) was instrumental in helping the NEE creators develop a method of evaluating evaluation data to help teachers create a professional development plan based on the evaluation data from their summative evaluation. The summative evaluation provides teachers with an average of the indicator scores from the administrator's observations over the school year. Teachers could use indicator scores that revealed areas of concern or areas where they wanted to continue to grow professionally as the rationale for choosing a particular professional development topic for the next school year. This practice is

in accordance with the research of DeNisi and Sonesh (2011), which studied the importance of using data when setting new professional goals.

One main indicator of teacher effectiveness is student engagement. Griffey and Housner (1991) found significant differences between an experienced teacher's ability to impact student engagement vs an inexperienced teacher's ability to impact student engagement. Experienced teachers were shown to have more advanced lesson plans that promoted higher levels of student engagement. Before planning a lesson, they tended to contemplate student struggles with the lesson. This step allowed for more variety in questioning and follow-up questions during the lesson, which resulted in higher levels of student engagement.

Teachers are often evaluated on their ability to cognitively engage students in the learning process. (EdhHub personal communication, May 6, 2020), the online resource library for NEE, expounded upon the definition of student engagement in Indicator 1.2 (Cognitively engages students in the content.) as, "The teacher cognitively engages students in the content. Cognitive engagement in the classroom refers to students' active mental involvement in the learning activities or mental effort, such as meaningful processing, strategy use, concentration, and metacognition. Cognitive engagement is different from behavioral engagement, which is cooperative participation, or adhering to classroom rules. Cognitive engagement is a key goal of many school reform efforts because it predicts achievement." (EdhHub personal communication, May 6, 2020)

In the classroom, this might look like:

- Teacher incorporates appropriate learning and instructional strategies to encourage deep thinking
- Teacher supports students in monitoring their own levels of cognitive engagement
- Teacher recognizes if some students are not cognitively engaged, and tries alternate strategies to increase or maintain students' thinking about content
- Teacher uses cognitive engagement strategies such as advanced organizers, K-W-L charts, share-out, shoulder-partner work
- Teacher cognitively engages students so that they are active in the lesson or activity
- Teacher is able to build activities appropriate for all depth of knowledge levels
- Teacher assesses student understanding often. (EdhHub personal communication, May 6, 2020)

Park (2005) published research that showed a positive correlation between student engagement and monthly student growth in math for a variety of demographic groups including low socioeconomic status, race, and gender. Finn (1993) noted the importance of student engagement as early as elementary associated with students becoming at risk of dropping out of school. Harbour, Evanovich, Sweigart, and Hughes (2014) marked student engagement as one of the most important teacher strategies that impact student engagement.

Furthermore, Hattie's 2008 meta-analysis of educational tools listed student engagement as a top five strategy.

NEE (2020), expounded upon the definition of formative assessment in Indicator 7.4 as "Monitors effect of instruction on individual and class learning. This indicator addresses the teacher's ability to monitor the effect of instruction on individual students and the whole class. It is about formative assessment of a particular kind. Formative assessment has multiple meanings, but in NEE we use the term to refer to quick checks for understanding as the lesson is progressing. The purpose is to inform the modification of teaching and learning activities in real-time. Thus, it is information used to guide instruction as part of the instructional process. Questioning is the most common form of this kind of formative assessment. However, other kinds of formative assessment might include solving problems on a whiteboard or answering spot quizzes with fist-to-five, thumbs up, or clicker techniques.

In the classroom, this might look like:

- Teacher monitors the learning of the whole class and many individuals
- Teacher uses multiple checks for understanding
- Teacher engages in effective formative assessment
- Teacher monitors learning progress
- Teacher uses assessment for learning
- Teacher uses systematic monitoring of learning progress

- Teacher uses strategies such as questioning, whiteboarding, thumbs up, fist-to-five, observing student work, etc.
- On-the-spot assessment is seamless throughout instruction
- Strong, appropriate corrective action is taken to ensure learning of almost all students. (EdHub, personal communication, May 6, 2020).

Black and Wiliams (1998) noted the importance of formative assessment as an effective teaching strategy. Their review of formative assessment has been cited over 1000 times as an effective teaching strategy to improve academic achievement. Apthorp, H., Klute, M., Petrites, T., Harlacher, J., & Real, M. (2016) reviewed the work of Black and Wiliam (1998) and found that after expanding the broad definition of formative assessment, formative assessment still had a positive effect on student achievement. Ozan and Kincal (2018) published a study that showed formative assessment practices "had a significantly higher academic achievement levels and better attitudes toward the class than the students did in the control group." Tibbitt (2020) noted the importance of formative assessment as an effective tool in approving the achievement of all students. Her research showed formative assessment's positive impact on students of color, students with varying cognitive abilities, and students with disabilities. The results of the McMillan, Venable, & Varier (2013) meta-analysis study of formative assessment concluded more quality research is needed to substantiate the notion of formative assessment as a key strategy in improving student achievement.

Summary

The information in this chapter explored the history of teaching in America as it relates to the research for this study. The literature included major paradigm shifts in education from the 19th and early 20th centuries. Information was also included on the government's impact on education as U.S. presidents began to enact new educational reforms. The chapter described the controversies surrounding the teacher evaluation process. An explanation of highly qualified teachers as it relates to the teacher's years of experience and level of education was also discussed. Finally, a history of the NEE teacher evaluation system was explored as it relates to measuring teacher effectiveness and maintaining data for the years of experience and level of education for teachers within their network.

The information provides a broad understanding of the teacher evaluation system in America and its connection to determining teacher effectiveness. It further explains the need for a method to determine if the teachers' years of experience and levels of education have correlations to academic success. In chapter 3 the researcher describes the methods used to conduct the research for this study.

Chapter 3

Methods

The purpose of this study was to determine the extent the NEE teacher evaluation scores were higher than the threshold of 5 on the NEE indicators for student engagement and conducting formative assessment. An additional purpose of this study was to determine if teachers' years of experience or teachers' level of education had an impact on their NEE evaluation scores for the student engagement indicator and conducting formative assessment indicator. Chapter 3 includes the research design, selection of participants, measurement, data collection procedures, data analysis and hypothesis testing, and the limitations of the study.

Research Design

A quasi experimental research design was utilized for this study. Creswell (2014) stated, "Quasi experimental research is an experimental research in which the research subjects were not likely to be assigned randomly" (p. 142). The participants in this research study were not assigned randomly. The research design involved the variables of teacher observation scores for the NEE formative assessment indicator, the teacher observation scores for the NEE student engagement indicator, the teachers' years of experience, and the teachers' level of education. This approach was suitable to explore the differences based on the teachers' years of experience and level of education on the two NEE evaluation indicators for student engagement and conducting formative assessment.

Selection of Participants

Purposive sampling was used in this study. Lunenburg and Irby (2008) defined purposive sampling as a process that "involves selecting a sample based on the researcher's experience or knowledge of the group to be sampled" (p. 175). The participants for this study were chosen based on the researcher's knowledge of the NEE data tool and its design to collect, store, and compare teacher data related to their levels of experience, levels of education, and NEE evaluation scores. The population for this study consisted of teachers in a suburban school district near Kansas City, Missouri, who were evaluated using the NEE evaluation tool. The sample included secondary teachers who taught in the district between 2015-2019.

Teachers were included in the sample for this study if they met the following criteria

1. The teacher taught at the secondary level.
2. The teacher had been scheduled for a summative evaluation. Teachers are scheduled for a summative evaluation every 4 years. (School District A Board Policy GCN - AP1)
3. The teacher received a minimum of six or more administrator observations. Teachers on cycle for a summative evaluation should have 6-8 administrator observations (EDHub, personal communication, May 6, 2020). However, there are occasions when the evaluator is not able to conduct the minimum 6 observations. The

data for teachers without six observations was not used in this data collection.

4. The teacher had at least six evaluator scores for the formative assessment evaluation indicator. There are occasions where the teacher had six observations, but the evaluator chose not to score a particular indicator because there was no opportunity to observe that indicator during the time the evaluator was present in the room. This practice is one of the trained protocols taught by NEE (EdHub, personal communication, May 6, 2020). If this occurred, the teacher evaluation scores were not used in the data.

Measurement

All data from the six secondary schools in the district was retrieved from the NEE database. NEE's online data tool was designed to match teacher profile data with teacher evaluation scores (NEE, 2015). At the beginning of each school year, teachers are required to update or verify their years of experience and level of education in their NEE profile before gaining access to other features of the NEE profile page. The Human Resources department for each school district can then cross reference the teacher-reported information with their employment records for the teachers. As administrators are conducting the teacher observations during the school year, the NEE data tool stores the observation scores in the teachers' profiles connected to their years of experience and level of education.

For the first component of the study, the researcher included data that was retrieved from the NEE database to divide the teachers into three categories by their years of experience: 0 - 5 years, 6 - 14 years, and 15 years or more. For the second component of the study, the researcher organized the data from the NEE database into two categories to group teachers by level of education: (a) bachelor's degree, and (b) master's degree. No identifiable personal information was released with this data. The NEE system automatically reported averages for the two indicators, student engagement and conducting formative assessment, as the administrators submitted evaluation scores during the in-class teacher observations.

The NEE measurement tool is a practical evaluation tool to evaluate teachers. It provides a wider range of scores to rate teachers. The previous teacher evaluator tool used in Missouri, Performance Based Teacher Evaluation (PBTE) only had three levels in the rating scale to measure a teacher's performance: *Does Not Meet Standards*, *Meets Standards*, and *Exceeds Standards*. However, the NEE evaluation tool utilizes a scale with seven levels. Using this model, the evaluator could give a more precise rating of the teacher's performance on the indicator. In the NEE evaluation model, each evaluator watches for evidence of the indicator to determine how often and to what quality the indicator is utilized. Evaluators mark one of the following:

____ 0 - No evidence of the indicator,

____ 1 - Seldom demonstrates evidence of the indicator,

____ 3 - Less than half of the time and/or less than half of the students,

____ 5 - More than half of the time and/or more than half of the students,

____ 7 - Almost all of the time and/or almost all of the students.

The evaluator also has the option to use the even numbers not listed on the scale when a teacher demonstrates more or less evidence than the labeled numbers on the scale.

NEE (2021) created rubrics for each of the 27 indicators used by administrators to score teachers during the teacher observation. The rubric also has a "look for," which explains examples of why the score may be marked at the score assigned by the evaluator. The design of the scale has a unique vocabulary for each indicator. For example, in the figure below for Indicator 1.2, Cognitively engages students in the subject, each score on the left-hand column uses the phrase "cognitively engage(s) students" and assigns a score (0, 1, 3, 5, 7) based on the number of times the teacher utilized the indicator during the time the evaluator was in the room. If the teacher occasionally utilized the indicator for less than half of the time or with half of the students, the evaluator could assign a score of 3. If the teacher utilized the indicator for more than half of the time, a higher score of 5 could be assigned.

The evaluator can also give a score in between the designated odd scores of 1,3,5,7 on the left-hand side of the figure. The right-hand column has a list of descriptors for each score in the right-hand column. These are called "look fors." These descriptors provide additional information for the evaluator to assign even scores (2,4,6) that suggest the teacher's use of the indicator was slightly higher or lower than the scores in the left-hand column. For example, if the teacher

cognitively engages more than half of the students more than half the time, but does not use any alternative strategies to engage non-engaged students, the evaluator might assign a score of 4 instead of the 5 marked on the rubric. (How does the NEE classroom Observation work? p 4-5).

Indicator 1.2: Cognitively engages students in subject	
The teacher…	
0 - does not cognitively engage students in the content.	• Does not use instructional strategies to promote thinking about the content • Students are not cognitively engaged in the subject matter
1 - seldom cognitively engages students in the content.	• Ineffectively uses at least one potentially weak instructional strategy to promote thinking about the content • Only cognitively engages one student at a time
3 - occasionally cognitively engages students in the content, less than half of the time, or less than half of the students.	• Uses cognitive engagement strategies (e.g., advanced organizers, K-W-L charts, share-out, shoulder-partner), but not very effectively • Missed opportunities for thinking about the content • Some students are cognitively engaged minimally
5 - occasionally cognitively engages students in the content, more than half of the time, or more than half of the students.	• Most students are cognitively engaged much of the time • Recognizes if some are not cognitively engaged, and tries alternate strategies to increase or maintain students' thinking about content • Uses specific processing structures with students
7 - almost always cognitively engages students in the content, or engages almost all the students.	• Almost all students spend most of the time cognitively engaged with the content • Effectively uses strategies to promote thinking about the content • Supports students in monitoring their own levels of cognitive engagement and in employing personal strategies to increase their engagement

Figure 1. Indicator 1.2: Cognitively engages students in subject. Adapted from NEE training materials by NEE, 2022, p. 2. (EdHub, personal communication, May 6, 2022).

Indicator 7.4: Monitors effect of instruction on individual and class learning	
The teacher…	
0 - does not check the effect of instruction on whole class or individual learning.	• Does not assess whether students have achieved the lesson objective • Does not engage in on-the-spot assessment.
1 - seldom conducts formative, on-the-spot assessment of learning for either the whole class or individual students or does not take needed corrective action.	• Seldom monitors learning progress • May superficially use question and answer as assessment • Minimal follow-up or checking for understanding • Monitors learning somewhat, but does not take corrective action.
3 - conducts formative, on-the-spot assessment of learning less than half-of the-time or for less than half of the students and takes corrective action as needed.	• Occasionally quickly assesses understanding of some students before moving on to the next learning activity • Occasionally uses techniques to monitor learning progress such as observing classroom interactions or student work, questioning, thumbs up, fist-to-five, white boarding, exit slips • May monitor progress of the class as a whole • If needed, some corrective action is taken • Must take corrective action to score above a "2"
5 - conducts formative, on-the-spot assessment of learning more than half of the time or for more than half of the students and takes corrective action as needed.	• Occasionally monitors learning progress of most students • Monitors the whole class and many individuals May use multiple checks for understanding • Often adjusts instruction using students' responses to questions and discussions, correcting misconceptions, or monitoring other feedback • If needed, corrective action appropriate to most students is taken
7 - almost always conducts formative, on-the-spot assessment of learning for both the whole class and almost all individual students and takes corrective action as needed.	• Systematically monitors learning progress • Continuously monitors progress of attaining instructional objectives of the whole class and of each student • On-the-spot assessment is seamless throughout instruction Strong, appropriate corrective action is taken to ensure learning of almost all students

Figure 2. Indicator 7.4 Monitors effect of instruction on individual and class Adapted from NEE training materials by NEE, 2022, p. 27. (EdHub, personal communication, May 6, 2022).

Each teacher included in this 2019 study received a minimum of six or more observations on the two indicators, student engagement and conducting formative assessment. Each indicator is scored separately, and scores from one indicator are not combined with evaluation scores from a different indicator. The

six scores for each indicator were averaged to provide a summative score at the end of the evaluation cycle. During a summative evaluation, which occurs near the end of each academic school year, the teacher receives a summative report that contains the teacher's score on each indicator. The report also compares the teacher's scores with those of other teachers in the building, other teachers in the district, and other teachers in Missouri. This data was used to determine if there was a difference in the range of possible evaluation scores based on the teachers' years of experience and level of education.

According to Lunenburg and Irby (2008), validity is "the degree to which an instrument measures what it purports to measure" (p. 181). The NEE data tool is a valid measuring tool for the data collected for this study. EdHub (2015), the company that manages the evaluation data for each school district using NEE, conducts annual audits of each district's data to ensure the accuracy of all information in the database. All teachers are required to annually verify their years of experience and level of education. School districts can then verify this information against their own records. These steps validate the teacher self-reported data used to measure the teachers' years of service and level of education.

Lunenburg and Irby (2008) defined reliability "as the degree to which an instrument consistently measures whatever it is measuring" (p. 182). The NEE data tool is a reliable method for administrators to evaluate teachers. School administrators using the NEE evaluation tool receive an initial 3-day training from NEE representatives. NEE describes the following about the training, "At

these educator training sessions, evaluators practice scoring videos of teachers. Scores are then compared and evaluators are given feedback about their scoring practices. Evaluators take a qualification assessment at the end of their training, which helps to increase inter-rater agreement in a building, in a school district, and across the country" (NEE "Classroom Observation & Walkthrough Training"). During the training, the administrators watch a series of videos containing 8 -10-minute teacher observations. The observations have been scored by master scorers or expert scorers of the NEE indicators. The administrators must score the videos within one score above or below the master scorer's score for the observation. This process is part of the administrator certification test. Administrators must pass this test to be certified to conduct teacher observations for the upcoming school year.

According to research analysis for NEE, to ensure the rating reliability of the NEE data, NEE requires administrators to attend a recertification training each year, where they take a "recertification exam on their proficiency/accuracy in conducting classroom observations each summer." (T. Hairston, personal communication, February 4, 2022). Administrators participate in a half-day training module with four specific goals.

1. Score classroom observations accurately.
2. Learn about updates to NEE for the (next school year).
3. Utilize the NEE Guide to Effective Instructional Change to plan teacher professional growth processes

4. Qualify on their own, which will complete recertification requirements for the (school year). (EdHub, personal communication, May 6, 2020)

Similar to the initial training, administrators must take another qualification assessment at the end of the training. The researcher believes this process suggests the NEE data tool is a reliable and valid method to evaluate teachers because it ensures all administrators have the same initial training regime and recertify using the same assessments each year.

Data Collection Procedures

The researcher submitted a request to the Institutional Review Board at Baker University to receive permission to collect data for this study. The submission was approved in July 2022 (see Appendix A). A formal request was submitted to the human resource department of School District A and permission was granted to conduct the study and to use data for district teachers in the research on July 21, 2021 (see Appendix B). The school district contacted the NEE organization to arrange a meeting between the researcher and a research analyst from NEE to discuss the archival data needed for the research study. NEE granted permission to use their teacher evaluation data and information related to teachers' years of experience and teachers' level of education. To use the NEE teacher evaluation data and evaluation training documents from EdHub (the subscription-only database) in the research study, NEE required the use of the NEE Data Sharing and Use Agreement (see Appendix C).

Data Analysis and Hypothesis Testing

One-sample t tests were used to determine whether the teachers' evaluation scores were higher than the threshold value of 5. One-factor analyses of variance (ANOVAs) were conducted to identify the differences in the teachers' indicator scores based on their years of experience. Independent-sample t tests were used to identify the differences in the teachers' indicator scores based on their level of education.

RQ1. To what extent are NEE scores on the indicator for student engagement different than the threshold value of 5?

H1. The NEE scores on the indicator for student engagement are different from the threshold value of (5).

A one-sample *t* test was conducted to test H1. The sample mean was compared to a test value of 5. The one-sample *t* test was chosen for the hypothesis testing because it involves the comparison of one group mean with a known value, and the group mean is calculated from a numerical variable. The level of significance was set at .05. When appropriate, the effect size, as indexed by Cohen's *d*, is reported.

RQ2. To what extent are NEE scores on the indicator for conducting formative assessment higher than the threshold value of (5)?

H2. The NEE scores on the indicator for conducting formative assessment are different from the threshold value of (5).

A one-sample *t* test was conducted to test the H2. The sample mean was compared to a test value of 5. The one-sample *t* test was chosen for the

hypothesis testing because it involves the comparison of one group mean with a known value, and the group mean is calculated from a numerical variable. The level of significance was set at .05. When appropriate, the effect size, as indexed by Cohen's *d*, is reported.

RQ3. To what extent is there a difference in scores on the NEE indicator for student engagement based on the teachers' years of experience?

H3. There is a difference in scores on the NEE indicator for student engagement based on teachers' years of experience.

A one-factor analysis of variance (ANOVA) was conducted to test H3. The categorical variable used to group the dependent variable, scores on the NEE indicator for student engagement, was years of experience (0-5, 6-14, 15 or more). The results of the one-factor ANOVA can be used to test for differences in the means for a numerical variable among three or more groups. The level of significance was set at .05. When appropriate, an effect size, as indexed by eta squared, is reported.

RQ4. To what extent is there a difference in scores for the NEE indicator for conducting formative assessments based on the teachers' years of experience?

***H4*.** There is a difference in scores for the NEE indicator for conducting formative assessment based on teachers' years of experience.

A one-factor ANOVA was conducted to test H4. The categorical variable used to group the dependent variable, scores on the NEE indicator for conducting formative assessment, was years of experience (0-5, 6-14, 15 or more). The results of the one-factor ANOVA can be used to test for differences in the means

for a numerical variable among three or more groups. The level of significance was set at .05. When appropriate, an effect size, as indexed by eta squared, is reported.

RQ5. To what extent is there a difference in scores for the NEE indicator for student engagement based on the teachers' level of education?

***H5*.** There is a difference in NEE indicator scores for student engagement based on the teachers' level of education.

An independent-samples *t* test was conducted to test H5. The two sample means for NEE indicator scores for student engagement were compared. An independent-samples *t* test was chosen for the hypothesis testing. The hypothesis test involves the examination of the mean difference between two mutually exclusive independent groups, and the means are calculated using data for numerical variables. The level of significance was set at .05. When appropriate, an effect size, as indexed by Cohen's *d*, is reported.

RQ6. To what extent is there a difference in scores for the NEE indicator for conducting formative assessment based on the teachers' level of education?

***H6*.** There is a difference in scores for the NEE indicator for conducting formative assessment based on the teachers' level of education.

An independent-samples *t* test was conducted to test H6. The two sample means for NEE indicator scores for conducting formative assessments were compared. An independent-samples *t* test was chosen for the hypothesis testing. The hypothesis test involves the examination of the mean difference between two mutually exclusive independent groups, and the means are calculated using data

for numerical variables. The level of significance was set at .05. When appropriate, an effect size, as indexed by Cohen's *d*, is reported.

Limitations

Lunenburg and Irby (2008) defined limitations as "factors that may have an effect on the interpretation of the findings or the generalizability of the results" (p. 133). The factors are beyond the control of the researcher. Limitations of this study included

1. The administrators all attended the same initial administrator evaluation training and recertification training each year. However, administrators can choose to attend an in-person or vertical training session, which can impact their level of engagement and potential gains of the training.
2. Despite the training, administrators still have the opportunity to score teachers outside of the parameters of what they were taught.

Summary

The purpose of this chapter was to describe the research method used to conduct this quantitative research study. The researcher explained the methods for selecting the participants, gathering and measuring the data, and analysis for the research questions. The purpose of Chapter 4 is to describe the results of the descriptive statistics and the study.

Chapter 4

Results

The purpose of this study was to determine the extent the NEE teacher evaluation scores were higher than the threshold of 5 on the NEE indicators for student engagement and conducting formative assessment. In addition, the purpose of this study was to determine if a teachers' years of experience or teachers' level of education had an impact on their NEE evaluation scores for the student engagement indicator and conducting formative assessment indicator. Chapter 4 describes the descriptive statistics for the sample. The chapter also provides the results of the hypothesis testing for research questions.

Descriptive Statistics

Lunenburg and Irby (2008), described descriptive statistics as the "mathematical procedures for organizing and summarizing numerical data" (p. 63). The researcher divided the teachers into two categories. The first category for this study was the teachers' years of experience. Teachers were organized into three groups to compare teachers with beginning, middle, and long-term teaching experience. The middle group was initially organized for teachers with 6-15 years of experience, but it was changed to 6-14 years to increase the number of teachers in the long-term sample group.

Table 1

Descriptive Analysis for Teachers' Years of Experience

Years	*N*	%
0-5	58	46.4
6-14	49	39.2
15+	18	14.4

The second category for this study was the teachers' levels of education. For levels of education, the teachers were initially organized into three groups. However, this category was reduced to two groups to be more consistent with the wording in the research that noted teachers with a bachelor's degree versus teachers with advanced degrees. In addition, the teachers were only included in the sample group if they had six NEE evaluations for each of the two indicators: student engagement and conducting formative assessment. There were 125 teachers who met this criterion. However, if the teacher did not identify their level of education in the NEE teacher portal, they were excluded from the data results. Therefore, the number of participants for the level of education decreased to 113.

Table 2

Level of Education	*N*	%
Original		
Bachelor's	25	22.2
Master's	80	70.8
Specialist	18	7.0
Recoded		
Bachelor's	25	22.2
Master's or Higher	88	77.9

Note: *N* = 12 teachers were omitted from this analysis because they did not list level of education.

Hypothesis Testing

One-sample t tests were used to determine whether the teachers' evaluation scores were higher than the threshold value of 5. ANOVAs were conducted to identify the differences in the teachers' indicator scores based on their years of experience. Independent-samples t tests were used to identify the differences in the teachers' indicator scores based on their level of education.

RQ1. To what extent are NEE scores on the indicator for student engagement different than the threshold value (5)?

H1. The NEE scores on the indicator for student engagement are different from the threshold value (5).

A one-sample *t* test was conducted to test H1. The sample mean was compared to a test value of 5. The one-sample *t* test was chosen for the hypothesis testing because it involves the comparison of one group mean with a

known value, and the group mean is calculated from a numerical variable. The level of significance was set at .05. When appropriate, the effect size, as indexed by Cohen's *d*, is reported.

The results of the one sample *t* test indicated a statistically significant difference between the group mean and the test value, $t(124) = 11.822$, $p = .000$, Cohen's $d = 1.057$. The sample mean ($M = 5.76$, $SD = .72$) was significantly higher than the test value (5). H1 was supported. The NEE scores on the indicator for student engagement are different from the threshold value (5). The effect size indicated a large effect.

RQ2. To what extent are NEE scores on the indicator for conducting formative assessment higher than the threshold value (5)?

H2. The NEE scores on the indicator for conducting formative assessment are different from the threshold value (5).

A one-sample *t* test was conducted to test the H2. The sample mean was compared to a test value of 5. The one-sample *t* test was chosen for the hypothesis testing because it involves the comparison of one group mean with a known value, and the group mean is calculated from a numerical variable. The level of significance was set at .05. When appropriate, the effect size, as indexed by Cohen's *d*, is reported.

The results of the one sample *t* test indicated a statistically significant difference between the group mean and the test value, $t(124) = 6.872$, $p = .000$, Cohen's $d = 0.615$. The sample mean ($M = 5.49$, $SD = .80$) was significantly higher than the test value (5). H2was supported. The NEE scores on the indicator

for conducting formative assessment are different from the threshold value (5). The effect size indicated a large effect.

RQ3. To what extent is there a difference in scores on the NEE indicator for student engagement based on the teachers' years of experience?

H3. There is a difference in scores on the NEE indicator for student engagement based on teachers' years of experience.

A one-factor (ANOVA) was conducted to test H3. The categorical variable used to group the dependent variable, scores on the NEE indicator for student engagement, was years of experience (0-5, 6-14, 15 or more). The results of the one-factor ANOVA can be used to test for differences in the means for a numerical variable among three or more groups. The level of significance was set at .05. When appropriate, an effect size, as indexed by eta squared, is reported.

The results of the analysis indicated there was not a statistically significant difference between at least two of the means, $F(2, 122) = 0.761$, $p = .469$. See Table 3 for the means and standard deviations for this analysis. H3 was not supported. There is not a difference in scores on the NEE indicator for student engagement based on teachers' years of experience.

Table 3

Descriptive Statistics for the Results of the Test for H3

Years of Experience	*M*	*SD*	*N*
0-5	5.68	0.70	58
6-14	5.85	0.71	49
15 or more	5.80	0.82	18

RQ4. To what extent is there a difference in scores for the NEE indicator for conducting formative assessment based on the teachers' years of experience?

H4. There is a difference in scores for the NEE indicator for conducting formative assessments based on teachers' years of experience.

A one-factor ANOVA was conducted to test H4. The categorical variable used to group the dependent variable, scores on the NEE indicator for conducting formative assessment, was years of experience (0-5, 6-14, 15 or more). The results of the one-factor ANOVA can be used to test for differences in the means for a numerical variable among three or more groups. The level of significance was set at .05. When appropriate, an effect size, as indexed by eta squared, is reported.

The results of the analysis indicated there was not a statistically significant difference between at least two of the means, $F(2, 122) = 0.153$, $p = .858$. See Table 4 for the means and standard deviations for this analysis. H4 was not supported. There is not a difference in scores on the NEE indicator for conducting formative assessment based on teachers' years of experience.

Table

Descriptive Statistics for the Results of the Test for H4

Experience	*M*	*SD*	*N*
0-5	5.48	0.71	58
6-14	5.53	0.82	49
15 or more	5.41	1.02	18

RQ5. To what extent is there a difference in scores for the NEE indicator for student engagement based on the teachers' level of education?

***H5*.** There is a difference in NEE indicator scores for student engagement based on the teachers' level of education.

An independent-samples *t* test was conducted to test H5. The two sample means for NEE indicator scores for student engagement were compared. An independent-samples *t* test was chosen for the hypothesis testing. The hypothesis test involves the examination of the mean difference between two mutually exclusive independent groups, and the means are calculated using data for numerical variables. The level of significance was set at .05. When appropriate, an effect size, as indexed by Cohen's *d*, is reported.

The results of the independent samples *t* test indicated no difference between the two means, $t(111) = -0.359$, $p = .720$. The sample mean for teachers with a bachelor's degree ($M = 5.70$, $SD = 0.72$, $n = 25$) was not different from the sample mean for teachers with a master's degree or higher ($M = 5.76$, $SD = 0.74$, $n = 88$). H5 was not supported. There is not a difference in NEE indicator scores for student engagement based on the teachers' level of education.

RQ6. To what extent is there a difference in scores for the NEE indicator for conducting formative assessment based on the teachers' level of education?

***H6*.** There is a difference in scores for the NEE indicator for conducting formative assessment based on the teachers' level of education.

An independent-samples *t* test was conducted to test H6. The two sample means for NEE indicator scores for conducting formative assessments were

compared. An independent-samples *t* test was chosen for the hypothesis testing. The hypothesis test involves the examination of the mean difference between two mutually exclusive independent groups, and the means are calculated using data for numerical variables. The level of significance was set at .05. When appropriate, an effect size, as indexed by Cohen's *d*, is reported.

The results of the independent samples *t* test indicated no difference between the two means, $t(111) = -0.258$, $p = .797$. The sample mean for teachers with a bachelor's degree ($M = 5.43$, $SD = 0.77$, $n = 25$) was not different from the sample mean for teachers with a master's degree or higher ($M = 5.48$, $SD = 0.82$, $n = 88$). H6 was not supported. There is not a difference in NEE indicator scores for conducting formative assessments based on the teachers' level of education.

Summary

Chapter 4 started with a summary of the descriptive statistics used to explain the quantitative data for this research study. The results of the hypothesis testing for research questions 1 and 2 showed a statistically significant difference between the group mean and the test value. This means the NEE scores for the student engagement indicator were different from the threshold value of 5. The results of hypothesis testing for research questions 3 through 6 revealed there was not a statistically significant difference between at least two of the means. Therefore, there was not a difference in scores on the NEE indicator for student engagement based on the teacher' years of experience and level of education. In addition, there was not a difference in scores on the NEE indicator for conducting formative assessment based on the teachers' level of education.

Chapter 5

Interpretation and Recommendations

The purpose of this study was to determine if NEE evaluation scores for student engagement and conducting formative assessment were different based on the teacher's years of experience and level of education. The teacher evaluation process is an important tool in determining teacher effectiveness. In 2009, as part of the Race to the Top (RTTT) federal education program, school districts were charged to establish and maintain highly qualified teachers (The U.S. Department of Education, 2016). However, educational leaders often debate if a teacher's level of education or a teacher's years of experience has a more positive correlation with student achievement. Ladd (2013) and Podolsky (2016) found that experienced teachers had a positive effect on student achievement than experienced teachers. Other research from Darling-Hammond (2000) indicated that the student's academic performance increased when the teacher had an advanced degree when compared to teachers with just a bachelor's degree. Chapter 5 includes a summary of the study, the findings related to the literature, and the conclusions.

Study Summary

Chapter 5 provides an overview of the problem school districts encounter as they meet the federal government's mandate to hire and maintain highly qualified teachers. It also includes the purpose of the study and the research questions created to conduct the research. A summary of chapters one through

four is included, followed by the key findings from the literature, the results of the study, implications for research, and recommendations for future studies.

Overview of the problem. School districts compete for highly qualified teachers by paying higher salaries for experienced teachers and teachers with advanced degrees. However, the previous research on whether the teachers' years of experience or the teachers' level of education mattered produced conflicting results. Most researchers studied the two factors in isolation by focusing on years of experience or level of education, but not both in the same study. To provide more insight into this debate, school districts need a teacher evaluation system that reviews the teachers' level of education and the teachers' years of experience in the same measurement tool. Doss (2016) created the NEE teacher evaluation system as a way to provide teacher evaluation data that includes the teachers' level of experience and years of education. This data could help school districts take a more critical look at teacher effectiveness based on their level of experience and education.

Purpose statement and research questions. The purpose of this study was to determine if the NEE teacher evaluation scores were different than the threshold of (5) on the NEE indicators for student engagement and conducting formative assessment. NEE considers a score of 5 or higher an acceptable evaluation score. The second purpose of the study was to determine to what extent the NEE teacher evaluation scores were different on the student engagement indicator and the conducting formative assessment indicator based on teachers' years of experience. Using data from the administrator evaluations and

student surveys, the study was conducted to identify whether teachers with 0-5, 6-14, and 15 or more years of experience scored higher on the two NEE indicators: student engagement and conducting formative assessments. The third purpose of the study was to determine to what extent the NEE teacher evaluation scores were higher on the student engagement indicator and the conducting formative assessment indicator based on teachers' level of education. Teachers were divided into groups of those with advanced degrees versus those with a bachelor's degree to also determine how varying levels of education impacted scores on the two NEE indicators.

The following research questions were used to conduct the research.

RQ1. To what extent are NEE scores on the indicator for student engagement different from the threshold value of 5?

RQ2. To what extent are NEE scores on the indicator for conducting formative assessment different from the threshold value of 5?

RQ3. To what extent is there a difference in scores on the NEE indicator for student engagement based on the teachers' years of experience?

RQ4. To what extent is there a difference in scores for the NEE indicator for conducting formative assessment based on the teachers' years of experience?

RQ5. To what extent is there a difference in scores for the NEE indicator for student engagement based on the teachers' level of education?

RQ6. To what extent is there a difference in scores for the NEE indicator for conducting formative assessment based on the teachers' level of education?

Review of the methodology. A quasi experimental research design was utilized for this study. The research design involved the variables of teacher observation scores for the NEE formative assessment indicator, the teacher observation scores for the NEE student engagement indicator, the teachers' years of experience, and the teachers' level of education. This approach was suitable to explore the differences based on the teachers' years of experience and level of education on the two NEE evaluation indicators for student engagement and conducting formative assessment.

The participants for this study were chosen based on the researcher's knowledge of using the NEE data tool and its design to collect, store, and compare teacher data related to their level of experience, level of education, and NEE evaluation scores. The population for this study consisted of teachers in a suburban school district near Kansas City, Missouri, who were evaluated using the NEE evaluation tool. The sample included secondary teachers who taught in the district between 2015-2019.

For the first component of the study, the researcher included data that was retrieved from the NEE database to divide the teachers into three categories by their years of experience: 0 - 5 years, 6 -14 years, and 15 years or more. For the second component of the study, the researcher organized the data from the NEE database into two categories to group teachers by level of education: (a) bachelor's degree, and (b) master's degree. No identifiable personal information was released with this data. The NEE system automatically reported averages for the two indicators, student engagement and conducting formative assessment, as

the administrators submitted evaluation scores during the in-class teacher observations.

Major findings. The major findings from the study were mixed based on the research that was conducted. The first two research questions asked whether teachers in District A scored higher than a threshold of 5 on the teacher evaluation indicators for student engagement and conducting formative assessment. The results indicated that teachers in District A significantly scored higher than the threshold of 5 for teacher evaluator scores related to student engagement and conducting formative assessments. In addition, the effect size for both indicated a large effect.

Research questions three and four assessed whether there was a difference in the teachers' NEE indicator scores for student engagement and conducting formative assessment based on the teachers' years of experience. The research findings indicated there was not a significant difference based on the teachers' years of experience as it relates to the NEE indicators for student engagement and conducting formative assessments. Teachers of all experience levels scored within the same range on the two NEE indicators.

Research questions five and six focused on whether there was a difference on the teachers' NEE indicators scores for conducting formative assessment based on the teachers' level of education. The findings failed to show a significant difference in the teachers' evaluation scores based on their level of education as it relates to the NEE indicators for student engagement and conducting formative

assessment. Ultimately, teachers scored similarly on the two NEE indicators regardless of possessing a bachelor's degree or an advanced degree.

Findings Related to the Literature. The findings from the research indicated teachers in district A typically scored high on the NEE evaluation indicators regardless of their years of experience or level of education, however, there was not a significant difference in evaluation scores based on the teachers' years of experience or level of education. The findings from this research provide continued ambiguity over whether the teachers' years of experience and/or level of education are important. Horn and Jang (2017) found mixed results regarding the teachers' level of education and student learning. They concluded degrees were more impactful in subjects like math and science, but less important in English and reading. Likewise, Selke's (2001) research concluded that advanced levels of education were primarily significant when the teacher received a degree in the subject area in which they taught, but not as impactful if it was in another field like administration.

Several research studies from the literature section noted the positive correlations between the teachers' years of experience and level of education as it relates to teacher effectiveness in the classroom. Kini and Podolsky (2016) found that teachers with more experience favorably impact student achievement on standardized tests, school attendance, and classroom behavior. Clotfelter et al. (2007) found that 3rd and 5th grade students in their study had higher math scores when they had teachers with advanced levels of teaching experience and education. Selke (2001) research study concluded that teachers with a master's

degree in the field they taught had better strategies to help their students improve academically.

The findings from the research study showing similar teacher evaluation scores regardless of the teachers' years of experience or level of education exemplify the literature from Chapter 2 which detailed some negative perceptions associated with teacher evaluation systems. Callahan and Sadeghi (2015) noted shortcomings of teacher evaluation systems if they rely heavily on a checklist of what the evaluated observed rather than timely feedback connected to state standards and professional development opportunities. Warring (2015) also noted the weaknesses in teacher evaluation systems that fail to connect teacher evaluation scores to student achievement. Warring (2015) and Punyanunt-Carter and Carter (2015) also found perceptions of bias in teacher evaluation systems. Bias in rating teachers was prevalent in student and administrator scores based on the teacher's reputation or likeability. Weisber, et al. (2009) outlined flaws in teacher evaluation systems that use the same standards to evaluate experienced and novice teachers regardless of educational level and previous professional development opportunities.

Conclusions

This section summarizes the conclusions from the current study on the relationship between the teachers' years of experience and level of education on the two NEE indicator scores for student engagement and conducting formative assessment. The implications for action and recommendations for future research are explained.

Implications for action. Since the findings of this current research study indicated there were no significant differences in teacher evaluation scores based on the teachers' years of experience or level of education, there are several implications. District A could study why teachers of all educational backgrounds scored similarly on the NEE indicators. With 80% percent of the teachers holding an advanced degree and 16 years of average teaching experience, District A could determine if additional administrator training is needed to differentiate teaching strategies expected by experienced teachers versus novice teachers for measuring teacher effectiveness on student engagement and conducting formative assessment.

Recommendations for future research.

The research findings for this study indicated that there was not a significant difference in the NEE evaluation scores for the NEE indicators on student engagement and conducting formative assessment as it relates to a teacher's years of experience and level of education. The researcher recommends additional research studies using data from a teacher evaluation system different from the NEE system to determine if there may be a different outcome. Other teacher evaluation tools may use different indicators to evaluate teachers, which may result in a different outcome in the data.

The researcher also recommends a research study in an urban or rural school district as compared to the suburban school district in which this study was conducted. Factors like the teacher turnover rate, higher or lower rate of teacher education levels above the bachelor's degree, and varying attitudes about

education from urban and rural school districts may yield different results in the data.

Concluding remarks.

The debate on whether the teacher's years of experience or level of education have a positive correlation with teacher effectiveness in the classroom will continue. Most of the previous research on this topic reviewed the two factors in isolation. However, this study reviewed the research on both factors and provided data that measured years of experience and level of education in the teacher evaluation process. This information could be valuable to school districts and education leaders as they continue to hire and maintain highly qualified teachers in efforts to increase student achievement. More studies are needed to study the factors in other school districts who use the NEE evaluation system. Studies within other suburban school districts similar to District A as well as urban and rural school districts could also provide more definitive answers regarding the impact of the teachers' years of experience and level of education on student achievement.

References

Anonymous School District. (2022). Board policy GCN - AP1. Retrieved from DESE report on Anonymous School District A.

Apthorp, H., Klute, M., Petrites, T., Harlacher, J., & Real, M. (2016). *Valuing a more rigorous review of formative assessment's effectiveness*. Retrieved from ERIC database. (ED567511)

Aragon, S. (2018). *Teacher evaluations: What is the issue and why does it matter? Policy snapshot.* Retrieved from hecs.org/wp-content/uploads/Teacher_Evaluations.pdf

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principals. Policy, & Practice*, 5(1). 7-74.

Brill, S. (August 31, 2009). The rubber room. *The New Yorker*. Retrieved from https://www.newyorker.com/magazine/2009/08/31/the-rubber-room

Callahan, K., & Sadeghi, L. (2015). Teacher perceptions of the value of teacher evaluations: New Jersey's ACHIEVE NJ. *International Journal of Educational Leadership Preparation*, 10(1), 46-59. Retrieved from ERIC database. (EJ1060978)

Chen, G. (2019) Understanding No Child Left Behind [Blog post].Retrieved from www.publicschool review.com.

Close, K., Amrein-Beardsley, A., & Collins, C. (2019). Mapping America's teacher evaluation plans under ESSA. *Phi Delta Kappan, 101*(2), 22-26. Retrieved from https://journals.sagepub.com/doi/abs/10.1177/0031721719879150

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). *How and why do teacher credentials matter for student achievement*? Retrieved from https://www.nber.org/papers/w12828

Cocoran, S. P. (2010). Can teachers be evaluated by their students' test scores? Should they be? The use of value-added measures of teacher effectiveness in policy and practice. *Education policy for action series*. Retrieved from ERIC database. (ED522163)

Cohen, J., & Goldhaber, D. (2016). Building a more complete understanding of teacher evaluation using classroom observations. *Educational Researcher*, *45*(6), 378-387.

Cuban, L. (2019). *Whatever happened to Madeline hunter? Larry Cuban on school reform and classroom practice*. Retrieved from https://larrycuban.wordpress.com/2019 /10/08/whateve r-happened-to-madeline-hunter-lesson-plans/

Cubberley, E. P. (1916). *Public school administration: A statement of the fundamental principles underlying the organization and administration of public education*. Boston, MA: Houghton Mifflin.

Danielson, C. (2007). *Enhancing professional practice: A framework for teaching* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Danielson, C. (2013). *The framework for teaching 2013 edition evaluation instrument*. Retrieved from https://issuu.com/timfarquer/docs/danielson_f armework_ for_ teaching_eval_instrument. p. iv

Darling-Hammond L. How teacher education matters. *Journal of Teacher Education*. 2000;51(3),166-173. doi:10.1177/0022487100051003002

DeAngelis, K., White, B., & Presley, J. (2010). The changing distribution of teacher qualifications across schools: A statewide perspective post-NCLB. *Education Policy Analysis Archives, 18*, 28. doi:https://doi.org/10.14507/epaa.v18n28.2010

DeNisi, A. S., & Sonesh, S. (2011). The appraisal and management of performance at work. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology: Vol. 2. Selecting and developing members for the organization* (pp. 255–279). Retrieved from ERIC database. (ED058166)

Department of Elementary and Secondary Education (2020). Faculty Information. Retrieved from https://apps.dese.mo.gov/MCDS/home.aspx?categoryid=1&view=2

Department of Education. (2023). Missouri – Highly qualified teachers state plan. Retrieved from https://www2.ed.gov/programs/teacherqual/hqtplans/mo.doc

Drake, G. (2018). *When more is less* [Blog post]. Retrieved from https://www.nctq.org/blog/When-more-is-lesstin

Education First. (2023). Reform support network – Quality evaluation rollout work group. Retrieved from https://www.education-first.com/impact/featured-story/reform-support-network-quality-evaluation-rollout-work-group/

Education Week Staff. (2011). AYP - Adequate yearly progress. *Education Week*. Retrieved from https://www.edweek.org/policy-politics/adequate-yearly-progress/2004/09

Elam, S (1971). *Performance based teacher. What is the state of the art*? Retrieved from ERIC database. (ED058166)

Every Student Succeeds Act. (2015). Retrieved from ttps://www.ed.gov/essa?src=rn

Finn, J. D. (1993). *School engagement & students at risk*. Retrieved from ERIC database. (ED362322)

Great Schools Partnership. (2014a). *Classroom observation*. Retrieved from https://www.ed glossary. org/classroom-observation/#:~:text=A%20classroom% 20observation %20is%20a, classroom %20or%20other%20learning%20 environ ment.& text= School%20 administrators%20also%20regularly%20observe, of%20formal% 20job%2Dperformance%20evaluations

Great Schools Partnership. (2014b). *Formative assessment*. Retrieved from https://www. edglossary. org/formative-assessment/

Great Schools Partnership. (2018). *Student engagement*. Retrieved from https://www. edglossary.org/student-engagement/#:~:text=LAST%20 UPDATED% 3A%2002. 18.16,and%20 progress%20in%20their%20education.

Goe, L., Holdhelde, L., & Miller, T. (2011). *A practical guide to designing comprehensive teacher evaluation systems*. Retrieved from ERIC database. (ED520828.pdf)

Goldhaber, D., &Anthony, E. (2003). *Indicators of teacher quality*. Retrieved from ERIC database. (ED478408)

Goldhammer, R. (1969). *Clinical supervision: Special methods for the supervision of teachers*. New York, NY: Holt, Rinehart & Winston.

Griffey, D., & Housner, L. (1991). Differences between experienced and inexperienced teachers' planning decisions, interactions, student engagement, and instructional climate. *Research Quarterly for Exercise and Sport,* 62(2), 196-204, doi.org/10.1080/02701 367.1991.10608710

Hattie, J., & Yates, G. (2013). *Feedback*. *Effective feedback in the classroom*. Retrieved from https://visible-learning.org/2015/02/feedback-in-the-classroom/

Hattie, J. (2008). *Visible learning*. New York, NY: Routeledge.

Harbour, K., Evanovich, L., Sweigart, C., & Hughes, L. (2015). A brief review of effective teaching practices that maximize student engagement. *Preventing School Failure: Alternative Education for Children and Youth, 59*(1), 5-13, doi:10.1080/1045988X.20 14.919136

Horn, A. S., & Jang, S. T. (2017). The impact of graduate education on teacher effectiveness: Does a master's degree matter (MHEC Research Brief)? *Midwestern Higher Education Compact*. ERIC. Retrieved from ERIC database. (ED587431)

Hunter, M. (1982). *Mastery teaching*. Thousand Oaks, CA: Corwin Press.

Kini, T., & Podolsky, A. (2016). *Does teaching experience increase teacher effectiveness? A review of the research.* Retrieved from https://learningpolicyinstitute.org/produc t/does-teaching-experience-increase-teacher-effectiveness-review-research .

Klein, A. (2015, April 10). No child left behind: An overview. *Education Week.* Retrieved from https://www.edweek.org/policy-politics/no-child-left-behind-an-overview/2015/04

Ladd, H. (2013). *Why experienced teachers are important – and what can be done to develop them.* Retrieved from https://scholars.org/sites/scholars /files/ssn_basic_facts_ladd_ on_the_importance_of_ experienced _teacgers_1.pdf

Linn, R. L., Baker, E. L., & Betebenner, D. W. (2002). Accountability systems: implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher, 31*(6), 3-16. https ://doi.org/10.31 02/0013189X031006003

Lunenburg, F., & Irby, B. (2008). *Writing a successful thesis or dissertation: Tips and strategies for students in the social and behavioral sciences*. Thousand Oaks, CA: Corwin Press.

Marzano, R. J. (2012). The two purposes of teacher evaluation. *Association for Supervision and Curriculum Development, 80*(3), 14-19.

Marzano, R. J., Frontier, T., & Livingston, D. (2011). *Effective supervision: Supporting the art and science of teaching*. Retrieved from

https://www.readpbn.com/pdf/Effective-Supervision-Supporting-the-Art-and-Science-of-Teaching-Sample-Pages.pdf

McMillan, J., Venable, J., & Varier, D. (2013). Studies of the effect of formative assessment on student achievement: So much more is needed. *Practical Assessment, Research, and Education,*18, 1-15. doi.org/10.7275/tmwm-7792

National Archives and Records Administration. (2016). Race to the top. National Archives and Records Administration. Retrieved August 8, 2022, from https://obamawhitehouse.archives.gov/issues/education/k-12/race-to-the-top

National Center for Educational Statistics. (2021). *Teacher qualifications.* Retrieved from https://nces.ed.gov/fastfacts/display.asp?id=58

National School Board Association. (2005). *Center for public education*. Retrieved from https://www.nsba.org/Services/Center-for-Public-Education

Network for Educator Effectiveness. (2015). Classroom observation & walkthrough training. Retrieved from neeadvantage.com/observation-of-teachers-in-the-classroom/

Nittler, K. (2018). *How do school districts compensate teachers for advanced degrees*? Retrieved from https://www.nctq.org/blog/How-do-school-districts-compensate-teachers-for-advanced-degrees

Ozan, C., & Kincal, R. Y. (2018). The effects of formative assessment on academic achievement, attitudes toward the lesson, and self-regulation

skills. *Educational Sciences: Theory and Practice*, (*18*)1, 85-118. Retrieved from ERIC database. (EJ1179831)

Park, S. Y. (2005). Student engagement and classroom variables in improving mathematics achievement. *Asia Pacific Education Review*, *6*(1), 87-97. Retrieved from ERIC database. (EJ728830)

Podolsky, A., Kini, T., Bishop, J., & Darling-Hammond, L. (2016, September 15). *Solving the teacher shortage: How to attract and retain excellent educators*. Retrieved from https://learningpolicyinstitute.org/product /solving-teacher-shortage

Punyanunt-Carter, N. & Carter, S. (2015). Students' gender bias in teaching evaluations. *Higher Learning Research Communications*, *5*(3), 28-37. Retrieved from ERIC database. (EJ1133098)

Reform Support Network. (2012). *Race to the Top at a glance. Evaluations of teacher effectiveness: State requirements for classroom observations*. Retrieved from http://www.isnetworked.org/uploads_regional/nyc/ Resources%20for%2 0Faculty/1373655392_Race%20to%20the%20 Top%20Reforms%20for%20Teacher%20Evaluations.pdf

Robinson, S. (2018). *Teacher evaluation: Why it matters and how we can do better*. Retrieved from https://www.frontline education .com/teacher-evaluation/

Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2008). *Can you recognize an effective teacher when you recruit one*? Retrieved from https://www.nber.org /papers/w14485

Ruff, R. R. (2019). State-level autonomy in the era of accountability: A comparative analysis of Virginia and Nebraska education policy through No Child Left Behind. *Education Policy Analysis Archives, 27,* 1-31. Retrieved from http://dx.doi.org/10 .14507/ep aa.27.4013

Selke, M. (2001). The professional development of teachers in the United States of America: The practitioners' master's degree. *European Journal of Teacher Education*, *24*(2), 205-214. doi:10.1080/02619760120095598

Stallings, J., & Krasavage, E. (1986). Program implementation and student achievement in a four-year Madeline Hunter follow-through project. *The Elementary School Journal, 87*, 2. https://www.journals.uchicago.edu/doi/abs/10.1086/461483

Taylor, E., & Tyler, J. (2021) Can teacher evaluation improve teaching? Evidence of systematic growth in the effectiveness of midcareer teachers. *Education Next, 12*(4). Retrieved from https://www.educationnext.org/can-teacher-evaluation-improve-teaching/

Tibbitt J. (2020). Formative assessment: A tool for closing achievement gaps in diverse classrooms. *New Directions in Deaf Education, 2*, 72-75.

Toch, T., & Rothman, R. (2008). *Rush to judgement: Teacher evaluation in public education.* Education Sector Reports. Retrieved from ERIC database. (ED502120).

Tracy, S. (1995). How historical concepts of supervision relate to supervisory practices today. *The Clearing House, 68*(5), 320-325. Retrieved from http://www.jstor.org/stable /30189094

Tran, H., & Buckman, D. G. (2020). The relationship between districts' teacher salary schedule structures and the qualifications of their teacher staffing profile. *Journal of School Administration Research and Development*, 5(1), 6-15.

U.S. Department of Education. (2013). *Nine districts submit waiver for relief from NCLB.* Retrieved from https://www.ed.gov/news/press-releases/states-granted-waivers-no-child-left-behind-allowed-reapply-renewal-2014-and-201

Warring, D. F. (2015). Teacher evaluations: Use or misuse? *Universal Journal of Educational Research, 3(*10), 703-709. doi:10.13189/ujer.2015.031007

Wetzel, W. A. (1929). Scientific supervision and curriculum-building. *The School Review*, *37*(2), 117–123. Retrieved from jstor.org/stable/1078902

Williams, K., & Herbert. D. (2020). Teacher evaluation systems a literature review on issues and impact. *Research Issues in Contemporary Education*, *5*(1), 42-50. Retrieved from ERIC database. (EJ1255555).

Appendices

Appendix A. Baker University IRB

Baker University Institutional Review Board

July 15th, 2022

Dear Michelle Edwards and Jim Robins,

The Baker University IRB has reviewed your project application and approved this project under Exempt Status Review. As described, the project complies with all the requirements and policies established by the University for protection of human subjects in research. Unless renewed, approval lapses one year after approval date.

Please be aware of the following:

1. Any significant change in the research protocol as described should be reviewed by this Committee prior to altering the project.
2. Notify the IRB about any new investigators not named in original application.
3. When signed consent documents are required, the primary investigator must retain the signed consent documents of the research activity.
4. If this is a funded project, keep a copy of this approval letter with your proposal/grant file.
5. If the results of the research are used to prepare papers for publication or oral presentation at professional conferences, manuscripts or abstracts are requested for IRB as part of the project record.
6. If this project is not completed within a year, you must renew IRB approval.

If you have any questions, please contact me at npoell@bakeru.edu or 785.594.4582.

Sincerely,

Nathan Poell, MLS
Chair, Baker University IRB

Baker University IRB Committee
Sara Crump, PhD
Nick Harris, MS
Susan Rogers, PhD

Appendix B. District 1 Approval Letter

[redacted]

July 20, 2023

To Whom It May Concern,

This letter is to confirm our support of Michelle Edwards, Program Coordinator at [redacted] High School, as she completes her dissertation. Michelle's research topic is "The Relationship Between Teacher Evaluation Scores and Teachers' Years of Service and Level of Education."

In conjunction with her research, Michelle received permission to utilize [redacted] District data.

If you have any questions regarding this matter, please do not hesitate to call me at 816-986-1004.

Sincerely,

David Carlson, Ed.D.
Assistant Superintendent of Human Resources

Appendix C. NEE Data Sharing and Use Agreement

NETWORK for EDUCATOR EFFECTIVENESS

NEE Data Sharing and Use Agreement

I. Introduction

This Agreement is entered into between The Curators of the University of Missouri on behalf of the College of Education's Network for Educator Effectiveness (collectively "NEE") and Michelle Edwards ("Data Recipient"). The purpose of this Agreement is to document the terms under which NEE will provide access to certain data and the Data Recipient will use the data.

In signing this Agreement, the Data Recipient agrees to the following terms and conditions of data use:

II. Terms and Conditions

1. Data Recipient agrees to use the data for research, analytics, and instructional purposes only.

2. Data Recipient agrees that the research plan has been approved by an appropriate research oversight, IRB, or human subjects office. NEE is not liable for any unauthorized use of the data by Data Recipient or any third party.

3. Data Recipient will not use, nor permit others to use, the data in any manner except that explicitly stated in this Agreement.

4. Data Recipient shall require all employees, agents, or contractors affiliated with Data Recipient that use or access the data (each an "Individual Data User") to sign this Agreement and will submit the signed Agreement(s) to NEE.

5. Data Recipient agrees not to modify or create derivative works of the NEE data without the express written consent of NEE and further agrees not to sell, disclose, transfer, share, or rent any NEE data to third parties. Data Recipient also agrees not to translate, convert, adopt, alter, modify, enhance, add to, delete, or tamper with any data unless expressly authorized by NEE in writing. Data Recipient agrees to refer all requests for access to the data to NEE.

6. Data Recipient will not attempt any linkage or combination of NEE data to any other data set for any other purpose, unless agreed upon in writing.

7. Data Recipient understands that NEE has de-identified the data set to the best of its ability. Data Recipient agrees not to attempt, in any way, to re-identify or otherwise discern any person or school included in these data.

8. Data Recipient agrees to make no disclosure or use of the identity of a person or school discovered inadvertently (or by any other means) and will advise NEE of any such discovered identity within two (2) business days of the date of discovery. If such a discovery is made, the information that would identify the individual or school will be safeguarded or destroyed by Data Recipient as requested by NEE.

9. Data Recipient agrees not to report information on any small cells (less than 10 subjects).

10. Data Recipient agrees to the following security procedures:
    a. Each Individual Data User will password protect any analysis files, such as those produced by a statistical analysis package.

    b. Data Recipient will treat the data as confidential and not give other persons access to it, except as under condition #4 above.

    c. Data Recipient will ensure that the data only resides on computers and systems that Individual Data Users own and operate, and only those Individual Data Users who have signed this Agreement.

11. Neither the Data Recipient nor the Individual Data User shall acquire any ownership interest in the data.

12. The Data Recipient is liable to NEE for any actions, inactions, and compliance with the terms of this Agreement, and Data Recipient agrees to indemnity, defend and hold harmless NEE, its officers, employees, and affiliates from any and all claims and expenses, including legal expenses and reasonable attorneys' fees, arising out of Data Recipient's use of the data or a violation or breach of this Agreement.

13. THE DATA IS DELIVERED "AS IS" IN EVERY RESPECT. NEE, ITS CURRENT OR FORMER CURATORS, OFFICERS, EMPLOYEES, AND AFFILIATES MAKE NO REPRESENTATIONS AND EXTEND NO WARRANTIES OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO WARRANTIES OF COMMERCIAL UTILITY, MERCHANTABILITY, FITNESS FOR ANY PARTICULAR PURPOSE, THE ABSENCE OF LATENT OR OTHER DEFECTS.

III. Project Details

Project Title: The Relationship Between the Teacher Evaluation Scores and Teachers '

Project Sponsor: Baker University

IRB approval was provided by:

Project Description:

Project Description: The purpose of this study was to determine if th

IV. Publication and Authorship

The Data Recipient will have the right to publish research based on NEE data (but not the NEE data itself unless expressly agreed to by NEE in writing) as long as conditions of this Agreement are met.

1. Data Recipient agrees to include relevant NEE personnel who have contributed substantially to designing the data collection, conducting analysis, and reporting of the data as co-authors of any publications or public dissemination of findings, with the permission of the relevant NEE personnel. Authorship shall be determined according to guidelines in the APA Publication Manual (6th edition), pp. 18-19.

2. Data Recipient agrees not to imply or state that interpretations based on the data are those of NEE without NEE permission in all publications and presentations.

3. Data Recipient agrees to provide NEE with a courtesy copy of any results and presentations of any Individual Data User analysis prior to their release.

V. Service Provided by NEE

Pursuant to this Agreement, NEE will pull from its archives data that it deems appropriate, in NEE's sole discretion, to Data Recipient's research project in an agreed upon format (EXCEL, SPSS, etc.). NEE will provide a codebook as agreed upon by NEE and the Data Recipient. NEE will use a secure data transfer mechanism (e.g., UM Secure Transmit) as agreed upon by NEE and the Data Recipient.

- Description of Data (including academic year):

  NEE summative evaluation data from 2015- 2019 (5 years) for indicators 1.2 and 7.4 for the three middle schools and three high schools in the Lee's Summit school district. Broken apart by teachers' level of experience and teacher level of education

- Data Transfer Mechanism:

  Sent by email to Michelle Edwards school email address.

- Permitted Use by Data Recipient:

  Dissertation to be submitted to Baker University

The Data Recipient will not use the data for any other applications or use. Should the Data Recipient desire to use the data for any purpose other than the permitted use expressly described herein, a separate and additional Agreement must be executed by the parties to include any additional terms and conditions related to usage and fees.

VI. Fees

The Data Recipient will provide a fee of $ 0.00 to NEE to cover administrative and personnel cost of generating and providing the data and codebook. Data will be transferred only upon receipt of the fee.

VII. Term and Termination

The term of this Agreement will be effective upon IRB app. and terminate after one year. NEE and Data Recipient may renew this Agreement by mutual consent, which shall be in their sole discretion, or the parties may enter into a new agreement.

Upon any failure of the Data Recipient to abide by the terms of this Agreement (including, but not limited to violations reported to the IRB or Federal Office for Human Research Protections), NEE shall have the right to immediately terminate this Agreement without an opportunity to cure. Upon termination or expiration of this Agreement, the Data Recipient shall destroy all copies of data in the Data Recipient's possession, as well as in the possession of any of the Data Recipient's employees, agents, assigns, and subcontractors, as directed by NEE.

In any action brought by NEE under this Agreement in which NEE prevails, NEE shall be entitled to its attorney's fees and court costs.